Investigating the role of effector proteins in the rice blast fungus *Magnaporthe oryzae*

Submitted by Vincent Mbashira Were to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Biological Sciences

In February 2018

I certify that all material in the thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature…….

Vincent Were

Abstract

To cause disease in plants, microbial pathogens secrete effector proteins that can suppress basal plant immunity mechanisms and help facilitate proliferation of the pathogen in plant tissue. The rice blast fungus, *Magnaporthe oryzae*, causes the most serious disease of cultivated rice. During early biotrophic colonisation, this fungus evades the plant immune system via the action of secreted effector proteins, allowing penetration of individual rice cells by invasive hyphae. The ability of the *Magnaporthe* isolates to infect weeping lovegrass (*Eragrostis curvula*) is controlled by a single gene, *PWL2*. Pwl2, like other effector proteins and several biotrophy-associated secreted (Bas) proteins, is secreted into a structure referred to as the biotrophic interfacial complex (BIC) before translocating into the cytoplasm [1-3].

In this study, I have used comparative genomics to analyse the relationship between the effector repertoire of *M. oryzae* and their potential recognition as avirulence determinants in rice blast populations in Sub-Saharan Africa (SSA). During effector-triggered immunity, effectors can be recognised as avirulence gene products by rice resistance (R) proteins. Currently, 25 rice blast resistance genes have been cloned along with 10 cognate avirulence genes [4]. I have used third-generation Pacbio RSII sequencing technology to generate improved genome assemblies of two *M. oryzae* strains, wild type Guy11, and KE002, a Kenyan isolate that is avirulent on selected rice monogenic lines and is thought to carry many avirulence genes. I show that by using improved contiguous genomes, in combination with RNA-sequencing data, it is possible to establish a gene prediction pipeline that can identify isolate-specific or novel

small secreted protein encoding genes. With this approach, I have identified additional 49 and 590 genes in Guy11 and KE002 genomes respectively. Three of them, *MEP13, MEP15* and *MEP14*, have been confirmed to encode secreted effector proteins.

I also provide new insight into how effectors are secreted and delivered across the fungal plasma membrane. I report the use of high resolution microscopy and fluorescently labelled effector proteins to show that the biotrophy interfacial complex (BIC) is a plant derived structure through which the fungus continually secretes effectors in one direction into the targeted plant organelles. Using fluorescently labelled Pwl2 and Bas1, I present results suggesting that effectors translocate across the plasma membrane packaged in vesicles which are visible in the BIC, and that these effectors are sorted into distinct vesicles before crossing the plasma membrane. I then use CRISPR/Cas9 gene editing to generate *Δpwl2* mutants in *M. oryzae* strain Guy11 which has three copies of the gene resulting in deletion of all copies of the effector gene and gain of virulence towards weeping lovegrass. This confirmed the function of *PWL2* as a host range determinant that is conserved in most rice blast isolates. These results show that genes with multiple copies in the genome can be functionally characterised through either disruption or replacement using CRISPR/Cas9 in *M. oryzae*.

Contents

Appendix

Mutiga, S. K., Rotich, F., Devi Ganeshan, V., Mwongera. D.T, Mgonja, E.M., Were. V., Harvey, J., Zhou, B., Wasilwa, L., C. Feng., Ouedraogo I., Wang, G-L., Mitchell, T., N. Talbot., and Correll, J. C (2017) Assessment of the virulence spectrum and its association with the genetic diversity in a *Magnaporthe oryzae* population from sub-Saharan Africa. Phytopathology

List of Figures

List of Tables

Abbreviations

bp	base pair
cDNA	complementary DNA
CM	complete media
DNA	deoxyribonucleic acid
DIC	differential interference
EDTA	ethylenediaminetetraacetic acid
EIHM	extra-invasive hyphal membrane
g	grams
GFP	green fluorescence protein
CTAB	hexadecyltrimethylammoniun bromide
Hpi	hours post infection
Hpi	hours post inoculation
IH	invasive hyphae
kb	kilo base
L	litre
Mg	microgram
µL	microlitre
µM	micromolar
mg	milligrams
mL	mililitre
mM	milimolar
min	minute
M	molar
ORF	open reading frame
%	percentage

w/v	weight by volume
PCR	polymerase chain reaction
QRT-PCR	quantitative real-time polymerase chain reaction
RFP	red fluorescence protein
mRFP	monomeric red fluorescence protein
RPM	revolutions per minute
RNA-seq	RNA sequencing
WT	Wild type
NGS	next generation sequencing
Mb	Megabases (Mb)
NLR	Nucleotide-binding and leucine repeat domain
HMA	heavy-metal associated
BIC	biotrophic interfacial complex
HMW	high molecular weight
Sec	second
LB	Lysogeny broth
°C	degrees Celsius
coIP	co-immunoprecipitation
BLAST	Stand-alone Basic Local Alignment Search Tool
SSA	Sub-Saharan Africa
MAX	*Magnaporthe* AVRs and ToxB like

Acknowledgements

I would like to thank the BBSRC, Bill and Melinda Gates Foundation for providing me with the funding to carry out this research. In a special way, I would like to thank Les and Claire Halpin, for the generous sponsor of the Halpin Trust that paid my scholarship and gave me the opportunity to attend the best scientific courses, conferences and workshops during my PhD. I would also like to express my sincere heartfelt gratitude to my supervisor Professor Nick Talbot for his continuous support, motivation, encouragement, and for sharing his immense scientific knowledge, and challenging me throughout my PhD. I am forever grateful for the opportunity to learn from the best molecular genetics scientist, it's been an honor. I would to thank everyone involved in the SCPRID project, Jim, Guo-liang, Tom, Bo, Samuel, Felix, Ibrahima and Moussa for sharing the pathotype data and rice monogenic lines used in my studies.

I have learnt a lot in the past four years from my laboratory colleagues. First, I would like to thank Mick, from whom I have learnt a lot, both professionally and personally. I would like to thank Darren for all bioinformatics support and training that expanded the breadth of my PhD research. I would also like to thank George for his help with Laser-confocal microscopy and Xia for her help in cloning. Special thanks to Lauren, Miriam, Magdalena, Yogesh, David, Clara, Richard, Bozeng, Wasin, Tina, Andy, Sabrina, Kate, Alice and to all present and past member of the Halpin Lab who have been amazing, supportive, helpful colleagues and friends. An extended gratitude goes to Jane Usher for sharing her experience with DNA extraction, and to Karen, Paul, Jeremie and Aaron at Exeter sequencing services. I would also like to thank my previous supervisors Yasmina and Mary at the University of Queensland, for

supporting me during my Masters' degree studies and to Jagger for supporting me during the start of my career and being a good mentor. Many thanks to my good friends in Exeter, who have made my time here special and have made me, feel at home. Special thanks to Rik, Will, Mark, Dominic and Freddie who have been there for me throughout the four years.

I would also like to thank my family back home for always supporting and believing in me. Thanks to my mum for your prayers, I know you lacked the opportunity to do this, I did it for you! Thanks to my dad, my sisters Edith and Anne and my brothers Patrick and Michael, and finally Tenaya, may you grow up and achieve your dreams. Many thanks to Silo, Kush, Paul, Seth, for the years of friendship and to all my friends who've been there for me despite the distance, I do not take you for granted. Finally, I would like to thank my lovely wife, Christine for your love, support, patience, encouragement and being with me throughout this journey.

Chapter 1 General Introduction

1.1 *Magnaporthe oryzae* is an immediate concern for global food security

The global human population has been projected to increase to almost 9 billion by the year 2050, which represents a 6-fold increase since Food Agricultural Organisation (FAO) statistics in 1990 [5, 6]. This creates a need for a three-fold increase in food production in the next 30 years to meet growing demand. Currently, more than 3 billion people are believed to depend on rice as a source of their calorific intake [7]. Africa is slowly becoming a major player in the international rice trade because it counts for up to 32 % of global imports, an equivalent of 9 million tonnes annually (as per 2006 statistics) [8]. The increase in importation is explained by the fact that rice is rapidly growing as a food source in sub-Saharan Africa, to satisfy the consumer preference of a growing population, especially in urban areas [8, 9].

Rice consumption in Africa has surpassed domestic rice production. The development of new rice cultivars - New Rice for Africa (NERICA), a cross between *Oryza glaberrima* Steud (African rice) and *Oryza sativa* L. (Asian rice), was welcomed by farmers [10]. These varieties have a high yield, early maturity, and possess increased resistance to pests [10]. However, rice blast disease remains one of the biggest threats to global rice production [11-13]. This disease is caused by the fungus *Magnaporthe oryzae,* which despite efforts to control it, still destroys 10-35% of the global rice harvest, or sufficient rice to feed 60 million people [14]. Efforts to control the disease, including use of fungicides, avoiding excessive use of nitrogen-based fertilisers and planting resistant rice varieties, have not been successful [15].

1.2 Rice blast disease

Rice blast disease is caused by *M. oryzae* (synonym of *Pyricularia oyzae*) [16, 17], a filamentous ascomycete capable of affecting almost all parts of the rice plant including the panicle, neck and stem. The fungus has also been reported to be able to infect roots and proliferate systemically into rice plants [17]. *M. oryzae* is able to infect more than 50 grass species, including economically important crop species, such as finger millet (*Eleusine coracana*) barley (*Hordeum vulgare*) and wheat (*Triticum aestivum*), but some host-limited strains are generally restricted to a single host species [13, 18-20]. Wheat blast has been reported to have devastating effects especially in the province of Parana', Brazil, where it was first observed in 1985 [20, 21]. Wheat blast disease is now one of the most significant diseases affecting wheat and outbreaks have been recently reported in Bangladesh and India [22, 23]. The ability to culture and genetically manipulate the rice blast fungus has made *M. oryzae* a very important model organism in understanding the plant-pathogen interactions. In addition, the availability of the complete genome of both the fungus and its host offers a molecular toolkit for gene identification and characterisation [24, 25].

1.3 Life cycle of *M. oryzae*

Rice blast infection commences when a three-celled spore called a conidium, lands on the leaf surface either through wind or splash dispersal [12]. To enable attachment and anchorage to the waxy hydrophobic surface of the leaf, the conidium releases adhesive spore tip mucilage [26]. Approximately 2-3 h after landing on the leaf surface, in the presence of water, the conidium

germinates, producing a single polarised germ tube [12, 27]. The germ tube then flattens on the leaf surface and starts to swell at its tip. This stage allows the conidium to recognize the rice leaf surface, prior to appressorium development [26]. Under laboratory conditions, a hydrophobic surface can be used to induce formation of an appressorium as shown in Figure 1.1 [28].

The appressorium is defined as a dome-shaped pressurised cell with a highly differentiated cell wall that contains a layer rich in chitin and a thick layer of melanin sandwiched between the cell membrane and the cell wall. The melanised layer allows accumulation of glycerol inside the cell generating high internal turgor pressure [29, 30]. Additionally, the spore collapses and the nuclei are degraded by autophagy and the spore contents transferred into the appressorium [31, 32]. Autophagy has been shown to be important for appressorium maturation and penetration [31, 32]. The pressure accumulated in the appressorium is sufficient to physically puncture the rice cuticle [29]. Once in the initial epidermal rice cell, the penetration peg undergoes differentiation, which results in development of primary invasive hyphae. These later differentiate into bulbous invasive hyphae that colonise the whole cell and moves into neighbouring cells [12]. Four to five days after infection, disease symptoms in the form of large necrotic spots appear on the surface of infected leaves. The fungus then sporulates from these necrotic lesions and spores are dispersed to begin the cycle again [12].

Development of the appressorium is regulated by two independent S-phase checkpoints of the cell cycle [33]. Initial appressorium formation is mediated by an S-phase checkpoint acting through the DNA damage response

(DDR) pathway, a process shown to require the cds1 kinase [33]. However, recent studies have shown that the next step of appressorium repolarisation involves a DDR -independent S-phase checkpoint triggered by melanisation and turgor pressure generation [33]. This checkpoint also regulates septin-dependent reorientation of the F-actin cytoskeleton at the base of the appressorium, essential for host tissue penetration [33]. The appressorium is made of a highly differentiated chitin rich cell wall and contains a melanised layer between the cell wall and cell membrane [34]. Melanin prevents escape of compatible solutes from the appressorium, allowing accumulation of high turgor pressure [29, 35, 36].

Accumulation of up to 3.2 M glycerol within the appressorium causes influx of water, resulting in internal hydrostatic pressure estimated to be 8.0 Mpa - an equivalent to fifty times the pressure of car tyre. This immense pressure allows formation of a penetration peg at the bottom of the appressorium and physical piercing of the tough rice cuticle [29, 36]. At 72 h post-infection, it is estimated that the fungal biomass constitutes up to 10 % of the total leaf, at which point disease symptoms start to appear. Typical disease lesions are characterised by ellipsoid necrotic lesions on the leaf surface [12, 13, 27].

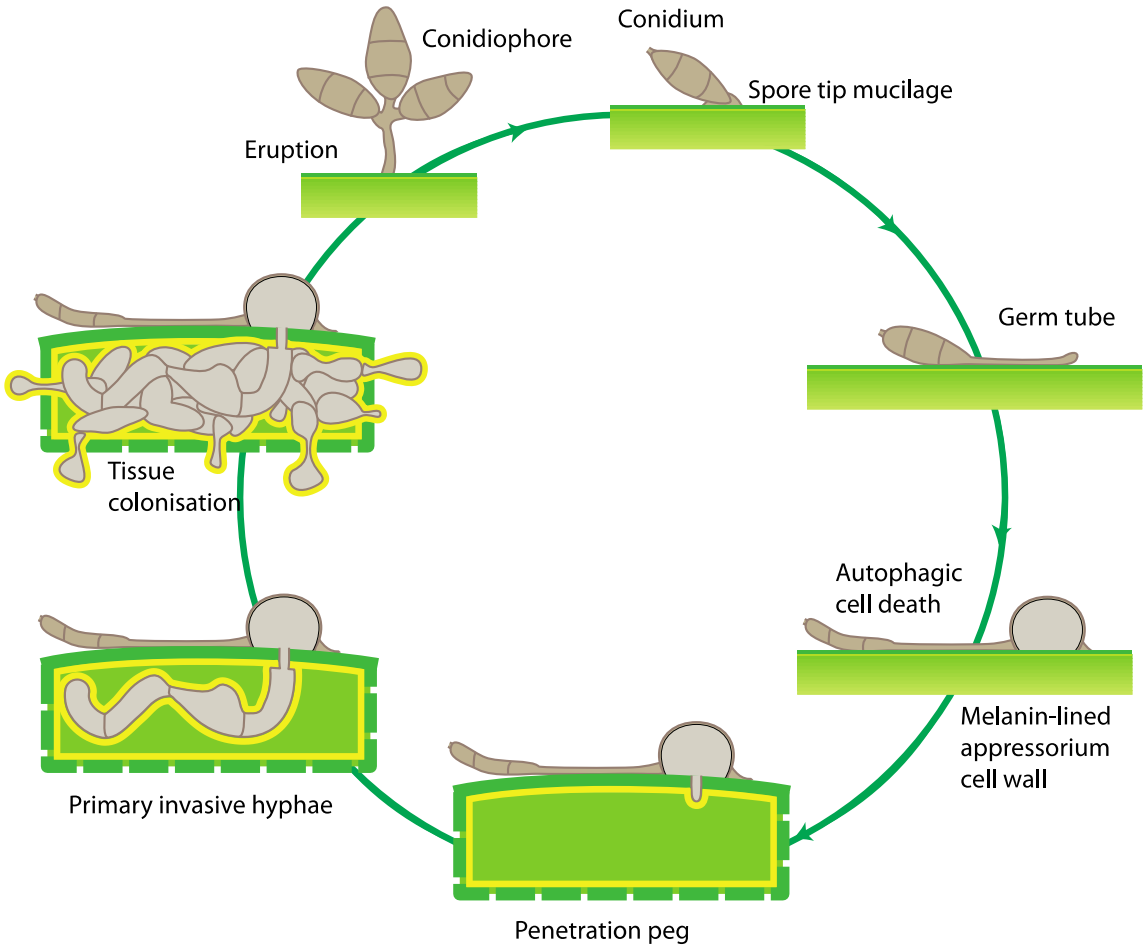

Figure 1.1 Life cycle of rice blast fungus *Magnaporthe oryzae*

The life cycle begins when a three-celled conidium lands on the rice leaf surface and germinates to form an appressorium. This leads to tissue invasion and then culminates in the next round of sporulation. Figure courtesy of Marion Littlejohn.

1.4 Genome sequencing in understanding fungal pathogen interactions and evolution

With the advent of next generation sequencing (NGS) and continuous development in this field, there have been a lot of advances in comparative genome studies in different fungal pathogens [24, 37-40]. In *M. oryzae*, more than 50 genome assemblies have been generated through NGS [41]. This has played a big role in enhancing the knowledge of the dialogue between the fungus and its host, rice, during infection, how the fungus undergoes host adaptation, specificity and, recently, host jumps and host range expansion [17, 20, 42-44]. The draft *M. oryzae* genome was originally published as a result of a whole-genome shot-gun sequencing approach [24, 42, 43]. Several comparative studies of NGS-assembled genome sequences show that some key features that determine virulence may not be well represented [41]. Next generation sequencing yield highly fragmented assembled genomes that lack good representation of isolate-specific regions [38, 45]. These regions are enriched with repeat sequences and effectors which contribute to genome plasticity [38, 45]. Genome variations have been observed in different isolates of a single species and unique regions carrying virulence factors or effector genes have been reported [43] [46].

In some unique cases, a variation in the number of chromosomes was also observed by comparative genomic studies of *Fusarium oxyporum*, *Mycosphaerella graminicola*, *Alternaria alternata*, *Leptosphaeria maculans*, *Nectria haematococca* and *Cochliobolus heterostrophus* [38, 47-52]. The variation in the number of chromosomes has been attributed to supernumerary

chromosome variation often associated with isolate-specific virulence [53], with small chromosomes ranging from 470 kb - 2.2 Mb identified [54]. A comparative genomics study using a field isolate, Ina168 and the laboratory isolate 70-15 was used to identify three avirulence gene *AVR-Pia*, *AVR-Pii* and *AVR-Pik* alleles located on a 1.6 Mb isolate specific region of Ina168 [46]. To elucidate the genetic basis of virulence and understand genomic variability, a comparative genome study of two field isolates, Y34 and P131, was carried out in 2012 [43]. This study identified regions of the genome under selection pressure and the importance of transposon-like elements in sequence diversification [43].

Genome sequences of two other field isolates, FJ81278 from Fujian and HN19311 from Hunan in China, were also compared to the reference genome 70-15 by Chen et al 2013 [55]. This study identified differences between the isolates in terms of single nucleotide polymorphisms, regions under positive selection and variation in genome structure. It was possible to identify inter-chromosomal translocation events and isolate-specific genome regions. The majority of observed inter-chromosomal translocation events occurred in the telomeric region. Moreover, more than 200 putative effectors and virulence determinant genes were identified [55]. Using RNA sequencing, it was possible to improve the quality of gene predictions [55]. High quality genome sequencing and assemblies are therefore urgently required to improve the outcome of studies in plant-microbe interactions.

1.5 Plant Immunity and defence against filamentous pathogens

Plants present different levels of defence against pathogens attack [56]. The first level involves the epidermal cell walls and waxy cuticle that offers the physical barriers [56, 57]. The next line of defence includes active recognition of the pathogen by the host plant [57, 58]. Plasma membrane-localised receptor proteins can recognize conserved molecules from micro-organisms including fungi [59-61]. These receptors can also recognise conserved protein motifs that might be shared in different plant pathogens and these molecules are referred to as pathogen-associated molecular patterns (PAMPs) and trigger the first line of the host immune system, which is known as PAMP-triggered immunity (PTI) [59-61]. The recognition of PAMPs by pattern recognition receptors (PRRs) located at the host cell membrane culminates in immune response signalling. This triggers production of reactive oxygen species, reprogramming of the host transcriptional profile, activation of ion-channels and induction of defence-related, mitogen-activated protein (MAP) kinase-signalling pathway that leads to activation of WRKY transcription factors [58].

In *Arabidopsis thaliana*, fungal chitin oligomers for example are recognised as PAMPs and directly bound by three extracellular LysM domains in the PRR and then perceived by the LysM-RLK (receptor like kinases) CERK1/RLK1/LYK1 receptor as shown in Figure 1.2 [60]. In bacteria, conserved molecules like flagellin, elongation factor Tu (EF-Tu), peptidoglycans and lipopolysaccharides are recognised by the plant immune system [62]. Although the mechanism through which microbial conserved molecules are recognised is partially conserved across the plant kingdom, the perception of

some epitopes might vary in different plant families [62]. Flagellin for example, contains a conserved 22-amino-acid epitope termed flg22, that is recognised by most plants through a leucine-rich repeat (LRR) receptor kinase called FLS2 [62, 63]. However, Flagellin epitopes can also be perceived via a FLS2-independent manner in tomato plant *Solanum lycopersicum* (including other species in the *Solanaceae* family) and in *Oryza sativa* [64]. Plants from the *Brassicaceae* family can perceive an 18-amino-acid epitope termed elf18, displayed by ET-Tu, through the LRR-receptor kinase EFR [62, 63]. However, in rice, a different epitope of ET-Tu comprising of Lys176 to Gly225 called EFa50 is recognised [65].

In fungal and oomycete pathogens, perception by plants is mostly dependent on recognition of cell wall main component, chitin in fungi and β–glucans in oomycetes. In *A. thaliana* for example, these chitin oligomers induce homo-dimerisaton of CERK1 essential for activation of downstream signalling [60]. In rice, in addition to the CERK1, a chitin elicitor binding protein (CEBiP) containing a LysM domain, is required to perceive chitin and initiate signalling. Fungal pathogens are known to secrete many small proteins that function to suppress host PTI [60, 66-68]. For example, the *ECP6* effector in the tomato leaf mould fungus *Cladosporium fulvum* is a LysM domain protein that competes for chitin oligomers thereby suppressing chitin-triggered immunity [68].

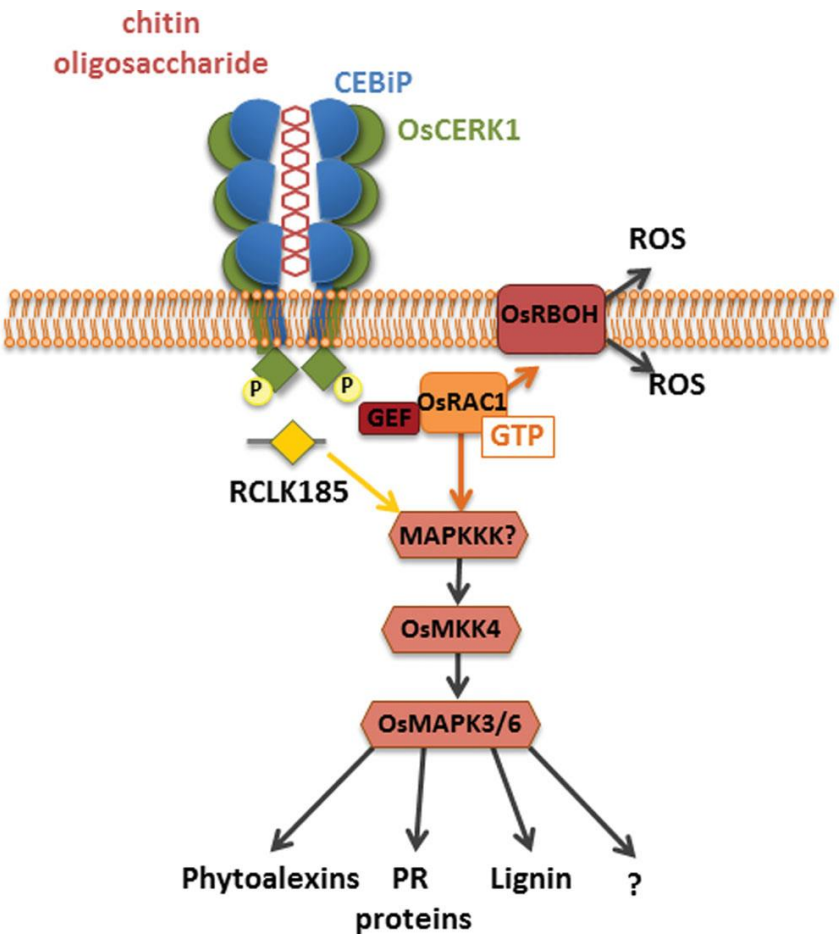

Figure 1.2 Schematic representation of the chitin receptor complex at the plasma and components involved in the cellular signalling pathway in rice.

The chitin receptor complex at the plasma membrane including all components involved in the rice cellular signaling pathway. LysM receptor CEBiP directly binds chitin; resulting in CEBiP homodimerization as the ectodomain of two CEBiP monomers bind the same chitin oligosaccharide from opposite sides. Furthermore, CEBiP heterodimerizes with OsCERK1 and its kinase domain is subsequently phosphorylated. In turn, OsCERK1 phosphorylates the downstream signaling component OsRacGEF1, which activates OsRAC1 which in turn activates a MAPK cascade, culminating in the expression of defense executers. OsCERK1 can also phosphorylate RCLK185 that activates the same MAPK cascade. This figure was adapted from Sanchez-Vallet *et al* [60].

To counter pathogens, plants have developed a third layer of defence called effector-triggered immunity (ETI) that is directed towards specific pathogenic strains [58, 59, 69]. This involves recognition of pathogen-specific effector proteins (avirulence proteins) by plant resistance proteins (R). Most R genes encode proteins containing a centrally situated nucleotide-binding site (NB-ARC) domain and leucine-rich repeats (LRRs) in the C- terminal. These proteins are referred to as nucleotide-binding oligomerisation domain-like receptors, or in short NOD-like receptors (NLRs). Many NLRs carry an additional coiled-coil (CC) or TOL1/interleukin receptor (TIR) domains in the N-terminal. These domains have been shown to be involved in the formation of an NLR homo complex, crucial for downstream signalling. Only effector proteins that translocate into the host cytoplasm can be recognised by the products of these R genes [70-73]. Initially, it was thought that recognition of these effectors was mediated by a single NLR, but there is increasing evidence of cases whereby more than one NLR can co-operate to recognise a pathogen and initiate an immune response [74-76] [77]. Some R genes like the Cf-resistance proteins that confer resistance to leaf mold caused by *C. fulvum* in resistant tomato lines, belong to a distinct class of R proteins that are collectively referred to as receptor-like proteins (RLPs) [78, 79]. This type of R genes encode proteins that possess an extracytoplasmic domain containing primarily of LRRs, a C-terminal membrane anchor and lack a nucleotide-binding site [78, 79].

In plant genomes, some genes encoding for NLRs can occur in paired, inverted tandem arrangements. In few cases, two NLRs may act as hetero-complexes; where only one NLR directly recognises the effector while the other plays a crucial role of downstream signalling [75, 76] [77]. Recently, several

experiments have demonstrated that NLRs can indirectly interact with effectors [77]. In this case, they can either recognise a modified protein targeted by the effector, also referred to as a ‘guardee’ or a modification in a host protein that mimics the effector targeted protein termed as “an integrated decoy” [75, 80, 81].

In the last 20 years 25 different rice blast resistance genes have been cloned, of which 24 code for NLRs [4, 82]. In rice, resistance against blast can be conferred by paired NLRs which cluster in a tandem organisation in the genome. *Pi-CO39/Pia* encode a pair of NLR Rga4/Rga5 which interact with either AVR-CO39 or AVR-Pia secreted by the rice blast fungus [83]. In this pair, Rga4 acts by constitutively inducing disease resistance and cell death [83]. However, Rga5 acts by repressing this activity in the absence of pathogen [83]. Additionally, the direct binding of Rga5 to *M. oryzae* secreted effectors AVR-CO39 or AVR-Pia causes de-repression of Rga4 which results in the activation of cell death and immune response signalling [83]. These effectors are recognised after binding to the ATX1 (RATX1) domain that is present in the C-terminal of Rga5 and is like a heavy-metal associated (HMA) domain from *Saccharomyces cerevisiae* [83, 84]. The rice NLR, Pik-1, has also been shown to contain an integrated HMA domain and acts together with the NLR Pik-2 in the recognition of another *M. oryzae* effector AVR-Pik (Figure 1.3) [85]. It is now hypothesised that effectors may target HMA proteins for disease development since another HMA containing protein Pi21 is known to be a susceptibility factor targeted for disease development [86]. Thus RATX1 has been termed an integrated decoy [70, 87, 88] [74] [83] [85] [80].

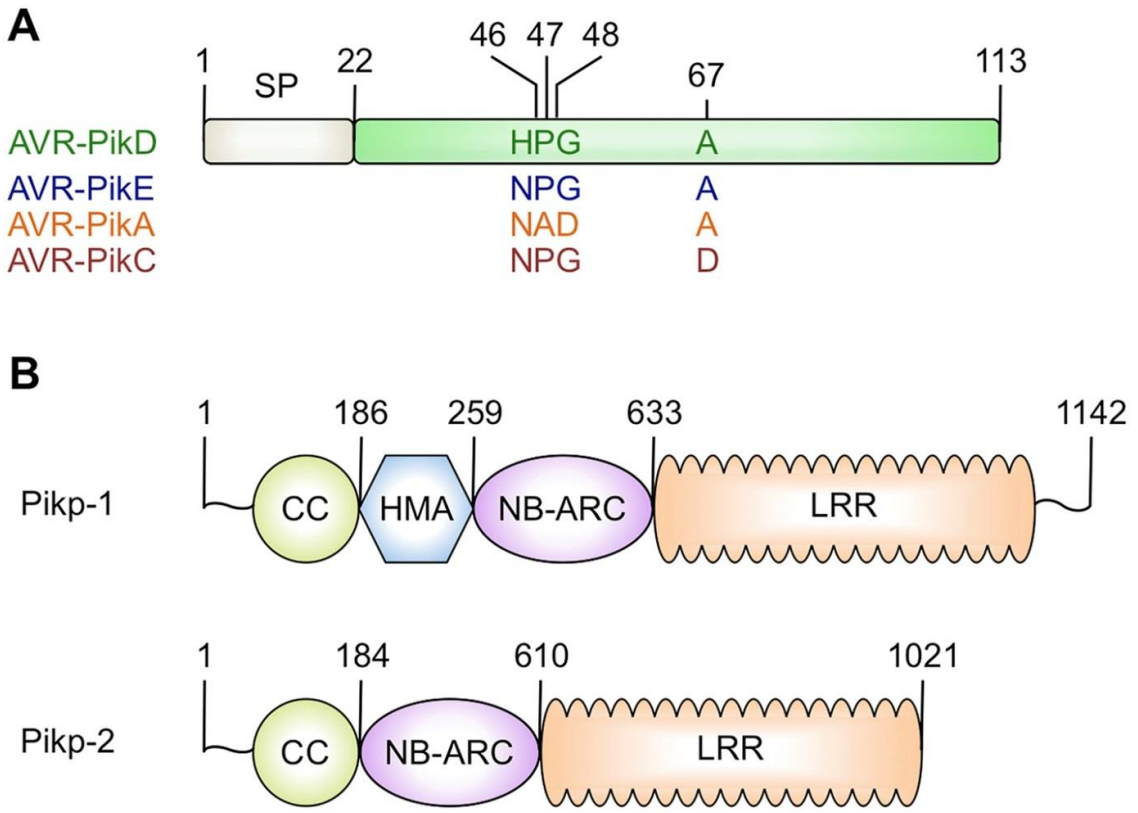

Figure 1.3 A simple representation of pathogen-host co-evolutionary interaction

(A) The *Magnaporthe oryzae* avirulence gene AVR-Pik has four distinct alleles. The position of polymorphic residues is indicated as 46, 47, 48 and 67. The signal peptide is shown in grey while green denotes the mature protein. (B) Rice Pik R proteins. Pik-1 contains an integrated HMA domain situated between the coiled domain and nucleotide-binding site (NB-ARC): and acts together with the NLR Pik-2 in the recognition of *M. oryzae* expressing AVR-Pik. This figure was adapted from Kanzaki *et al* [85]

1.6 Effector secretion, translocation and localisation

In the process of colonizing their hosts, fungal species have adapted different lifestyles that range from biotrophic, hemi-biotrophic to necrotrophic growth [89]. For a compatible reaction and successful proliferation, most fungi must avoid eliciting the host PTI, suppress the response or cope with plant defence mechanism. They do this via secretion of effector proteins that alter host cell physiology for the benefit of the pathogen [57]. Some of these effectors may be secondary metabolites with toxic activities, or necrotic effect on the host plant, especially in necrotrophic fungi or during the necrotic phase for hemi-biotrophic pathogen infection [90]. As a hemi-biotrophic pathogen, *M. oryzae* sustains host cell viability initially, before switching to a necrotic growth phase, and causing necrosis and disease lesion formation [90, 91]. During its early biotrophic colonisation, this fungus is known to secrete many effector proteins that help it evade the plant immune system and alter host physiology, allowing proliferation of invasive hyphae in rice cells. It's biotrophic phase of growth is characterised by presence of a biotrophic interfacial complex (BIC) in invaded cells, a structure thought to be the point where certain effectors are first secreted from invasive hyphae before delivery into host cells [3, 57, 91].

The BIC is a membrane-rich structure that appears at the tip of primary invasive hyphae and is later positioned sub-apically as the invasive hypha differentiates into a bulbous branched structure before penetrating adjacent cells [1, 2]. BIC formation and the secretion of effectors recur when invasive hyphae penetrate a new neighbouring living cell. When a *M. oryzae* strain expressing a fluorescently-labelled, BIC-localising effector, Pwl2, was

inoculated onto a rice line with a fluorescently-labelled, rice plasma membrane marker, LTi6B, the BIC was seen to co-localise with the host plasma membrane [2]. The structure was therefore demonstrated to be distinct from the fungal plasma membrane and cell wall.

The BIC is therefore a host plant-derived structure located at an interface outside the fungal cell wall and bounded by plant plasma membrane [2]. The exact function of the BIC is yet to be discerned but is thought to be a point at which effectors are concentrated before cytoplasmic translocation. Importantly, accumulation of effectors in the BIC is a fundamental characteristic of successful infection that is not observed in plants with cognate resistance genes, and is therefore currently used as a way of characterising novel effectors [1, 2]. Effectors that are expressed through the BIC are sequentially delivered into the host cytoplasm, where they can be recognised by resistance (R) proteins [1, 2]. Consequently, recognition of effectors (avirulence proteins) that can trigger a hypersensitive response (HR) will stop proliferation of invasive hyphae and lead to plant immunity from infection.

The transcription signal that initiates expression of effectors is thought to be triggered when *M. oryzae* lands on, and recognises, the host leaf surface suggesting that effector secretion starts before host cell penetration [1]. In the anthracnose pathogen *Colletotrichum higginsianum*, a subset of effectors has been shown to be expressed in the appressorium prior to penetration [92]. This has also been shown in *M. oryzae* where an effector called Mep1 which is expressed in appressoria prior to infection (Xia Yan and N.J. Talbot, unpublished observations). Using fluorescently-labelled effector proteins and

host cell components, it is becoming clearer how effectors are secreted, translocated and, in some instances, localised at cellular level during invasion [2]. This phenomenon is well studied in *M. oryzae* by observing invasive hyphae in translucent rice leaf sheath preparation. Jones et al [93] used a fluorescent dye, fluorescein diacetate (FDA) that stains the cytoplasm, to identify live cells, and propidium iodide (PI) that stains the nuclei of dead cells to test cell viability after invasion. Jones et al [93] showed that newly invaded cells remain alive and then die when the fungus moves from the first invaded cell and invades neighbouring cells. In a live cell, invasive hyphae are enclosed by an extra-invasive hyphal membrane (EIHM) which remains intact before the invaded cell dies [93].

Two localisation patterns have been determined in the rice blast fungus. Some effectors will localise in the space between the fungal cell wall and the EIHM and these effectors have been named apoplastic effectors as shown in Figure 1.4 [94]. This type of effectors includes Slp1, a secreted LysM protein 1, and the biotrophy-associated secreted protein 4 (Bas4) [1, 94]. Another set of biotrophy-associated secreted proteins, such as Bas1 are secreted into the biotrophic interfacial complex (BIC) in primary invasive hyphae as shown in Figure 1.4 [1]. Apart from Ace1 whose synthesised secondary metabolite is recognised as an avirulence determinant, all other known avirulence proteins like Pwl2, AVR-Piz-t, AVR-Pia, AVR-Pii, AVR-Pi9 and AVR-Pita have been shown to localise to the BIC before translocating into the host cytoplasm [2, 3, 42, 57, 95, 96].

The two types of effectors have also been demonstrated to be secreted through two different secretory pathways. Giraldo *et al* [2] showed that secretion of the apoplastic effector Bas-4 was inhibited after treatment with Brefeldin A (BFA), a chemical that inhibits secretion through the Golgi apparatus. However, BFA did not affect BIC-localising effectors such as Bas-1 or Pwl2. Moreover, the *M. oryzae* exocyst complex components Sec5 and Exo70 were shown to be important for efficient secretion of effectors through the BIC [2]. In *Δsec5* and *Δexo70* mutants expressing fluorescent labelled effectors, partial retention of the BIC-localised effectors in BIC-associated hyphae cells was observed. However, these mutants secrete apoplastic effectors normally [2].

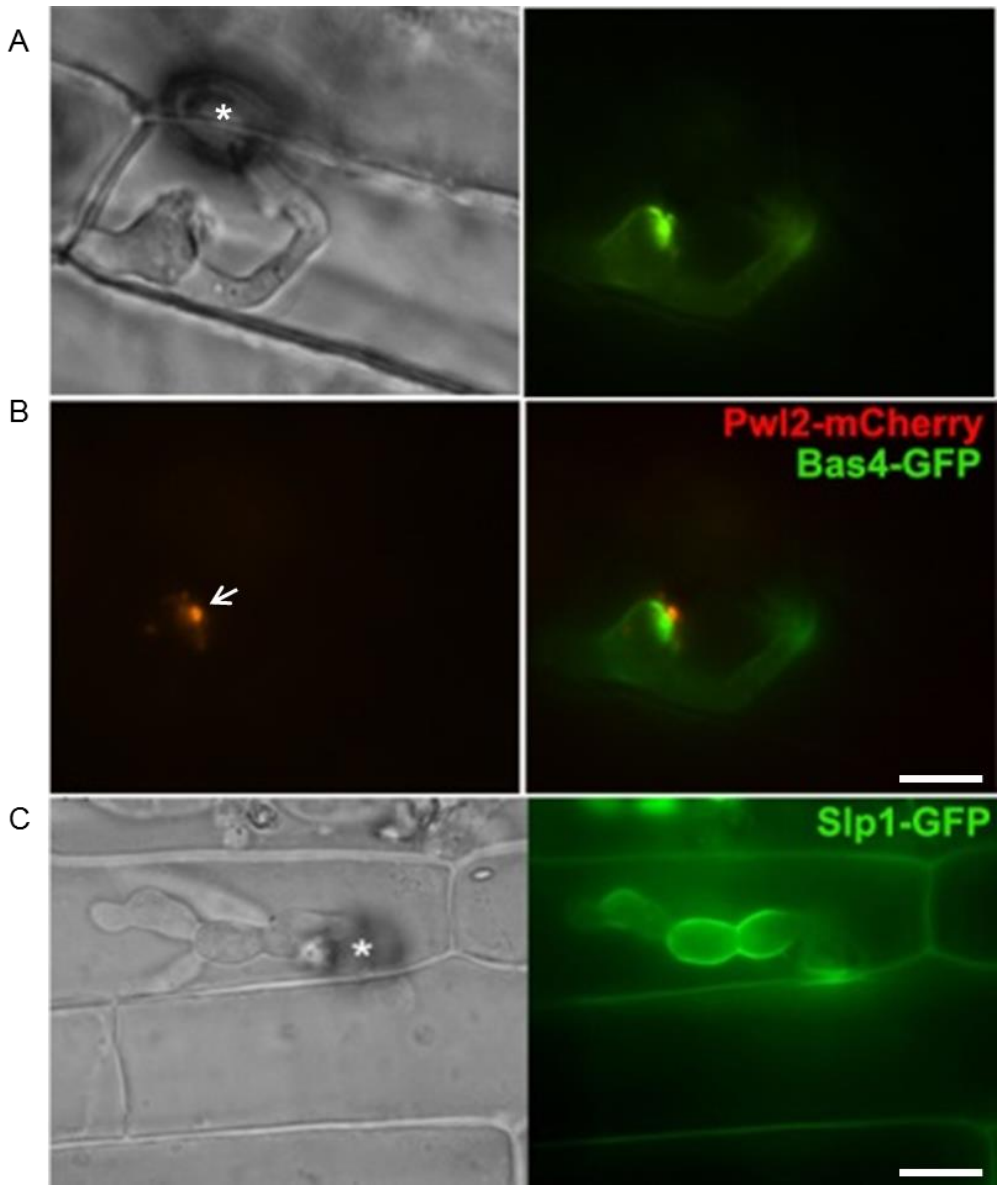

Figure 1.4 *M. oryzae* effectors exhibit different secretion and localisation patterns during rice blast infections

(A) Bas4 is an apoplastic effector that localises at the plant-fungal interface and also shows partial secretion into the BIC (B) Pwl2 is a cytoplasmic effector that localises into the BIC (Figures from this study). (C) Secretion of the apoplastic effector Slp1 in the plant-fungal interface. Slp1 has two LysM domains that bind chitin with high affinity and competitively inhibits CEBiP PRR. Figure from Mentlak *et al* [94]. White asterisks indicate the site of appressoria formation and the white arrow indicates the site of the BIC. Scale bar represent 10 µm

1.7 Effectors can act to alter host cell physiology and enhance host susceptibility

Most fungal effectors lack clear motifs or homology to known proteins that could be used to predict their function [97, 98]. In most cases, deletion of effector genes for functional analysis rarely shows any virulence phenotype and this is thought be due to functional redundancy of the resulting effector repertoire [95, 99]. For example, in a study targeting 78 *M. oryzae* genes for gene replacement, only one, *MC69* was associated with virulence [95]. Its orthologue in *Colletotrichum orbiculare* was also shown to be essential for pathogenicity [89, 95]. The function of some effectors has been characterised in different plant pathogens. In the bacterial pathogen *Pseudomonas syringae,* a ubiquitous virulence effector HopI1, for example, has been shown to localise to the chloroplast, the site for salicylate and jasmonic acid synthesis, causing remodelling of the thylakoid membrane and leading to inhibition of SA accumulation [100]. This effector interacts directly with Hsp70 to subvert its involvement in host defence by stimulating ATP hydrolysis [100]. Hopl1 possesses an N-terminal putative chloroplast targeting sequence that is also a characteristic of another *P. syringae* chloroplast targeting effector, AVRRps4 [100].

In *Ustilago maydis*, an apoplastic effector, Pep1, has been shown to protect invasive hyphae from reactive oxygen species (ROS) by acting as a peroxidase inhibitor [101]. In the same pathogen, an effector, Cmu1, was identified as targeting defence signalling and is important for virulence during colonisation. This effector is highly expressed during the biotrophic phase of

infection and possesses chorismate mutase activity. Cmu1 inhibits production of the defense signalling hormone, salicylic acid, by catalysing conversion of chorismate to prephenate. This shifts the shikimate pathway towards production of aromatic amino acids, rather than salicylic acid [102]. Maize plants infected with *Δcmu1* mutants therefore exhibited high levels of salicylic acid accumulation [102].

Another *U. maydis* effector, Tin2, has been shown to inhibit production of lignin by reducing production of p-coumaric acid, a precursor in the lignin biosynthetic pathway. This is through binding and stabilising ZmTTK1 in the host cytosol to promote anthocyanin biosynthesis and negatively affect p-coumaric acid levels [89, 103]. Effectors can also target the host ubiquitination system, which is important in both positive and negative regulation of the plant immune system. In *P. infestans* for example, AVR3a, is characterised as having the capability to bind and stabilize CMPGI, a U-box E3 ubiquitin ligase in potato [104]. This eventually blocks Inf1-induced cell death, thereby promoting biotrophic growth.

A *M. oryzae* avirulence effector AVR-Pizt has been shown to act in a similar way, by suppressing chitin-induced PAMP-triggered immunity through binding and destabilizing APIP6, a rice RING E3 ubiquitin ligase [105]. Recently however, it has been reported that AVR-Pizt interacts with a cell death suppressor APIP5 in rice cells lacking Piz-t. APIP5 is a bZIP-type transcriptional factor and inhibiting its transcriptional activity leads to effector-triggered necrosis required at the necrotic stage. Piz-t binds and stabilizes APIP5 and inhibits effector-triggered necrosis [106]. This effector appears to target two very

distinct host proteins to suppress immunity. Another example of effector-host interaction is that of AVR-Pii and the host protein, OsExo70-F3 during Pii-dependent resistance [107]. Study of this interaction has increased the growing evidence of integrated decoy/helpers' involvement in AVR/R gene interactions. Recently, Singh *et al* [108] used a Y2H screen to identify another putative AVR-Pii interactor, a rice NADP-malic enzyme (Os-NADP-ME2). Os-NADP-ME2 serves an important function in rice innate immunity against the rice blast fungus by inhibiting its enzymatic activity, and the fungus thereby suppresses the ROS burst [108].

The *M. oryzae* apoplastic secreted LysM effector Slp1 acts by preventing rice from recognising chitin oligomers released by the fungus. Slp1 possesses very high chitin binding affinity and will scavenge for, and bind to, chitin fragments to prevent recognition by the host chitin receptor CEBiP [94]. Necrosis and ethylene-inducing peptide (Nep) effectors have been characterised in necrotic fungal pathogens. In the hemi-biotrophic *C. higginsianum* for example, effector expression has been shown to occur in four different waves during the infection stage or pathogenesis: 1. pre-penetration into the appressorium, 2. post-appressorium – by initial invasive hyphae, 3. at the biotrophic-necrotic switch and 4. during late necrotrophy. Some of the early secreted effectors are involved in maintaining or promoting cell viability and late secreted effectors in promoting cell death [109].

The *C. higginsianum* Nep-like effector protein, ChNLP1, is highly expressed during the switch from biotrophic colonisation to necrosis and can induce cell death when transiently expressed in a non-host *N. benthamiana*

[109]. By contrast, Mogga *et al* [110] characterised two *M. oryzae* effectors that can inhibit necrosis induced by a Nep1-like effector when transiently expressed in *N. benthamiana* [110]. Effectors that have a protease activity are common in fungal pathogens and oomycetes. This includes Pit2 in *U. maydis* [111], AVR2 in *C. fulvum*, and two well characterised *P. infestans* effectors EPIC2B and EPIC1 that accumulate in the tomato apoplast and inhibit Rcr3 protease. EPIC1 can also inhibit two other proteases closely related to Rcr3 [112-114].

Recent studies have identified effectors that serve as suppressors of plant cell death (SPD) in *M. oryzae* [115, 116]. These effectors include SPD3 (also named MoHEG13/Bas52), SPD5 a paralogue of Bas4, SPD6 (also named Bas3) Iug6 and Iug9 [110, 115, 116]. These effectors inhibit host cell death induced by necrosis and ethylene-inducing-protein-1(Nep1) and mammalian BAX-mediated cell death in *N. benthamiana* [110, 115, 116]. However, little is known about the functions of most *M. oryzae* effectors. Pwl2 for example, is an avirulence protein that prevents infection of weeping lovegrass (*Eragrostis curvula*) but the function served when it is secreted into a rice cell is largely unknown [117, 118]. However, Pwl2 has been shown to diffuse into the neighbouring cells before the invasive hyphae. Among biotrophy-associated proteins, Bas-1 has been characterised as a cytoplasmic effector and like Pwl2 moves into the neighbouring cells to serve an unknown function. Bas-2 and Bas-3 are cytoplasmic effectors that are secreted into the host cell and accumulate at the point where the invasive hyphae cross to the neighbouring cell [1, 2]. These effectors may be involved in the opening of plasmodesmata [1]. Bas-4 accumulates into host apoplastic space but serves an unknown function during biotrophic growth [1, 2].

Table 1.1 List of known *M. oryzae* avirulence/effector genes.

Avirulence/effector	ID number	Function/localisation	R-gene
AVR-Pita	AF207841	Zinc metalloprotease	*Pita*
AVR-Piz-t	HE578813	Target rice U3 ubiquitin ligase	*Piz-t*
AVR-CO39	AF463528	Interacts with RGA4/RGA5	*CO39*
AVR-Pia	AB498873	Interacts with RGA4/RGA5	*Pia*
AVR-Pii	AB498874	Interacts with Exo70 or NADP-malic enzyme	*Pii*
AVR-Pik/km/kp	AB498875-79	In planta expressed/cytoplasmic	*Pik*
AVR-Pi9	MGG_12655	In planta expressed/cytoplasmic	*Pi9*
AVR-Pib	KM887844	BIC/Cytoplasmic	*Pib*
PWL2	MGG_04301/MGG_13683	In planta expressed/BIC	-
ACE1	AJ704622	Polyketide synthase	*Pi33*
AVR-Pi54	MGG_01947	In planta expressed/appressorium	*Pi54*
SLP1	MGG_10097	LysM domains, suppress PTI	-
MC69	MGG_02848	In planta/apoplast	-
MSP1	MGG_05344	Cerato-platanin family	-
BAS1	MGG_04795	In planta expressed/BIC	-
BAS2	MGG_09693	cell wall crossing points	-
BAS3	MGG_11610	suppressor of plant cell death	-
BAS4	MGG_10914	Apoplastic	-
BAS107	MGG_10020	BIC/Nucleus	-

1.8 Pathogen-host interaction

Pathogenic fungi and host plants are known to interact in an attack and counter-attack arms race-like evolutionary dynamic that has imposed a strong interactive selection and shaped the genomes of both host and pathogen [119-121]. In this type of interaction model, both the pathogen and host are in continuous development cycles that result in new effectors and host target R gene alleles being temporarily or permanently fixed in the population [121, 122]. Evolution of host R-genes helps the plant select for an incompatible (resistance) interaction [85]. The principle behind this selective co-evolution is based on a gene-for-gene interaction model in which the plant encodes for R proteins to detect pathogen secreted AVRs, leading to an incompatible reaction and resistance [85, 123]. On the other hand, absence of cognate AVRs will result in a compatible interaction [46, 124]. Moreover, allelic variation of a cognate AVR resulting from selection pressure imposed by the plant pathogen-interaction consequently leads to plant susceptibility [46, 121, 124]. A subset of AVRs will be selected for when cognate R genes are rare in the population, while the R genes will be selected for when a subset of AVRs are common in a population [46, 124, 125].

The evolution of effectors is thought to be a balance between escape from detection and maximisation of virulence. Continuous emergence of effector proteins to substitute those lost during the co-evolution will determine the fitness and survival of a fungal species in a population [89]. Several fungal genomics studies have indicated a higher selection pressure on effector proteins as compared to other non-secreted proteins. Cases of presence/absence

polymorphism are more common in effectors than other genes as demonstrated in barley powdery mildew pathogen *Blumeria graminis*, leaf rust pathogen *Melampsora larici-populina* and smut fungus *Melanopsichium pennsylvanicum* [126-128]. The gain and loss of effectors in fungi has been associated with effector genes being situated in flexible genomic regions resulting in rapid effector gene evolution [38, 128, 129]. Moreover, cases of host jump have caused a change in effector repertoire promoting enhanced diversification [48][128]. This means that the effector repertoire of pathogenic fungi is likely to be determined by the host.

Selection polymorphism in *AVR-Pik* was reported by Yoshida *et al* [46] through the cloning of three alleles *AVR-Pik/km/kp*. The rice resistance protein Pik has also been shown to have multiple encoding alleles in the form of *Pik*, *Pikm, Pikp, Piks* and *Pikh* which have varied recognition to the AVR-Pik alleles [85]. A study to understand variation on the NLR pair *Pik* locus (*Pik-1 and Pik-2*) revealed extensive polymorphism in the coiled-coil region of *Pik-1* with low polymorphism observed in the *Pik-2* region. The coiled region possesses an integrated HMA domain that binds effectors for recognition. AVR-Pik exhibits polymorphism at amino acid 46-48 and 67, while polymorphism in *Pik-1* is exhibited in the integrated HMA domain [74]. These variations can be attributed to the co-evolution of both the pathogen and host, resulting from direct protein-protein interactions at the molecular level [74].

Some effectors such as *PWL2* belong to a family of more than one gene which tend to diverge from a common ancestor [117]. The expansion or loss of members in these gene families may either be specific to one species, or

spread across the fungal kingdom [117, 130]. Gene family expansion and diversification can lead to host specification, as observed in the Irish potato famine pathogen *Phytophthora infestans* and its sister species *Phytophthora mirabilis,* responsible for infecting *Mirabilis jalapa* (four o'clock flower). A study by Dong *et al* [131] characterised diversification occurring in a cystatin-like effector EPIC1 that inhibits the activity of a cysteine protease in their respective hosts. Amino acid changes in this effector led to its specificity towards the host target protein, a process thought to have facilitated a host jump [131]. In these interactions, both the pathogen and the host appear to be selected to evade recognition by each other in a biological battle [99, 119]. The host recognises the pathogen invasion before initiating an effective immune response while on the other hand; the pathogen must effectively regulate its arsenal of effector proteins for efficient colonisation [130]. Some fungal genes may be down-regulated to avoid recognition, while some genes responsible for subverting the host metabolism, for detoxifying defence compounds and suppressing immune system will be up-regulated during colonisation [130].

1.9 Genomics for pathogen monitoring, surveillance and understanding the effector repertoire of *M. oryzae*

Phenotypic and genetic analysis of field isolates of *M. oryzae* from five different geographic regions in China revealed sequence variation within the *AVR-Pib* gene. Different levels of selection were identified; complete gene deletion, segmental deletion, point mutation and transposable elements (TEs) insertions. An analysis of isolates within one region in China revealed different genetic and phenotypic structures. Isolates from Liaoning were different from

those collected from Jilin and Heilongjiang, for instance, an observation related to host selection pressure rather than geography. The southern regions of China had more virulent pathotypes, an observation that was related to introduction of *Pib* into the region by IRRI in the early 1960s [124]. Transposable elements insertions were shown to be likely to account for loss of avirulence rather than gene deletions and point mutations. Wu *et al* 2015 [42] also reported the ability of field isolates to gain virulence towards *Pi9* through insertion of Mg-SINE within the *AVR-Pi9* coding sequence. This confirms the importance of transposable elements occurring in the *M. oryzae* genome in response to host selection pressure [42, 124].

AVR-Pita has been reported to be genetically unstable, with frequent occurrence of mutations that leads to gain-of-virulence on *Pi-ta* rice cultivars [132]. Cases of partial deletions, ranging from 100bp to 12.5 kb and point mutations have already been documented [125, 133]. The potential of deletion of AVR genes or mutations leading to a host jump remains a major concern. There are a small number of known AVRs for example that prevent *M. oryzae* pathotypes from infecting wheat. To date only strains from *Lolium* infecting isolates that cause the gray leaf spot in turf grasses, have been reported to infect wheat [19, 134].

Among the *Oryza, Avena* and *Setaria* pathotypes of *M. oryzae*, the presence of five effector-like genes *AVR-PWT*1-5 prevents infection on wheat [135]. Most recently, Inoue *et al* [20] cloned two avirulence genes *PWT3* and *PWT4* that control infection of *Magnaporthe oryzae* isolates towards *Triticum aestivum* (Wheat). An investigation on historical data regarding Brazil's wheat

cultivation led to a conclusion that deployment of cultivars lacking *Rtw3* resulted in susceptibility towards *Lolium* isolates [20]. In the late 1970s to early 1980s, IAC-5 (*Rtw3* carrier) was the most planted wheat cultivar in Brazil. Introduction of a high yielding cultivar called Anahuac that lacks *Rtw3* coincided with the first outbreak of wheat blast in 1985 [20]. This was later followed by loss of function of *PWT3* in more wheat infecting isolates [20].

An avirulence gene *AVR-CO39* prevents *Magnaporthe oryzae* infection towards rice lines that possess R-gene *CO39*. This gene has been deleted in rice infecting isolates through transposon insertions leading to gain-of-virulence on rice [135, 136]. However this gene is also present in other non-*Oryza* infecting pathotypes [135, 136]. Another host determinant gene *PWL2* was cloned from rice infecting isolates and it prevents isolates from infecting weeping love grass (*Eragrostis curvula*). This gene is absent in isolates infecting weeping love grass with some of them carrying the virulent allele of *pwl2* [118]. *Magnaporthe* host specificity is controlled by effector-host immunity interaction rather than other factors related to host physical difference. Loss of AVR genes could lead to host jump and disease outbreaks such as wheat blast.

1.10 Challenges in identification of effectors/AVRs

A specific amino acid motif associated with effector translocation in host cells has been characterised in oomycetes. This motif is referred to as the RXLR-dEER motif (X representing any amino acid) and situated on the N-terminus of the signal peptide [137, 138]. However in fungi, effectors are only classified as small secreted proteins and lack any known motifs associated with either their function or translocation into host cells [97]. Effectors are generally

secreted *in-planta* and gene expression levels are typically associated with a specific time during infection [139]. This means that expression of effector genes will be switched on only after the fungus is in contact with the host. Most often effectors are secreted through a conventional mechanism via the endoplasmic reticulum and Golgi-apparatus [2]. To undergo this process, effectors must therefore possess a signal peptide sequence at the N-terminus that facilitates its translocation. This forms the first criterion by which candidate effector genes can be identified using bioinformatics.

Fungal effectors have also been identified as small secreted proteins, ranging from 50-300 amino acids, although sometimes larger proteins can also act as effectors [39, 102, 140]. Some of these secreted proteins possess higher cysteine content and have stable tertiary structure with disulphide bridges [90]. This gives them the ability to resist the harsh physiological stress in a plant apoplast and provides another criterion for characterising apoplastic effectors. Finally, another characteristic used to identify effectors is the absence of protein orthologs outside the genus [95, 141]. For now, the definition of a fungal effector however remains ambiguous, which means any secreted, differentially expressed fungal protein, is classified as a potential effector.

1.11 *PWL2,* host range determinant gene

The ability of *M. oryzae* strains that can infect weeping lovegrass, to infect its host, was shown to be controlled by a single gene first identified in a *M. oryzae* laboratory strain, 4360 [118]. This strain is a genetic cross between two rice pathogenic laboratory strains. One of the parent strains, 4224-7-8 was able to infect weeping love grass and lacked *PWL2,* while the other strain,

6043 was non-pathogenic and possessed the *PWL2* locus [118]. In the genetic cross, each of the five tetrads had four ascospore progenies that were pathogenic on weeping lovegrass and four that were non-pathogenic. This suggested single-gene segregation of the ability to infect weeping lovegrass [118]. Spontaneous mutant strains in the study, lacking *PWL2*, were also able to infect weeping love grass, suggesting that *PWL2* determines pathogenicity of the *M. oryzae* strains to this host. When strains lacking *PWL2* were transformed with the gene, which was cloned by map-based cloning, the pathogenicity towards weeping love grass was lost, but the strain still retained pathogenicity towards barley and rice cultivars [118]. This indicates that *M. oryzae* strains did not have a defect in their ability to infect plants but were avirulent on weeping love grass due to the presence of *PWL2*.

PWL2 belongs to a family of genes that includes three additional putative effectors *PWL1*, *PWL3* and *PWL4*. *PWL1* has also been identified as being involved in the incompatible reaction of *M. oryzae* against weeping love grass (*Eragrostis curvula*) [118]. *PWL2* encodes for a protein containing 145 amino acids with a molecular weight of 16.17 kDa. An allele of *PWL2* obtained from one *M. oryzae* strain was unable to confer avirulence and was termed as a divergent *pwl2* allele [118]. A substitution of a guanine to adenine in *PWL2* allele causes an amino acid change from aspartic acid to asparagine at residue 90 [118]. The amino acid sequence in normal *PWL2* gene product, usually DKS, is altered to NKS, which is a putative signal-sequence for glycosylation [118]. Examination of different alleles of *PWL2* causing virulence to weeping love grass show that this gene is however, very polymorphic, depending on isolate and geographic origin. While field isolates with spontaneous *PWL2* deletions do

not show any known fitness issues, most of the field isolates possess one or two copies of the gene [118]. It is not clear what role, if any, *PWL2* plays in rice blast disease, although as it occurs in a high percentage of rice blast field isolates it seems likely that it has a function during plant infection.

1.12 Introduction to the current study

I set out to investigate effectors in the rice blast fungus *M. oryzae.* This work was part of an international project aimed at understanding the rice blast population biology in Sub-Sahara Africa, in order to provide a guide for a rice blast-resistance gene profile for rice cultivars in that region. The broader aspects of the project involve the University of Arkansas, Biosciences eastern and Central Africa, Kenya Agricultural and Livestock Research Organisation, University of Ohio, Africa Rice Centre and International Rice Research Institute. My project had two parts.

First I analysed a large-scale set of genome sequence data from a population of rice blast isolates from Sub-Saharan Africa. This was aimed at developing new methods for mining for potential avirulence/effectors and then functionally analyse them. Using bioinformatics and molecular tools, I aimed at understanding how the presence or absence of putative secreted proteins is associated with virulence of some of these isolates on near-isogenic rice lines. To predict secreted proteins, open reading frames in genome sequence unmapped reads can be considered as possible genes followed by identification of signal peptides to identify putative secreted proteins [46]. However, with this method, there is a possibility of missing out on some candidate genes that maybe lost during the re-assembly of unmapped genes, or predicting many

fragmented genes during gene calling. I used a different approach to identify potential AVR/effectors. We further sequenced two *M. oryzae* rice pathogenic isolates, Guy11 and KE002, using Pacbio to improve the quality of their genome assemblies. Guy11 is a highly aggressive strain, while KE002 is a Kenyan strain that is avirulent on selected rice monogenic lines, and is thought to carry many avirulence genes. Moreover, I also used RNA-seq analysis of KE002 infection on a susceptible rice cultivar *Moukoto* to identify secreted protein-encoding genes that are differentially regulated during plant infection. This analysis has identified several putative effectors and three novel effector genes and this thesis will focus on attempts functionally characterise them.

Secondly, I carried out a fundamental study of a host range determinant gene, *PWL2.* The aim was to determine the biological function and to test its role in rice blast disease as well as the ability to infect weeping lovegrass. Using bioinformatics tools, I observed that more than 95% of the 80 analysed rice pathogenic strains carry at least one copy of *PWL2*. In the rice-infecting isolates, I found that the occurrence of the allelic version of *pwl2* was frequent especially in isolates from Asia. On contrary, 23 *M. oryzae* sequenced genomes from Sub-Saharan Africa all possessed *PWL2.* From this observation, I reasoned that *PWL2* might potentially be valuable as an avirulence gene, and if the cognate R-gene could be identified it can be introgressed into rice to achieve durable resistance. I report the mechanism by which this effector is regulated, expressed and translocated into rice cells. I also report on use of CRISPR/Cas9 gene editing approach to functionally characterise genes with multiple copies in *M. oryzae* like *PWL2.*

In chapter 3 and 4 I investigate

1. **Different methods of identifying and cloning effector proteins**
2. **Role of newly identified effectors as virulence determinant genes**

In chapter 5 I investigate

1. **The fitness cost related to targeted deletion of *PWL2***
2. **The role played by Pwl2 during biotrophic growth of *M. oryzae***

Chapter 2 Materials and methods

2.1 Fungal growth, maintenance and storage

M. oryzae isolates used in this study were collected from 9 different countries in Sub-Saharan Africa (Benin, Burkina Faso, Mali, Ghana, Kenya, Nigeria, Tanzania, Togo, and Uganda) and stored in the laboratory of Prof. N.J. Talbot (University of Exeter). A subset of the isolates was provided by Dr. Didier Thareau (Centre de Coopération Internationale en Recherche Agronomique pour le Développement, CIRAD, Montpellier, France). For additional studies, the African collection was compared to *M. oryzae* isolates from different parts of the world, including China, Egypt, French Guyana, Philippines, USA, Thailand and South Korea. For long term storage, fungal isolates and strains were grown on filter paper discs (3mm, Whatman International), desiccated and stored at -20 °C. Fungal isolates were routinely grown on complete medium (CM) at 24°C with a controlled 12 h light and dark cycle for up to 12 days [53]. CM contains 10g glucose, 2 g peptone, 1 g yeast extract (BD biosciences), 1 g casamino acids, 0.1 (v/v) trace elements (22 mg zinc sulphate heptahydrate, 11 mg boric acid, 5mg manganese (II) chloride tetrahydate, 5 mg iron (II) sulphate heptahydrate, 1.7 mg cobalt (II) Chloride hexahydrate, 1.6 mg copper (II) Sulphate pentahydrate, Sodium molybdate dehydrate, 50 mg ethylenediaminetraacetic acid), 0.1% (v/v) vitamin supplement (0.001 g biotin, 0.001 g pyridoxine, 0.001 g thiamine, 0.001 g riboflavin and 0.001 nicotinic acid in 1 L), 6g $NaNO_3$, 0.5 g KCl, 0.5 g $MgSO_4$, 1.5g KH_2PO_4 in 1 L, [adjusted to pH 6.5 with NaOH] and 15 g agar for 1 L. For liquid CM, the agar was not

added. All chemicals were supplied by Sigma (Poole, Dorset UK) unless stated otherwise.

2.2 Description of rice germplasms

Differential rice cultivars used in this study are genotypes that carry 24 known rice blast resistance genes. All genotypes were generated from a susceptible japonica cultivar Lijiangxituanheigu (LTH) [142, 143]. Seeds for these genotypes were generously provided by Dr. Bo Zhou from the International Rice Research Institute (IRRI). The *Pi9* donor called 75-1-127 was obtained from Dr. Guo-liang Wang at Ohio State University. African rice *Oryza glaberrima* (cultivar AR105), F6-36 and New Rice FOR Africa (NERICA) lines, were generously provided by Dr. Ibrahima Ouédraogo at Station de Recherches de Farako-Ba, Burkina Faso, Africa.

2.3 Pathogenicity and infection assays

2.3.1 Virulence analysis of fungal strains on rice

Conidia from 8-12 day old cultures grown on CM agar were removed from a Petri dish culture using a sterile disposable plastic spreader in 3 mL sterile distilled water. The conidial suspension was filtered through sterile Miracloth (Calbiochem) and centrifuged at 5000 *x g* (Beckman, JA-17) for 5 min at room temperature. The pellet of conidia was re-suspended in 0.2 % (w/v) gelatin (BDH) and the spore concentration determined using a haemocytometer (Improved NEUBAUER, Hawksley, UK). Spores were diluted to a final concentration of 5 $x10^4$ conidia mL^{-1}. The spore suspension was used for spray inoculation using an artist's air-brush (Badger. USA) or for leaf drop infections

and leaf sheath assays. Rice plants were grown for 21 days in 9 cm diameter plastic plant pots or seed trays. After spray inoculation, the plants were covered in polythene bags and incubated in a controlled plant growth chamber (REFTECH, Holland) at 24°C for 48 h with a 12 h light and dark cycle, and 85% relative humidity, before removing the polythene bags. The inoculated plants were incubated for 4 more days before scoring the lesions [144].

2.3.2 Assay for studying fungal invasive hypha proliferation in the host cell.

To observe the intracellular growth of fungal invasive hyphae, rice cultivar *Moukoto* leaf sheaths were inoculated with 4 mL of a suspension at 5 x 10^4 of conidia mL^{-1} in 0.2% (w/v) gelatin using a syringe as described in Kankanala et al 2007 [145]. The inoculated leaf sheaths were incubated at 24°C for at least 27 h before dissecting a thin layer of the inner leaf sheath using a blade and mounted on a glass slide for microscopy.

2.3.3 Conidial germination and appressorium formation assay

The appressorium assay was adapted from Hamer et al 1988 [26]. Conidial suspensions were prepared, as described earlier in Section 2.3.1. A 50 µL aliquot of conidial suspension was inoculated onto a borosilicate glass coverslip (Fisher Scientific UK Ltd.) and placed on a moist paper towel. This was incubated at 24°C for a minimum of 8 h before observing by epifluorescence microscope as described in Section 2.4.

2.4 Microscopy

An IX81 motorized inverted microscope (Olympus, Hamburg, Germany) was used to perform conventional and differential interference contrast (DIC) microscopy. To capture images from the microscope, a photometrics coolSNAP HQ camera system (MDS Analytical Technologies, Winnersh, Uk was employed. The microscopic epifluorescence settings were as follows, GFP (excitation 480nm, emission 510); RFP/mCherry (excitation 561nm, emission of 58). A Leica TCS SP8 laser confocal microscope was used for laser scanning fluorescence at X 40 with oil immersion objective lens. The lasers were set as follows: GFP and RFP tagged proteins were excited using 488 and 561 nm laser diodes and the emitted fluorescence detected using 495-550 and 570-620 nm respectively. The auto-fluorescence from chlorophyll was detected at 650-740 nm.

2.5 DNA manipulation

2.5.1 DNA preparation

2.5.1.1 CTAB DNA extraction

Fungal mycelium was generated by growing fungal culture on either cellophane discs or liquid culture for large scale extraction. Routinely, 7-12 days old mycelium was ground into powder using a mortar and pestle, and 500 µL of pre-warmed CTAB (2% (w/v) Hexadecyltrimethylammonium Bromide (CTAB), 100 mM Trisma base, 10 mM Ethylenediaminetraacetic acid (EDTA) and 0.7 M NaCl) added into a 1.5 mL microcentrifuge tube containing ground mycelial powder and incubated at 65°C with gentle mixing every 10 min. An equal vol of

chloroform iso-amyl alcohol (CIA) was added, mixed thoroughly and incubated with shaking for 30 min at room temperature. This was followed by centrifugation at 17000 *x g* for 10 min. This step was repeated twice by adding equal vols of CIA and mixing vigorously on a shaker before centrifugation. The final supernatant was transferred into a clean sterile microcentrifuge tube and of isopropanol (2 x vol) added before incubating at -20°C overnight. The samples were centrifuged at 17000 *x g* for 10 min and the supernatant (isopropanol) was gently pipetted off and the resulting pellet re-suspended in 500 µL sterile distilled water (SDW), and left to dissolve at room temperature with gentle tapping to mix. Sodium acetate (NaOAc) (0.1 vol) and of 100% ethanol (2 vol) were added to re-precipitate nucleic acids. The mixture was incubated at -20°C for 2 h and centrifuged at maximum speed, before washing with 400 µL of 70% (v/v) ethanol. The DNA was re-suspended in nuclease-free water. RNase (2 µL) was added and incubated at 37°C for 1 h to digest contaminating RNA.

2.5.1.2 High molecular weight (HMW) DNA isolation for Pacbio sequencing

2.5.1.2.1 QIAGEN DNeasy Plant Mini Kit DNA prep for Pacbio sequencing

To obtain high molecular weight DNA free from RNA, carbohydrates and protein contamination, a commercial kit (QIAGEN DNeasy Plant Mini Kit) was used, according to the manufacturer's instructions. Freshly isolated *M. oryzae* protoplasts were used as starting material. This step was adapted to reduce contamination from polysaccharides that are abundant in fungal cell walls. A 400 µL aliquot of buffer AP1 (lysis buffer) and RNaseA (100mg/ml) were added to pelleted protoplasts (5 X 10^8 mL^{-1}) in a 1.5mL microcentrifuge tube. This was thoroughly mixed by gently tapping the microcentrifuge tube or gently pipetting

to remove clumps in the mixture, and incubated at 65°C for 10 min. To precipitate detergent, polysaccharides and proteins, 130 µL of buffer P3 was added to the mixture and incubated on ice for 5 min. The lysate was centrifuged for 5 min at 20,000 *x g* and the supernatant transferred onto the QIAshredder mini spin column placed into a 2 mL collection tube and centrifuged for 2 min at 20,000 *x g*. Without disturbing the pellet, the flow through was transferred into a new sterile tube and 1.5 volumes of buffer AW1 (containing guanidine hydrochloride with ethanol) added. The mixture was transferred onto a DNeasy column, placed into a 2 mL collection tube and centrifuged at 6000 *x g* for 1 min. The flow through was discarded and the step repeated for the remaining sample. The DNeasy column was transferred into a new collection tube and 500 µL of buffer AW2 (containing 70% (v/v) ethanol) was added. This was centrifuged for 1 min at 6000 *x g* and the step repeated but centrifuged at 20,000 *x g* for 2 min to completely remove any ethanol residues. The column was then transferred into a clean collection microcentrifuge tube and 100 µL of TE buffer pH 8 added directly onto the membrane. This was incubated at room temperature for 5 min and centrifuged for 1 min at 6000 *x g*.

2.5.1.2.2 DNA clean and up and concentration

To further purify high molecular weight DNA from any inhibitors, a commercial kit (Zymo Research DNA Clean & Concentrator™ -5), was used according to the manufacturer's instructions. In a 1.5 mL microcentrifuge tube, 2 vols of DNA-binding buffer were added to a volume of DNA sample (2:1) and mixed gently. The mixture was transferred onto a Zymo-Spin™ column in a 2 mL collection tube, centrifuged for 30 sec and the flow-through discarded. The

column was washed twice by adding 200 µL of DNA wash buffer onto the column and centrifuging for 30 sec. The column was then transferred onto a new sterile micro-centrifuge tube. A 50-100 µL aliquot of TE buffer pH 8 was added directly onto the column and incubated at room temperature for 1 min. DNA was eluted by centrifuging for 30 sec at high speed. The sample was submitted for DNA sequencing.

2.5.1.3 High molecular weight (HMW) DNA isolation for optical mapping

2.5.1.3.1 Plug lysis DNA isolation for optical mapping

To extract high molecular weight DNA-free from contaminants, *M. oryzae* protoplasts were subjected to gentle lysis. Fungal protoplasts were isolated by incubating mycelium obtained from liquid culture, in OM buffer (1.2 M magnesium sulfate, 10 mM sodium sulfate, 5% (w/v) Glucanex (Novo Industries, Copenhagen), pH 5.8), as explained in Section 2.6. To increase the number of protoplasts recovered, the enzyme suspension was passed through four layers of sterile Miracloth to separate protoplasts from mycelium. The protoplasts were then transferred into a sterile falcon tube and washed twice using STC (1.2 M sorbitol, 10 mM Tris-HCl [pH 7.5]), 10 mM calcium chloride). The cells were then re-suspended to a concentration of 5×10^8 mL^{-1}.

2.5.1.3.2 Making agarose plugs

2% (w/v) low melting point agarose (BioRAD) was made in STC (1.2 M sorbitol, 10 mM Tris-HCl pH 7.5, 10 mM calcium chloride). Molten agarose was mixed with protoplasts to achieve 0.75% (w/v) agarose final concentration. The mixture incubated in a water bath at 50°C and gently mixed by pipetting. A plug

former was assembled and 85 µL of the molten mixture loaded into each well and incubated at 4°C for 15 min until the plugs solidified. A plug remover was used to transfer plugs into a 50 mL falcon tube and 2.5 mL Proteinase K digestion buffer (1% (w/v) N-Lauroyl sarcosine, 0.2% (w/v) Sodium Deoxycholate, 100mM EDTA, Proteinase K, 2mg/mL final concentration added and incubated at 50°C overnight. The plugs were washed three times in 1 x wash buffer (10 x wash buffer – 100mM Tris, 0.5M EDTA pH 8.0) and store at 4°C until use.

2.5.2 Restriction enzyme digestion of genomic or plasmid DNA

Restriction endonucleases used in this study were obtained from either Promega. (Southampton, UK) or from New England Biolabs (Hitchin, UK). DNA digestion reaction mix composed of 1-15 µg DNA, 5-10 units of enzyme, 5 µL of manufacturer supplied buffer with a final volume of 50 µL nuclease-free water. The mixture was incubated at 37°C overnight for genomic DNA and at least 4 h for plasmid DNA. Agarose gel electrophoresis was used to fractionate the digested DNA fragments.

2.5.3 Polymerase chain reaction (PCR)

The reaction was carried out using Applied Biosystems GeneAmp ® PCR system 2400 thermo cycler following manufacturer's instructions. Routinely, GoTaq ® Green Master Mix and 50-100 ng of DNA template was used and each reaction set to 40 µL final vol. The PCR reaction included an initial denaturation step at 94ºC for 2 min and 35 cycles of PCR cycling parameter: denaturation at 94ºC for 30 sec, annealing at 55-62ºC for 30 sec and extension

at 72ºC for 1min/1kb, followed by a final extension at 72ºC for 10 min. To obtain a high fidelity DNA amplicon, Phusion DNA polymerase (New England Biolabs, Thermo Scientific) was used. The reaction mix included 1 unit of Phusion polymerase enzyme, 10 µM of 5 X Phusion buffer, 200 µM dNTPs, 0.5 µM of each primer and 50-100 ng of template DNA. PCR conditions were as follows, initial denaturation at 98°C for 30 sec, and 35 cycles of PCR cycling parameter: denaturation at 98°C for 10 sec, annealing at 58°C for 30 sec and extension at 72°C for 30 sec/1kb for the desired fragment length. For colony PCR screening, SapphireAmp Fast PCR Master Mix (Clontech, USA) was used as per manufacturer's instructions. The reaction mix included 25 µL of SapphireAmp Fast PCR Mix, 0.2 µM (final conc.) of forward and reverse primers and adjusted with nuclease free water to a final volume of 50 µL. PCR conditions were as follows; denaturation at 94°C for 1 min and 30 cycles of PCR cycling parameter: denaturation at 98°C for 5 sec, annealing at 58°C for 5 sec and extension at 72°C for 10 sec/1kb for the desired fragment length.

2.5.4 DNA gel electrophoresis

PCR amplification products and restriction enzyme digestion products were fractionated by gel electrophoresis through a 0.8% (w/v) agarose gel in 1X Tris-borate EDTA (TBE) buffer (0.09 M Tris-borate and 2 mM EDTA). Ethidium bromide was added to molten agarose gel to 0.5 µg/mL final concentration to enable DNA visualisation under UV-light. To estimate the size of DNA fragments in the agarose gel, 1 Kb plus size marker (Invitrogen) was loaded alongside the samples. The separated DNA fragments in the gel were visualised and recorded using a UV transilluminator and gel documentation

system (Image Master VDS with a Fujifilm Thermal Imaging system FTI-500, Pharmacia Biotech).

2.5.5 DNA fragments gel purification

A commercial kit (Wizard Plus SV Gel and PCR Clean-up System, Southampton, UK) was used to purify fractionated DNA from agarose gels as per the manufacturer's instructions. Agarose containing the desired size of DNA fragment was cut using a sterile razor blade and weighed in a sterile 2 mL microcentrifuge tube. Membrane binding solution (4.5 M Guanidine isothiocyanate and 0.5 M Potassium acetate, pH 5.0) was added to the tube containing the cut gel at the ratio of 10 µL per 10mg cut gel. The sample was incubated in a 65°C water bath until the gel was completely dissolved in membrane binding solution. An 800 µL aliquot of the DNA in molten agarose and membrane binding solution was transferred onto the Wizard® SV Minicolumn placed on a supplied 2 mL collection tube and incubated at room temperature for 1 min. The sample was centrifuged at 13,000 *x g* for 1 min to allow dissolved DNA to bind onto the Wizard® SV Minicolumn and the flow through waste to be discarded. 700 µL of membrane wash buffer (with 100% ethanol added) was added straight onto the column and centrifuged for 1 min. This step was repeated by adding 500 µL of membrane wash buffer and centrifugation for 5 min. The column bound DNA was eluted by directly pipetting 25-50 µL of Nuclease-Free water onto the column and incubating at room temperature for 1 min before centrifuging at 13,000 *x g* for 1 min. DNA solution was stored at -20°C for long term storage.

2.5.6 DNA cloning and transformation of bacterial hosts

2.5.6.1 In-Fusion Cloning

For precise cloning of one or multiple DNA fragments, a homologous recombination technique involving an In-Fusion HD Cloning Kit (Clontech, USA) was used. Primers were designed to introduce a 15bp extension that overlapped with sequences at the restriction sites of the destination vector or adjacent insert fragments allowing the ends to fuse by homologous recombination during cloning. The reaction was prepared as follows: 2 µL of 5 X In-Fusion HD Enzyme Premix, 10-150ng of purified PCR fragments and 50-100ng linearized vector. Deionised water was used to adjust the final volume to 10 µL. The reaction mixture was incubated at 50°C for 15 min then placed on ice before proceeding on with bacterial cell transformation. A 2.5 µL aliquot of the reaction mix was added to thawed 50 µL of Stellar^{TM} competent cells, mixed gently and incubated on ice for 30 min. The cells were subjected to 45 sec of heat-shock at 42°C followed by 2 min incubation on ice. A 450 µL aliquot of pre-warmed SOC media (Clontech, USA) was added to the transformed cells and incubated shaking at 37°C for 1 h. After 1 h, 100 – 150 µL of the bacterial culture was plated on Lysogeny (LB) medium containing the appropriate antibiotic. This was incubated overnight at 37°C. To determine successful transformants, colony PCR and restriction enzyme digests were used.

2.5.6.2 Gateway cloning

To test sub-cellular localisation of effector proteins in plant cells, transient expression in model plant *Nicotiana benthamiana* was carried out. Two steps of Gateway cloning technology were used to acquire the destination vectors for *Agrobacterium* transformation. Gateway cloning technology takes advantage of the site-specific recombination properties of bacteriophage lambda. The gene of interest is flanked by attB1 and attB2 sequences that recombine with flanking regions attP1 and attP2 in the entry vector to replace a ccdB gene that is toxic to *E. coli* cells. This reaction is mediated by BP clonase enzyme. This is followed by the LR reaction. The gene of interest now in the entry vector flanked by attP1 and attP2 sequences recombines with flanking regions attL1 and attL2 in destination vector to replace a ccdB. This reaction is mediated by the LR clonase enzyme [146].

BP Reaction

This procedure was carried out at room temperature according to manufacturer's instructions. A 1 µL aliquot attB-PCR product (final concentration 30-300ng), 1 µL of Donor vector (150 ng/ mL), 2 µL of 5 X BP clonase reaction enzyme (Invitrogen) and appropriate volume of TE buffer, pH 8.0 were added in a 1.5 mL Eppendorf tube and mixed by gentle tapping. The mixture was incubated at room temperature for 1 h then 2 µL of proteinase K was added and incubated at 37°C to stop the reaction at which point 1 mL of BP reaction product was used to transform 50 µL of top 10 competent cells. The cells were incubated on ice for 30 min before being subjected to heat shock at 42°C for 30 sec. A 450 µL aliquot of SOC media was added and the mixture

incubated by shaking at 37°C for 1 h at which point 100 - 200 µL of the transformation was plated on LB medium containing appropriate selective antibiotic.

LR reaction

This procedure was carried out at room temperature according to the manufacturer's instructions. Aliquots of 1 µL entry clone (final concentration 100-150ng), 1 µL of Destination vector (150 ng/ µL), 2 µL of LR clonase reaction enzyme (Invitrogen) and appropriate volume of TE buffer, pH 8.0 were added in a 1.5 mL Eppendorf tube and mixed by gentle tapping. The mixture was incubated at room temperature for 1 h. Proteinase, 2 µL, was added and incubated at 37°C to stop the reaction, after which 1 µL of the reaction was used to transform 50 µL of top 10 competent cells. The cells were incubated on ice for 30 min before being subjected to heat shock at 42°C for 30 sec. Finally, 450 µL of SOC media was added and the mixture incubated shaking at 37°C for 1 h, after which 100-200 µL of the transformation was plated on LB medium containing the appropriate selective antibiotic.

2.5.7 Medium-scale plasmid DNA preparation

PureYield™ Plasmid Midiprep System (Promega, Madison, Wisconsin, USA) was used to recover high quality plasmid DNA for sequencing prior to fungal, yeast or *Agrobacterium* transformation. A single positive bacterial colony was inoculated in 50 mL liquid LB media containing the appropriate antibiotic and incubated shaking at 200 rpm overnight in an Innova 4000 rotary incubator (New Brunswick Scientific) set at 37°C. The bacterial liquid culture was

transferred to an Oakridge tube and fractioned by centrifugation at 5000 x *g* for 10 min and the supernatant discarded. Cell re-suspension buffer (3 mL) was added and mixed by pipetting to completely dissolve the pellet, after which 3 mL of cell lysis solution was added and mixed by inversion. The mixture was incubated at room temperature for 3 min before adding 5 mL of Neutralization solution. The mixture was centrifuged at 15,000 *x g* for 15 min to pellet cell debris. The supernatant was then passed through a column stack mounted as follows: PureYieldTM Clearing Column (blue) was nested on top of a PureYieldTM Binding Column (white), which was placed onto a vacuum manifold. Vacuum was then applied to allow the lysate to pass through the column. This process allows cell debris to bind to the clearing column and plasmid DNA to bind to the silica membrane in the binding column. The binding column was washed by passing through 5 mL of Endotoxin removal solution, followed by 20 mL of Column Wash Solution by use of vacuum. The membrane was dried by applying vacuum through the binding column for 30 sec to 1 min. To elute the plasmid DNA, the binding column was placed on 50 mL collection falcon tube and 400-600 µL of nuclease free water added directly onto the column and incubated at room temperature for 1 min. This was centrifuged for 5 min at 1500 – 2000 *x g* in a swinging bucket rotor. The plasmid DNA was be stored at -20°C for long-term storage. The plasmids were sequenced to confirm all inserts were correct.

2.5.8 Southern blotting

This procedure was adapted from the protocol of Southern [147] with adjustments as described by SambroOk *et al* [148]. Fungal genomic DNA (15 µg) was digested with the appropriate restriction enzyme overnight at 37°C. The digested genomic DNA was then separated by gel electrophoresis. The DNA embedded in the agarose gel was de-purinated by immersion in 0.25 M HCl with gentle shaking at room temperature for 15 min. The gel was then transferred into neutralisation solution (0.4 M NaOH) and incubated with shaking at room temperature for 15 min. The gel was then placed onto Whatman® 3 mm paper sheet supported with a perspex sheet with two ends of the paper submerged in 0.4 M NaOH allowing transfer of DNA onto Hybond-NX membrane (Amersham Biosciences) placed on top of the gel, by capillary method. For these 5 layers of wet Whatman® papers, 5 layers of dry Whatmann papers, and a stack of paper towels were added on top of the membrane. Extra weight was placed on the stack and incubated overnight at room temperature. After dismantling the blot, the membrane was exposed to UV light in a BLX-254 cross linker (Bio-Link®) to fix and immobilise the DNA. For Radioactivity blotting, after de-purination, the agarose gel was incubated in denaturing solution (0.4 N NaOH, 0.6 M NaCl) for 30 min followed by 30 min of neutralisation solution (1.5 M Nacl, 0.5 M Tris-HCl (pH 7.5)). The gel was placed onto Whatman® 3 mm paper sheet supported with a perspex sheet with each end of the paper submerged in 20 X SSPE solution (3.6 M NaCl, 0.2 M NaH2PO4H2O, 0.02 M EDTA) to allow transfer of DNA onto Hybond-NX membrane (Amersham Biosciences) placed on top of the gel, by capillary method.

2.5.8.1 DIG Southern blotting

For non-radioactive probing, DIG-labelled hybridisation probes were used to detect specific sequences on Hybond-NX membrane. To make a probe, DNA fragment of interest was PCR- amplified using Phusion DNA polymerase. The reaction included the standard PCR reagents except dNTPs mix, which was substituted by a DIG DNA labeling mix (Roche, UK). The PCR product was verified by agarose gel electrophoresis and the DNA purified using the described Wizard Plus SV Gel and PCR clean up kit protocol.

2.5.8.2 Radioactive probe preparation

The DNA labelled hybridisation probe was prepared using a ready-to-go kit (Amersham Biosciences, UK) according to manufacturer's instruction. The random primer method [149] was used. For this, 150 ng of DNA was diluted in water to a final volume of 47 µL. This was added to a tube containing ready-to-use reaction mix beads (buffer, dATP, dGTP, dTTP, FLPCpure^{TM} Klenow fragment and random oligodeoxyribonucleotides). The mixture was boiled for 5 min at 100°C to denature the DNA, immediately cooled on ice for 2 min and briefly centrifuged. Magenta polymerase, 2 µL, was added and mixed by pipetting gently, after which 2 µL of (α-^{32} P) dCTP (3,000 Ci/mmol) was added and incubated at 37°C for 10 min. The labelling dye (0.1% (w/v) SDS, 60 mM etylenediaminetetraacetic acid, 0.5% (w/v) bromophenol blue, 1.5% (w/v) blue dextran) was added to the reaction. To remove un-incorporated isotopes, the mixture was passed through a Biogel P60 (Bio-Rad, UK) column to collect the dextran blue-labelled fraction. The probe was then denatured by heating at

100°C for 5 min, cooled on ice for 2 min, before adding it to a hybridisation bottle.

2.5.8.3 Hybridisation condition

Hybridisation was carried out according to standard procedures [148]. The membrane was placed in a hybridisation bottle (Hybaid Ltd. UK) and 30 mL of pre-hybridisation solution (6 X SSPE diluted from a 20 X stock (3M Sodium chloride, 0.2M Sodium dihydrogen phosphate monohydrate, 25 mM ethylenediaminetetraacetic acid, pH 7.4 adjusted with 10 M Sodium hydoxide), 5 X Denhardt's solution (diluted from a 50 X stock; 5 g Ficoll - type 400 pharmacia), 5 g polyvinylpyrrolidone, dissolved in 500 mL distilled water, 0.5% (w/v) Sodium dodecyl sulfate) added. Herring sperm DNA was denatured by boiling for 5 min and incubated on ice for 2 min and 500 uL added and incubated at 65°C for 4 h. The denatured radioactive probe was added and incubated for further 18 h at 65°C, after which the solution was carefully discarded and 30 mL of 2 X SSPE (0.1% (w/v) SDS, 0.1% (w/v) Sodium pyrophosphate, 2 X SSPE (from 20 SSPE stock)) added. This was incubated at 65°C for 30 min in a hybridisation oven.

For non-radioactive hybridisation, the DNA bound membrane was rolled and placed in a hybridisation bottle (Hybaid Ltd. UK) and 30 mL of Southern hybridisation buffer (0.5 M NaPO4, 7% (w/v) SDS, adjusted to pH 7) added before incubating at 62°C for at least 30 min. The probe was prepared by adding the purified DIG-labelled PCR product in freshly prepared hybridisation buffer (25-50ng mL) in a 50 mL falcon tube. This was boiled at 100°C for 5 min and cooled on ice for 2 min to denature the DNA. The hybridisation buffer was

poured off and the probe (mixed in hybridisation) buffer added and incubated at 62°C for at least 6 h or preferably overnight. Used probe solution was stored at -20°C for re-use. The membrane was washed twice by adding Southern wash buffer (0.1 M NaPO4, 1% (w/v) SDS, pH 7) and incubated for 15 min at 62°C.

2.5.8.4 Chemiluminescent detection of DIG-labelled DNA

After hybridisation, the membrane was removed from hybridisation bottle and placed into a plastic tray containing 20 mL DIG was buffer (DIG Buffer 1 [0.1 M maleic acid, 0.15 M NaCl, 5 M NaOH adjust pH to 7.5], 0.3% (v/v) Tween-20). This was incubated on a rolling platform at room temperature for 5 min. DIG-wash buffer was removed and replaced with 25 mL blocking solution (DIG-Buffer 1, 1% (w/v) milk powder) and incubated on a rolling platform for at least 30 min. The blocking solution was poured off and a solution containing 2 µL anti-DIG-alkaline phosphatase antibody in 20 mL of blocking solution added and incubated for 30 min at room temperature. DIG wash buffer was used to wash the membrane twice for 15 min each time. The membrane was then placed on a plastic sheet of paper, 1 mL of CDP-Star solution (Roche) was added to the membrane and an equal size plastic sheet of paper used to cover it and allow the easy spread of the solution, avoiding bubbles. This was incubated at room temperature for 5 min and the CD-Star solution drained off without letting the membrane to dry completely. The membrane sandwiched between the two plastic papers was placed inside a film cassette and incubated at 37°C for 15 min. In a dark room, an X-ray film was exposed to the membrane and incubated at room temperature for a minimum 1 min and developed on X-ray processor (Protec OPTIMAX®, Germany).

2.6 Transformation of *M. oryzae*

The *M. oryzae* transformation protocol was adapted from Talbot *et al* 1993 [53]. A 1" square piece of mycelium was cut from the growing edge of a fungal colony on CM agar and blended in 150 mL liquid CM. The blended mixture was incubated at 24°C with shaking at 125 rpm for 48 h. Mycelium was recovered by filtering through sterile Miracloth (Calbiochem) and rinsed thoroughly with sterile distilled water. The mycelium was then transferred into a sterile falcon tube (Becton Dickinson) and 40 mL of OM buffer (1.2 M Magnesium sulfate, 10 mM Sodium sulfate, 5% (w/v) Glucanex (Novo Industries, Copenhagen), pH 5.8) added, and incubated at 30°C shaking at 75 rpm for 2-4 h. The OM buffer containing mycelial debris was then transferred into a sterile polycarbonate Oakridge high speed centrifuge tubes (Nalgene^{TM}) and carefully overlaid with an equal volume of ice-cold ST buffer (0.6 M sorbitol, 0.1 M Tris-HCl pH 7.0). This was centrifuged at 5000 *x g* for 15 min at 4°C in a swinging bucket rotor (Beckman J2.MC centrifuge). The protoplasts were recovered at the interface between OM and ST buffers and transferred into a new sterile Oakridge tube and overlaid with cold STC buffer (1.2 M sorbitol, 10 mM Tris-HCl pH 7.5, 10 mM calcium chloride). Protoplasts were pelleted by centrifugation at 3000 *x g* for 10 min at 4°C, after which the supernatant was removed and the pellet re-suspended in cold STC. The wash step was repeated twice with 10 mL of STC buffer, after which protoplasts were re-suspended in 1 mL of cold STC. Protoplasts concentration was determined using a haemocytometer (Improved NEUBAUER, Hawksley, UK).

Protoplasts were re-suspended 150 µl in STC to a concentration of 1 X 10^6 to 1.0 X 10^7 mL-1, and mixed with 2 - 6ug of DNA in a 1.5 mL Eppendorf tube. The mixture was incubated at room temperature for 25 min, after which 1 mL of PTC buffer 60% (w/v) PEG 4000, 10 mM Tris-HCl pH 7.5 and 10 mM Calcium chloride) was added, mixed by gentle inversion, and incubated at room temperature for 25 min. The mixture was transferred into a 12 mL falcon tube and 3 mL TB3 buffer (20% (w/v) sucrose, 0.3% (w/v) yeast extract) added and incubated with shaking for at least 16 h at 24°C. For Hygromycin B selection, the mixture was added into molten (45°C) osmotically stable CM (OCM) containing 0.8M sucrose and 1.5% (w/v) agar, mixed gently and poured into 5-6 sterile petri dishes (appx 20 - 25 mL/plate). The plates were incubated at 24°C in the dark for at least 16 h before overlaying with molten complete medium (CM) agar containing Hygromycin B (an aminoglycoside antibiotic against growth of prokaryotic and eukaryotic) to a final concentration of 200 µg mL^{-1} per plate. For selection against Sulfonylurea (Chlorimuron ethyl) or Glufosinate ammonium (Basta), molten (45°C) BDCM-bottom media (0.8 M sucrose, 1.7 g L-1 yeast nitrogen base without amino acids and Ammonium sulphate (Difco), 2 g L-1 Ammonium nitrate, 1 g L-1 asparagine, 10 g L-1 glucose [pH 6.0]) was used. Transformation mixture (protoplasts, DNA and 1 mL of PTC buffer) was added into molten (45°C) BDCM (bottom) and poured into 5-6 sterile petri dishes (appx 20-25 mL/plate). BDCM (top) medium without sucrose and containing a final concentration of 150 µg mL-1 chlorimuron ethyl or Glufosinate ammonium was used to overlay BDCM plates. This was incubated at 24°C for at least 14 days until transformants colonies emerged. Transformants were sub-cultured onto new BDCM (top) plates containing 100 µg/mL of Chlorimuron

ethyl (Sulfonylurea) or Glufosinate ammonium (Basta). The transformants were stored on sterile filter paper at -20°C for long term storage.

2.7 RNA Sequencing

2.7.1 RNA Isolation from *M. oryzae* infected rice leaf material

To obtain DNA-free RNA from fresh rice tissue infected with *M. oryzae,* a commercial kit (QIAGEN RNeasy Plant Mini Kit) was used according to manufacturer's instructions. This process involves a combination of guanidine-isothiocyanate lysis and silica-membrane mini spin column for RNA purification. For an extensive *in-planta* gene expression profile, three biological replicates of samples were collected at 5 time points of 24, 36, 48, 59 and 72 h post-infection (hpi) respectively. Leaf drop infection assays were also used to increase the fungal: plant material ratio. Small pieces of leaf tissue around the inoculated areas were collected into 1.5 mL Eppendorf tubes and quickly frozen in liquid nitrogen. Infected plant material, (100 mg) was ground into fine powder using a sterile nuclease-free mortar & pestle containing liquid nitrogen. The powder was then transferred into chilled 2 mL Eppendorf tubes and 450 µL of buffer RLT (containing 10 µL β-mercaptoethanol for every 1 mL of the RLT buffer) added and mixed thoroughly. In order to homogenise and filter the viscous plant and fungal material, lysate was transferred to a QIAshredder spin column placed in a 2 mL collection tube and was centrifuged at 16,000 *x g* for 2 min. The supernatant of the flow-through was transferred into a sterile eppendorf tube and mixed with 0.5 vols of absolute ethanol. The mixture was transferred onto the RNeasy mini spin column placed in a 2 mL collection tube, and centrifuged for 15 sec at 16,000 *x g*. After discarding the flow-through, 700 µL of Buffer

RW1 and 500 µL Buffer RPE were subsequently added to the RNeasy spin column with centrifugation for 15 sec at 16,000 *x g* discarding the flow-through each time. Additional 500 µL Buffer RPE was added onto the column and centrifuged at 16,000 *x g* for 2 min. The RNeasy spin column was transferred into a sterile 1.5 mL Eppendorf tube and 30-50 µL RNase-free water added to eluted bound RNA. The RNA was then stored at -80°C. The RNA quality was determined using a NanoDrop spectrophotometer (Thermo Scientific, UK). An aliqout of each sample was analysed on an Agilent 2100 Bioanalyser using an RNA nano-chip kit (Agilent Technologies, UK). Library preparation was carried out using Illumina® sequencing TruSeq Stranded Total RNA Library Prep Kit following manufacturer's instructions. The quality and quantity of the libraries was determined using Agilent 2100 Bioanalyser on a RNA 1000 chip kit. Individual sequencing libraries were prepared from each individual time point post-infection as well as mycelial RNA (three biological replicates per time point). Sequencing of 100 base paired-ends reads was carried out using Illumina Genome Analyser GXII platform by the Exeter Sequencing Service (University of Exeter).

2.7.2 Read mapping and determining differentially regulated genes in *M. oryzae*

Low-quality reads were removed and any sequences containing adaptor sequences trimmed or removed as well. To do this, the fastq-mcf program from the ea-utils package (http://code.google.com/p/ea-utils/) was used. Cleaned sequence reads were mapped to the *M. oryzae* 70-15 reference genome version 8 [24] using TopHat2 splice site-aware aligner [150]. The TopHat2

program uses an aligner called Bowtie. To identify isolate-specific differentially expressed genes, the cleaned sequence reads were also mapped onto genome sequence of strain KE002. Filtered sequence reads were also mapped to the *O. sativa* L. ssp *indica* genome [151] to identify rice-specific differentially expressed genes. Relative transcript abundance and differential gene expression was estimated using Cufflinks (http://cole-trapnell-lab.github.io/cufflinks/cuffdiff/)

Differential expression analysis was carried out using DESeq2 packages [152]. HTseq-count function of the HTSeq package [153] was employed to determine read counts mapping to each gene on the *M. oryzae* or *Oryza sativa* genome.

2.8 Whole genome Sequencing

Purified RNA-free DNA was obtained using Hexadecyltrimethylammonium Bromide (CTAB) procedure as described earlier in Section 2.5.1.1. Template quality was assessed using a NanoDrop spectrophotometer (Thermo Scientific) and Qubit BR assay (Thermo Fischer, NY, USA) indicating the concentration of double-stranded DNA. Sequencing was carried out at Exeter Sequencing services, University of Exeter, UK. NEXTflex^{TM} Rapid DNA-seq Library Prep Kit was used to prepare and index libraries before sequencing on HiSeq 2500 (Illumina), with two lanes per sample.

2.8.1 Read Mapping and genome assembly

The quality of sequencing reads was first checked using the FastQC toolkit (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Reads with a median quality score above Q20 were considered acceptable. From the raw data (*fastq* files), adaptor sequences were trimmed from sequences containing adaptors and low quality reads removed using the fastq-mcf program. The trimmed sequences were aligned to the reference genome (70-15) [24] using BWA (Burrow Wheeler Aligner) https://github.com/lh3/bwa [154].

2.9 Transient expression

2.9.1 Growth and maintenance of *Agrobacaterium tumefaciens*

Agrobacaterium tumefaciens strain GV3101 [155] was used for transient expression. This strain is resistant to gentamycin (50 µg/mL) and rifampicin (10 µg/mL) respectively, and was used for selection in liquid or solid media. Kanamycin and spectinomycin are toxic to GV3101 and can be used to select vectors conferring resistance to these antibiotics. GV3101 was grown at 28°C and stored at –80°C in 80% glycerol stock for long term storage. For transformation, 50 µL of competent *Agrobacterium* cells were thawed on ice and 300 ng of plasmid DNA added and gently mixed. The cells were incubated on ice for 5 min and quickly transferred into liquid nitrogen. Cells were then transferred to a 37°C water bath for 15 min. A 450 µL aliquot of SOC media was added and incubated with shaking at 28°C for 1 h. The transformed cells were plated on LB agar containing appropriate antibiotics and incubated at 28°C for 2-3 days. Three week old *N. benthamiana* leaves were infiltrated with

transformed *Agrobacterium* carrying T-DNA constructs, expressing the gene of interest. Plants were grown in a controlled room with a temperature range of 22-25°C and 85% relative humidity. Bacterial cultures were diluted to obtain a final OD 600nm of 0.25 -0.3 in agro-infilitration buffer (10mM MES, 10 Mm MgCl2, 150 µM acetosyringone, pH 5.6). To increase plant cells expression, plants were co-infiltrated (ration of 1:1 with *A. tumefaciens* strains expressing gene of interest and another expressing P19, a tomato bushy stunt virus protein that suppresses post-transcriptional RNA silencing. Leaf discs were cut from agro-infiltrated tissue 48 hpi and observed on a Leica TCS SP8 laser confocal microscope.

2.10 Yeast two hybrid [156]

To clone genes of interest into bait (pGBKT7 DNA-BD cloning) vector or prey (pGADT7 Activating domain) vector, In-Fusion HD Cloning Kit (Clontech) was used. The bait vector was digested using *Bam*H1 and *Eco*R1 while the gene to be inserted was PCR-amplified using primers containing 15 bp overhangs with homology to the two ends of the digested bait vector. The prey vector was digested using *Eco*R1 and *Sma*1 (New England Biolabs Hitchin, UK), while the gene to be inserted was amplified by PCR using primers containing a 15 bp overhangs with homology to the two ends of a digested prey vector. For transformation a single colony Y2H gold yeast strain was mixed in 1 mL of liquid yeast extract peptone dextrose (YPD). The mixture was transferred into an Erlenmeyer flask containing 50 mL of YPD and incubated at 30°C for 16-18h shaking (250 rpm). A 30 mL aliquot of this growing culture was transferred into 300mL of fresh YPD and incubated by shaking (230 rpm) at 30°C for 2-3 h.

The cultures were transferred into 50 mL falcon tubes and centrifuged at 700 X g for 5 min at room temperature. The supernatant was discarded and the cells re-suspended in 1.5 mL of 1.1 X TE/LiAc (1.1 mL 10 X LiAc, 1.1 mL 10 X TE buffer in 10 mL Milli-Q (MQ) water). The cells were then transferred into 1.5 mL microcentrifuge tubes and centrifuged at high speed for 15 sec. The supernatant was discarded and cells re-suspended in 600 µL of 1.1 X TE/LiAc ready for transformation. In a 1.5 mL microcentrifuge tube, 50 µL of Y2H gold yeast cells were mixed with plasmid DNA (at least 100 ng each of bait and prey vectors), 50 µL of herring sperm and PEG/LiAc. Herring sperm was first denatured by heating at 95-100°C for 5 min and then cooled on ice for 2 min before use. The mixture was incubated at 30°C for 30 min and heat shocked on a 42°C water bath for 15 min. The cells were then centrifuged at 4000 *x g* for 10 sec. The pellet was re-suspended in 150 µL of MQ water before plating on appropriate drop out medium. Plates were incubated at 30°C for 2-3 days until colonies appeared.

2.11 Co-immunoprecipitation from infected rice plant tissue

2.11.1 Protein extraction

For maximum accumulation of effector proteins in host rice cells, samples were collected at 36 and 48 h post-infection. Fresh rice leaf tissue inoculated with *M. oryzae* were collected and frozen in liquid nitrogen. Samples were ground in liquid nitrogen using mortar and pestle. Ground powder of 300 - 500 mg was quickly transferred into 1.5 mL a Eppendorf tube and 1 mL of ice-cold extraction buffer (GTEN [10% (v/v) glycerol, 25 Mm Tris pH 7.5, 1 mM EDTA 150 mM NaCl], 2% (w/v) PVPP, 10 Mm DTT, 1 X protease inhibitor

cocktail (Sigma), 0.1% (v/v) Tween 20 (Sigma) added and thoroughly mixed. This was centrifuged at 3, 000 *x g* for 10 min at 4°C to recover total proteins in the supernatant. The supernatant was transferred into a clean micro centrifuge tube and centrifuged further at high speed for 10 min at 4 °C. A 250 µL aliquot of the supernatant was transferred into a 2 mL microcentrifuge tube and the final volume brought up to 2 mL with dilution buffer (10 mM Tris/Cl pH 7.5; 150 mM NaCl, 0.5 mM EDTA). An aliquot of the sample was run on SDS-PAGE gel and used for Western blotting [157]. What remained of the sample was used for GFP-trap or stored at – 80°C.

2.11.2 Immunoprecipitation

GFP-trap beads (ChromoTek Biotechnology, Planegg, Germany) (mouse anti:GFP-antibody covalently bound on agarose beads surface) were prepared by cleaning twice with 5 X vol of dilution buffer. Total protein extract (50uL) was added to equilibrated, agarose beads re-suspended in ice cold dilution buffer. The sample and beads were incubated at 4°C with mixing for 1 h. The mixture was centrifuged at 2500 *x g* for 2 min and the supernatant discarded. The sample was washed three times using 500 µL ice cold wash buffer (10 mM Tris/Cl pH 7.5; 150 mM NaCl, 0.5 mM EDTA) and re-suspended in 100 µL 2 X SDS-sample buffer (120mM Tris/Cl pH 6.8, 20% (v/v) glycerol; 4% (w/v) SDS, 0.04% (v/v) bromophenol blue; 10% (v/v) β-mercaptoethanol). The re-suspended beads were boiled for 10 min at 95°C to unbound immobilised complexes from the beads and centrifuged at 2,500 *x g* for 2 min to recover the protein sample in the supernatant.

2.11.3 Western blotting

Protein sample (total protein or immune-precipitated) and loading buffer (12% (w/v) SDS, 30% (v/v) 2-mercaptoethanol, 50% glycerol, 0.012% (v/v) bromophenol blue, 0.375 M Tris-HCl, pH 6.8) were mixed in the ratio of 1:4. The mixture was loaded on 10% (w/v) protein gel (Mini-PROTEAN® TGX^{TM}). To estimate the size of protein fragments, a protein ladder was included in the run. The gel was initially run for 30 min at 100V followed by additional 40 min at 150 V.

2.11.4 Membrane transfer and antibody staining

The Amersham Protran Nitrocellulose membrane (GE Healthcare Life Sciences, Buckingham, UK) was washed by pre-wetting in MQ water for 10 sec and equilibrated in cold transfer buffer (25 mM Tris, 190 mM Glycine and 20% (v/v) methanol) for 5 min. The transfer set up was as follows: the cathode plate was placed in a tray pre-filled with transfer buffer. A sponge pad was placed on the cassette followed by blotting paper and the gel placed on top of the blotting paper and pre-soaked nitrocellulose membrane was placed on the gel and another blotting paper placed on top of the gel. A second sponge pad was added after the second blotting paper, and the assembly gently closed using the anode plate. The assembled cassette was placed in the transfer tank containing transfer buffer and ran at 400 mA for 1.5 h.

2.11.5 Immunostaining

After successful transfer the membrane was incubated in blocking solution (TBST (TBS + 0.1% (v/v) tween) and 5% (w/v) milk) for 1 h. The

membrane was then washed three times with TBST on an orbital shaker for 10 min per wash. The membrane was incubated in 20 mL primary antibody mix (TBST, 2% (w/v) milk 0.5 µL anti-GFP antibody) overnight at 4°C. The membrane was washed three times with TBST, before adding the secondary antibody conjugate (alkaline phosphatase) and incubated shaking for 2 h at room temperature. The membrane was washed 2 times with TBST before developing using the alkaline phosphatase conjugate substrate kit (Bio-Rad).

2.11.6 Detection

Detection working solution was prepared by mixing equal volumes of the Peroxidase solution and Luminol enhancer solution (Thermo Fisher Scientific) (0.125mL per cm^2 of membrane). The solution was added onto the membrane and incubated at room temperature for 1 h. The working solution was prepared immediately after use. Excess liquid was drained off and the membrane placed between two clear plastic wraps. To capture chemiluminescent signal from Luminol oxidisation on the membrane, cooled charge couple devices (CCD) cameras from G: BOX analysis system (Syngene, Cambridge, UK) were used. Images were captured using GeneSys image capture software.

Chapter 3 Single-Molecule Real-Time sequencing combined with RNA-sequencing improves annotation the *M. oryzae* of small secreted proteins in Guy11 and KE002 genome sequences

3.1 Introduction

Genomic studies have greatly enabled scientists to understand the mechanisms employed by pathogens to quickly adapt to new hosts or to different environments [39, 47, 128, 158-163]. Understanding how phytopathogenic fungi adapt to new hosts for example is key to predicting how they adapt to new environments, fungicides, climate change and host resistance [39, 47, 128, 158-163]. The plasticity of fungal genomes leading to genome expansion, chromosomal reshuffling or deletions, has been proposed as being a major reason contributing to rapid adaptations shown by fungi [38, 158]. Adaptive evolution includes production of secondary metabolism enzymes, secreted enzymes, including cell-wall-degrading enzymes, environmental sensors and small secreted proteins to help fungi colonise new subjects and adapt to changes in environmental conditions [164-166].

The genome plasticity of most fungi and oomycetes can also facilitate dispensable chromosomes, horizontal gene transfer, expansion of gene families and differences in global patterns of gene expression [39, 47, 128, 158-163]. Several gene families in the fungal genome have been reported to undergo adaptive evolution. A good example is the tomato pathogen *Cladosporium fulvum,* effector *ECP2* that belongs to a large multigene superfamily of effectors and spreads across the fungal kingdom [167]. Recent studies of pathogenic fungi suggest that analysis carried out using NGS data do not present an

entirely accurate representation of genomic structure variation among different strains of same species [37, 41]. Assemblies generated from NGS are fragmented and lack continuity [37]. Additionally, these assemblies lack information regarding host-specific regions which contain a higher percentage of repetitive DNA sequences and have been shown to contribute to genome plasticity [38, 158]. These regions are important in the biology of phytopathogens because they often contain effector genes and other virulence determinant factors like transposable elements and secondary metabolism enzymes [38, 158].

Transposable elements are involved in fungal evolution i.e. driving micro-rearrangements, induced point mutations in coding or non-coding regions, and modification of gene expression [168-170]. Moreover, proliferation of transposable elements plays a major role in the expansion of filamentous plant pathogen genomes [38]. For example *P. infestans* and *B. graminis* have high proportion of repetitive DNA sequences that make up 74% and 65% of their genomes respectively [38, 141, 171]. *P. infestans* and *B. graminis* are hemi-biotrophs and biotrophs respectively, which suggest that a high percentage of repeated DNA sequence might be associated with adaptation to biotrophic phase of growth [38, 171].

Since the advent of whole-genome sequencing technologies and use of next generation sequencing (NGS), several genome assemblies of *M. oryzae* have been generated [41, 43, 55]. The first draft of the rice blast fungus genome was sequenced and published in 2005 by the Broad Institute [24]. The resulting genome was assembled into 38.8 Megabases (Mb) of 2,273 contiguous

sequences longer than 2 kilobases (kb). Generated sequences were re-ordered into 159 scaffolds of which 50% of the nucleotide bases were assembled in scaffolds longer than 1.6 Mb (N50 1.6 Mb) [24, 172]. Using available genome maps, the draft assembly was then ordered to generate seven chromosomes by guidance of genetic markers [24, 172]. Importantly, 65% of the genome (19 scaffolds) contained several markers that aided orientation onto the maps [24, 172]. To identify the ends of chromosomes, the telomere repeat motif (TTAGGG)n was used [24, 172].

The rice blast reference genome 70-15 is a genetic cross between rice and *Eragrostis curvula* (weeping lovegrass) infecting strains of *M. oryzae* [24]. Additionally, genome of 70-15 was assembled from short-read sequences, which means the genomic plastic region may not be well represented in rice infecting *M. oryzae* isolates. Whole genome sequencing of several *M. oryzae* isolates that differ in host specificity (rice, wheat, foxtail millet and goosegrass), and a *Magnaporthe grisea* isolate with specificity to crabgrass, revealed that the genome sizes of *Magnaporthe* isolates have a close range of genome, between 39 and 45 Mb [44]. The number of genes predicted from these genome assemblies ranged from 12,283 – 14,781 [44]. These isolates shared 14,966 families from the whole set of predicted genes and 63% of these occurred in all the sequenced genomes [44]. Moreover, the genomic composition of pathogenicity determinant genes, for example secondary metabolites and effectors, was shown to be conserved in most *M. oryzae* isolates [44]. However, 529 gene families are shared only among the non-rice pathogens, while 86 gene families were specific only to rice pathogens [44]. The adaptation to different hosts is therefore thought to be caused by a small number of specific

genes and there is no major gene flow among host-limited forms of *M. oryzae* [44].

It remains a challenge to obtain in depth analysis of genomes generated by NGS either during genome annotation or comparative genomics [37]. For example, if the read-length obtained from DNA sequencing is shorter than a repetitive sequence, the genome assembly process will be hampered, because the short reads will be collapsed into a single entity. Single-molecule, real time (SMRT) sequencing from Pacific Biosciences has become an established technology used in current biological research studies to improve mammalian, plant, fungal and bacterial *de novo* genome sequence assemblies [37, 41, 173-175]. Using this technology, Bao *et al* [41] showed that the quality of two *M. oryzae* isolates, FJ81278 and Guy11 genome assemblies could for example be drastically improved. Assemblies from this study yielded larger genomes (~10%) compared to assemblies from Illumina-based short reads [41]. Additionally, fragmentation of the assembled genomes was reduced by 95% and the N50 was drastically improved from 0.156 Mb to 4.13 Mb and 0.18 Mb to 3.28 Mb for FJ81278 and Guy11 respectively [41]. Using such an approach resulted in improved contiguous assemblies of 54 contigs from 1415 and 56 contigs from initial set of 1182 for, FJ81278 and Guy11 respectively [41]. Moreover, additional 239 and 149 genes were predicted in FJ81278 and Guy11, respectively [41]. The study identified more transposable elements (10% more) that could not be detected by Illumina sequencing [41].

Currently, it is possible to successfully assemble smaller genomes such as prokaryotic genomes using long read sequencing [37, 174]. However,

assembling eukaryotic genomes is not straight forward. Optical mapping is a technique that can be used to improve assembly of genomes. Optical mapping involves generating a genome-wide ordered restriction maps from long DNA molecules [37, 176, 177]. This followed by either aligning in-silico generated restriction maps of the genome or ordering NGS long-read generated sequences into chromosomes [37, 176, 177]. However, this technique is rarely used since the advent of NGS. Optical mapping can be useful in constructing notoriously difficult regions to assemble, especially in eukaryotes [37, 176, 177]. Faino *et al* [37] used a combination of PacBio-generated long reads and optical mapping to generate a gapless genome assembly of a filamentous fungus *Verticillium dahlia*. A combination of such new sequencing technologies and assembly strategies can be used to generate either contiguous or complete genome assemblies of eukaryotes [37].

Fungal pathogens are known to secrete a large repertoire of secreted effector proteins which play a major role during fungal-plant interactions and function to inhibit, modify, alter or modulate activities in host cells to the benefit of the proliferating pathogen [89]. Effectors are described as small secreted proteins and lack any known motif associated with either their function or translocation into host cells [93, [85]. Mostly, secretion of effectors occurs *in-planta* and gene expression levels will be associated with a specific time during infection[178] [179]. More often, effectors are secreted through conventional mechanisms, via the endoplasmic reticulum and Golgi-apparatus [2]. To undergo this process, effectors must therefore possess a signal peptide sequence in the N-terminus that helps its translocation [180]. This forms the first criterion by which candidate effectors can be identified using bioinformatics.

Secondly, fungal effectors are identified as small secreted proteins, ranging from 50-300 amino acids, although sometimes larger proteins can also act as effectors [39, 102, 140, 181]. Some of these secreted proteins possess higher cysteine content and have stable tertiary structure with disulphide bridges [90]. This gives them the ability to resist the harsh physiological stress in a plant apoplast; and provides another criterion for characterising apoplastic effectors. Another criterion used to identify effectors is the absence of protein orthologs outside the genus. Finally, some effectors have been characterised to be located in repetitive sequence/TEs enriched regions in the genome or occur as gene clusters [38]. Up to now the definition of fungal effector remains ambiguous, which means any secreted, *in planta* expressed fungal protein, can be defined as a putative effector [179]. Some of these criteria for example protein size description lack or require set thresholds [182]. Recent experiments have verified that several fungal effectors lack cysteines and describing effectors as cysteine-rich can be misleading [182]. A specific amino acid motif associated with effector translocation in host cells has been characterised in oomycetes. The motif is referred to as the RXLR-dEER motif (where X represents any amino acid) and is situated at the N-terminus of the protein sequence [97, 138]. This type of motif has not been characterised in fungi.

In this chapter, a combination of Illumina HiSeq 2500 and third-generation Pacbio RSII sequencing was used to generate improved genome assemblies of *M. oryzae* isolates Guy11 and KE002. The aim of this approach was to improve the contiguous assembly of these two isolates and subsequently facilitate gene prediction. The Guy11 genome assembly was further improved by incorporating optical mapping to allow an in-depth analysis

of rice blast fungus genome. RNA sequencing was then used to determine sets of genes involved in the disease process, both in the rice blast fungus and in a susceptible rice cultivar. The aim was to determine genes upregulated during plant-pathogen interaction and to gain insight into the function of un-characterised secreted proteins in *M. oryzae* and on putative effector host targets in rice. I then used a range of different bioinformatics tools to improve effector gene prediction, annotation and characterisation. A pipeline of gene prediction process was set up using a program called Maker, to predict genes that were previously not predictable through protein and domain homology. The main objective in this chapter was to identify novel effector protein encoding genes that maybe involved in fungal infection and characterise effector-encoding genes in a less virulent Kenyan isolate KE002.

3.2 Material and methods

3.2.1 General material and methods

For standard procedures used in this chapter see Chapter 2

3.2.2 Pacbio sequencing

Genomic DNA was extracted from *M. oryzae* and sheared to prepare a 20 Kb library and size selected to remove shorter DNA fragments. The P6 polymerase and C chemistry (P6-C4 was applied to 8 SMRT cells per sample). Generated reads were assembled using the long read assembly program SMARTdenovo https://github.com/ruanjue/smartdenovo. To improve accuracy of the assembled contigs, sequences were further polished using Quiver.

3.2.3 Illumina sequencing

DNA sequencing was carried out at the Exeter Sequencing service, University of Exeter, UK, using HiSeq 2500 (Illumina), with two lanes per sample. Read quality was checked using a FastQC toolkit http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/. Reads with a median quality score above Q20 were considered acceptable. From the raw data (*fastq* files), adaptor sequences were trimmed and low quality reads removed using the fastq-mcf program. Trimmed sequences were aligned to the reference genome (70-15) using BWA (Burrow Wheeler Aligner) and SPAdes http://bioinf.spbau.ru/en/spades the used to generate de novo assemblies [183].

3.2.4 RNA Sequencing

Three biological replicates of infected rice leaf tissue were collected at 24, 36, 48, 59 and 72 hpi. Two controls of KE002 mycelium RNA and un-infected *Moukoto* leaf tissue RNA were also included in the study. Leaf tissue was collected at the inoculated spot to increase the *M. oryzae* biomass in the sample, to improve detection of transcriptomic changes in lowly expressed *M. oryzae* genes (see Section 2.7.1). Total RNA was isolated from samples before sequencing for 100 bp paired-end reads using an Illumina HiSeq 2500 Generated reads were filtered using fastq-mcf program from ea-utils package (http://code.google.com/p/ea-utils/) and aligned to 70-15, Guy11 or KE002 genome sequences. Reads were further aligned to the *Oryza sativa* indica genome (http://plants.ensembl.org/Oryza_indica/Info/Index) Relative transcript abundance and differential gene expression were estimated using Cufflinks and heat maps for transcript abundance at each time point generated using R https://www.r-project.org/. (See Section 2.7.2).

3.2.5 Construction of *M. oryzae* optical genome maps using the Irys system

To extract high molecular weight DNA free from contamination and to minimise DNA fragmentation, Guy11 protoplasts were used for DNA extraction. Samples were submitted for optical mapping at the Earlham Institute, Norwich, UK using Bionano Genomics technology. Agarose plugs containing high molecular weight DNA were melted at 70°C and digested using GElase™ (epicentre, Wisconsin, USA) at 43°C. The sample quality was analysed using an Opgen Argus Q-card (Figure 3.7) and 300 ng of sample used for Nick Label

Repair and Stain (NLRS) reaction using *Nt.BspQ*1 (New England Biolabs). The NLRS sample was loaded onto a single flow cell on a Bionano chip. Maps generated from BioNano chips were *de novo* assembled into a consensus map using Bionano Genomics IrysSolve software [184].

3.2.6 Generating an Ontology of *M. oryzae* differentially expressed genes

Bowtie2-build was used to build a *M. oryzae* index and reads from RNA-seq of KE002 infection on *Moukoto* data aligned to KE002 genome. Gene expression was estimated by number of transcript per million mapped reads (FPKM) with filtering set to disregard genes with less than 2 counts per million in more than 10% of samples. Pearson's Correlation Coefficient (PCC) was used to calculate a pairwise similarity matrix to provide an estimate of correlated gene expression during infection for each pair of genes. In this matrix, genes that have a similarity score close to 1 are considered to have highly correlated expression, while those with 0 score are termed as un-correlated. (Courtesy of Dr. Ryan Ames, University of Exeter)

3.3 Results

3.3.1 Pacbio sequencing of two isolates reveals variation in number of predicted secreted proteins and effectors

Initial attempts to predict secreted protein and effector-encoding genes from Guy11 and KE002 produced fewer genes than expected. This could be related to the highly fragmented genome assemblies obtained from short-read sequencing. To overcome this limitation, single molecule real-time (SMRT) sequencing from Pacific Biosciences (PacBio) was used to generate long reads and improve genome assemblies of the two *M. oryzae* isolates. We reasoned that the information regarding the effector repertoire of these two isolates could be contained in some of the non-assembled regions [37, 41, 45]. We reasoned that a well assembled genome would facilitate improved gene prediction and annotation. Guy11 was selected because it is virulent to 20 out of the 24 rice monogenic rice lines screened and is thought to lack most avirulence genes or has virulent alleles of these genes (Figure 3.6, C).

On the other hand, KE002 is a Kenyan isolate that is avirulent on a significant number of selected rice monogenic lines and is therefore thought to carry several avirulence genes as shown in Figure 3.6, C. The KE002 isolate produced incompatible reactions on most analysed rice monogenic lines and its predicted set of genes might contain putative *AVR-Pit, AVR-Piz, AVR-Piz-5, AVR-Pi1, AVR-Pi11 (t), AVR-Pi12 (t), AVR-Pi19 (t),* and *AVR-Pi20.* To obtain high molecular weight DNA free from RNA, carbohydrates and protein contamination; DNA isolation was carried out from protoplasts to avoid fungal cell wall contamination explained in Section 2.5.1.2. High molecular weight pure

DNA was submitted for sequencing at Exeter Sequencing service, University of Exeter, UK, on single molecule real-time (SMRT) RSII platform (Pacific biosciences). Generated reads ranged from 500 to 60,800 bases and with an N50 read length of 14,466 (Figure 3.1)

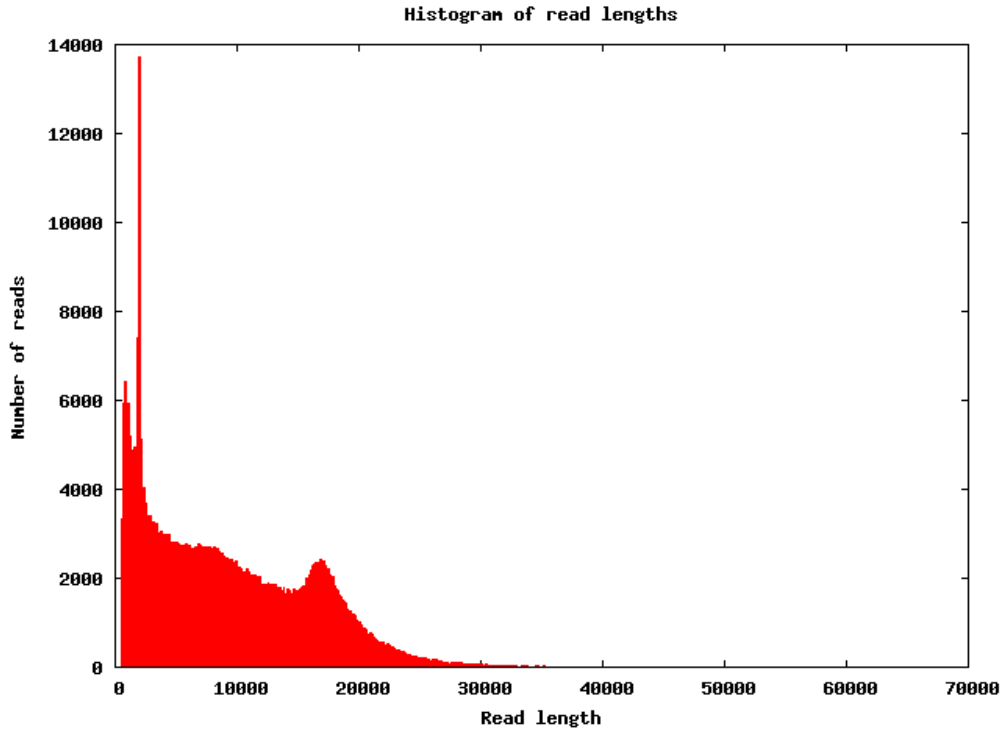

Figure 3.1 Read length sequencing data generated from a 20 Kb size-selected library prepared from Guy11 high molecular weight DNA

Reads obtained contained a minimum length of 500 bases, a maximum length of 60,800 bases and a sequence read length N50 of 14.466 Kb. Number of reads is shown on the Y-axis while the length of generated reads is shown on the X –axis

The Guy11 assembly generated by Pacbio sequencing comprised of 56 contigs with the longest contig approximately 4.5 Mb with an N50 contig length of 2.3Mb as shown in Table 3.1 and Figure 3.2. The KE002 assembly comprised of 42 contigs with maximum contig length of 6.9 Mb and N50 of 4.6 Mb (Table 3.1). The two assemblies therefore resulted in genome sizes of 43.3 Mb and 45 Mb for Guy11 and KE002 respectively as shown in Table 3.1. We observed that the genome sequences of the two isolates were assembled into sizes slightly larger than the reference *M. oryzae* genome, 70-15 as shown in Figure 3.2. The assembled genomes were also larger than genome assemblies generated on the Illumina platform which comprised 41004386 and 41111481 bp for Guy11 and KE002, respectively (Table 3.2). There is a possibility that with this approach, genome regions with difficulties to assembly were successfully assembled compared to the initial assemblies.

Table 3.1 Statistics of Guy11 and KE002 Pacbio *de novo* genome assembly using SMARTdenovo and polished using Quiver

Metric	Guy11	KE002
No. of contigs	52	42
Min contig (bp)	7,088	9,790
Max contig (bp)	4476857	6921104
N50 contig [1] (bp)	2324512	4650715
Total length (bp)	43304333	45074264
Reference length	**41027733**	**41027733**

[1] The N50 is a weighed median statistic in which 50% of the genome is contained in contigs equal or larger that given value, Guy11 (2.3 Mb) and KE002 (4.6 Mb).

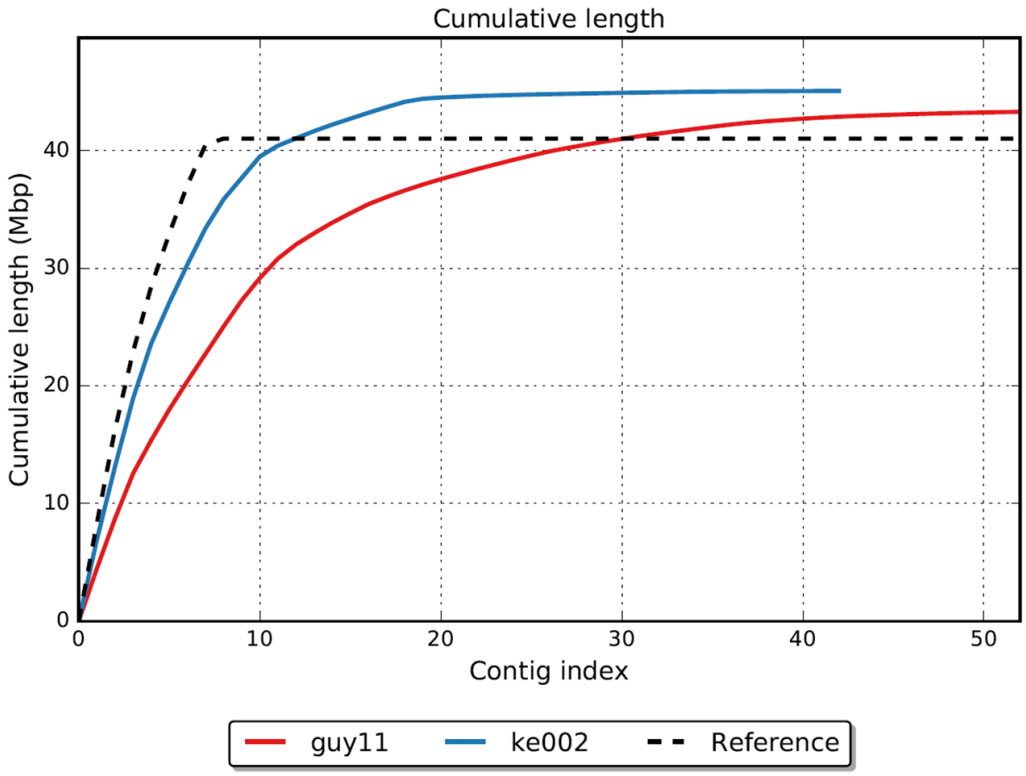

Figure 3.2 A comparison of *de novo M. oryzae* assembled genomes using Pacbio compared to the reference genome.

Analysis of the genome contiguous assemblies of Guy11 and KE002 compared to the reference genome 70-15 using SMARTdenovo together with Quiver. The predicted size of genome sequences is shown on the Y-axis while the contig index is shown on the X –axis. Plotted lines demonstrate the quality of assemblies in relation to reference genome 70-15. Both Guy11 and KE002 are predicted to have larger genome sizes compared to 70-15.

3.3.2 Pacbio sequencing and RNA-seq improves gene prediction in the *M. oryzae* isolate, KE002 genome

3.3.2.1 KE002 transcriptome analysis identifies putative effector protein-encoding genes

To facilitate a reliable mining of putative effector-encoding genes, we performed whole-genome sequencing of isolates that showed a range of virulence spectra towards differential monogenic rice series as shown in Figure 4.1 and Figure 4.2. Isolates BF5 and BF48 were selected as showing least number of susceptible interactions against monogenic rice lines and were hypothesised that they may carry most avirulence genes in the prevailing population. TZ090 was selected as the most virulent isolate while JUM1, BF17, BF32, BN0293, KE255, KE041, KE210, NG0135, NG0153, TG004 and UG08 had intermediate infection reactions. Other sequenced isolates KE002, KE016, KE017, KE019, KE021, KE029 and EG308 were also included in this study. Detailed virulence analysis of these isolates is discussed in Section 4.2. The gene calling program Augustus [185] was employed to predict genes from sequenced field isolate genome sequences of *M. oryzae*. Predicted genes from each isolate were aligned against predicted genes from other *M. oryzae* isolates using a stand-alone program, called *Proteinortho*, to group these genes into putative orthologs [186].

Fasta files containing predicted genes from the genome of a laboratory strain 70-15, the reference genome, and two other sequenced and well characterised field isolates Y34 and P131 were also included [43]. This step was taken to identify unique predicted genes that exist in only one isolate but

are missing in the rest (isolate-specific genes). A Fasta file containing protein sequences encoded by isolate-specific genes was then used to identify secreted proteins encoding genes by searching for the occurrence of a signal peptide in the N-terminus of each protein sequence using a standalone version of SignalP 4.1 program [187]. This analysis identified several predicted putative secreted protein-encoding genes specific to each of these isolates. Interestingly, not all molecularly cloned effectors/AVR genes could be predicted using this approach.

Most fungal effector proteins lack functionally characterised domains or homologs in closely related genome sequences or fungal species [188]. Therefore the chances of being able to identify them from an *ab initio* gene calling process is limited [188]. Most *ab initio* gene predictors work well for genes with homologs in other well studied genome sequences or highly conserved genes in eukaryotes and offer best guess for gene structures in less characterised genes/genomes [188]. However, *ab initio* gene predicting software can be configured to suit any genome and can be incorporated in other gene prediction pipelines such as Maker [189, 190].

The expression of most effector encoding genes is thought to be switched on during host colonisation [178] [179]. For this reason, RNA-seq reads/transcripts were incorporated into gene predictions to improve annotation. To determine genes encoding secreted proteins that are up-regulated during infection, conidia from KE002 were inoculated on a susceptible rice line *Moukoto* using the leaf drop method as described in section 3.2.4.

The GFOLD (generalized fold change) program was then used to determine changes in gene expression based on posterior distribution of log fold change https://www.tongji.edu.cn/~zhanglab/GFOLD/index.html [191]. Genes with a $\log_2$ ratio of expression $\geq$ 1 and GFOLD score (0.01) > 1 in comparison to mycelial transcripts were considered as up-regulated. This suggested a strong involvement of these genes during the interaction between KE002 and rice tissue. As expected, known biotrophy-associated secreted protein encoding genes, *BAS1*, *BAS4*, *BAS107* were among the highly up-regulated genes (Figure 3.3). All genes that showed no detectable expression in mycelium, but showed high-levels of expression during plant infection were selected. To improve *de novo* gene predictions for this study, a program called Maker was subsequently used to predict genes in Guy11 and KE002 genomes.

Maker is able to identify repeats, to align ESTs and proteins to a genome, produce *ab initio* gene predictions and incorporate generated data into protein-encoding gene annotations [189]. Moreover, outputs from previous runs or supplied files containing gene transcripts from RNA sequencing can be used for training. First, a program called RepeatMasker was used to screen and remove repeated elements before feeding the masked genomes into the pipeline (Figure 3.4). The program was trained using RNA seq data including genes up-regulated either in KE002 mycelium or during rice infection by the fungus. This was followed by incorporating the *ab initio* gene calling programs Augustus, SNAP and GeneMark to predict genes present in KE002. Predicted genes were aligned to the 70-15 predicted gene transcripts and *M. oryzae* Swiss-Prot protein database used for annotation.

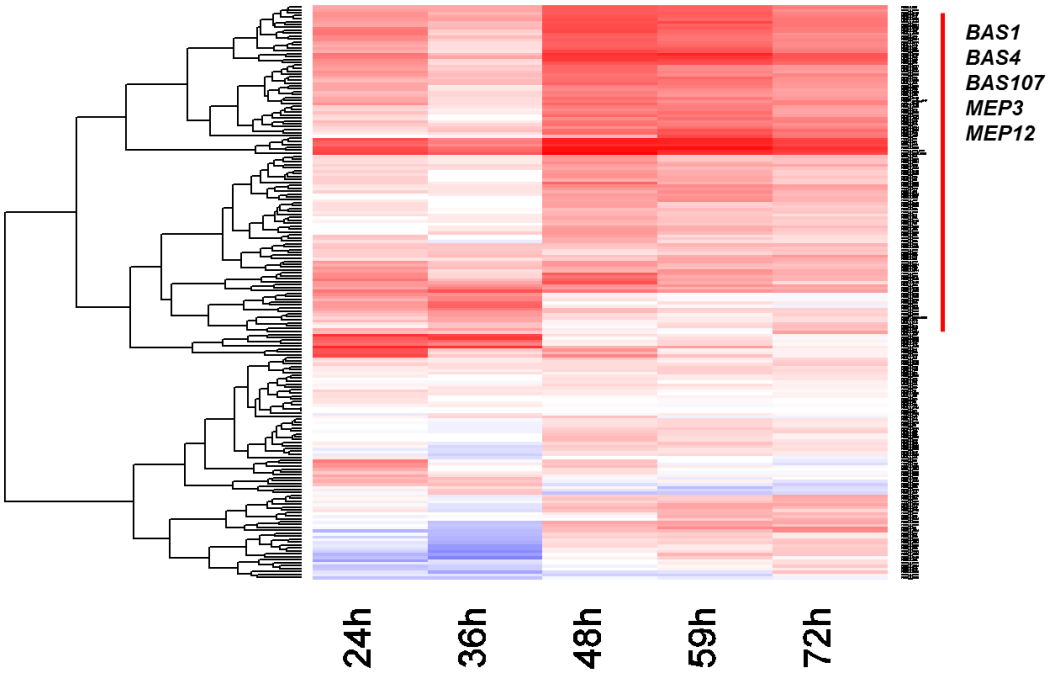

Figure 3.3 Heatmap showing transcript abundance of differentially expressed KE002 effectors and predicted genes encoding secreted proteins.

Levels of expression were calculated using $\log_2$ fold change during KE002 infection of *Moukoto* compared to KE002 mycelium gene expression (Blue= down-regulated Red= up-regulated). Genes that showed a 1-fold change expression were considered as up-regulated. Clade indicated with a red vertical line was enriched with known effector protein-encoding genes, and was thought to contain putative effector genes.

Several gene calling runs were conducted to increase the chances of predicting putative effector-encoding genes. The number of cloned effectors/avirulence-encoding genes predicted per run was then used to assess the accuracy of prediction as shown in Table 3.2. A fasta file containing predicted genes from two successful runs were clustered and a consensus list of predicted genes then obtained. The list of predicted putative effectors was used to conduct a BLASTn search in all sequenced genomes to determine presence/absence polymorphism. This made it possible to determine putative effectors-encoding genes specific to KE002, those missing in Guy11 and those shared by most sequenced isolates. Protein sequences of predicted genes were annotated using BLAST2GO [192].

In both Guy11 and KE002, it was possible to predict more genes in long-reads assembled genomes sequences than from short-reads assembled genomes (Table 3.3). In Guy11 for example, 49 more genes were predicted in long-reads assembled genome compared to short-reads assembled genome of the same strain. FASTA files containing predicted gene sets from the reference genome 70-15, Guy11 long-reads assembled genome and Guy11 short-reads assembled genome were aligned using a standalone program *Proteinortho*. A total of 11,319 (87.8%) genes annotated in 70-15 genome assembly were also predicted in Guy11 Pacbio assembly and Guy11 Illumina assembly and were thought to be conserved genes in the three *M. oryzae* genomes and easily identified during gene prediction. Strikingly, 319 genes predicted from short-reads assembled Guy11 genome did not align to genes predicted from the reference genome 70-15 or from Guy11 long-reads assembled genome (Figure 3.4). The 319 un-matched genes were further analysed using BLASTn and

most produced hits in the long-reads generated Guy11 genome sequence. We concluded that these were gene prediction artifacts or fragmentated genes and the alignment program *Proteinortho* could not match them to orthologous genes predicted from a contiguous Guy11 genome assembly generated using Pacbio sequencing. Poorly assembled genomic regions containing repeats might lead to prediction of fragmented putative genes. A set of genes that did not produce hits in the long-reads assembled Guy11 genome were further annotated using the NCBI online search tool. This search produced several hits from the *Enterospora canceri* genome and we concluded that there was a possible DNA sample contamination before sequencing.

In KE002, 590 more genes were predicted from long-reads generated genome assembly compared to short-reads assembled genome. These results suggests that these genomes contain 12,700 -13,000 genes. Further analysis using BLASTn showed that extra genes predicted from long-reads assembled genomes are not novel genes but families undergoing duplication and expansion. For example, a BLASTn search for *PWL2* in newly assembled showed a bigger event of expansion in Guy11 (3 copies) and 5 copies in KE002 as shown in Figure 5.5). Another effector gene, identified in this study, *MEP13* that encodes a BIC localised effector protein was also found to have undergone duplication and expansion in Guy11 (3 copies) (see Section 4.3.4.1**)**. BLAST search in the long-read generated Guy11 genome assembly produced three hits for *MEP13*. On contrary, during assembly of short-reads generated assembly, *PWL2* and *MEP13* collapsed into single copies and we could not detect duplication and expansion.

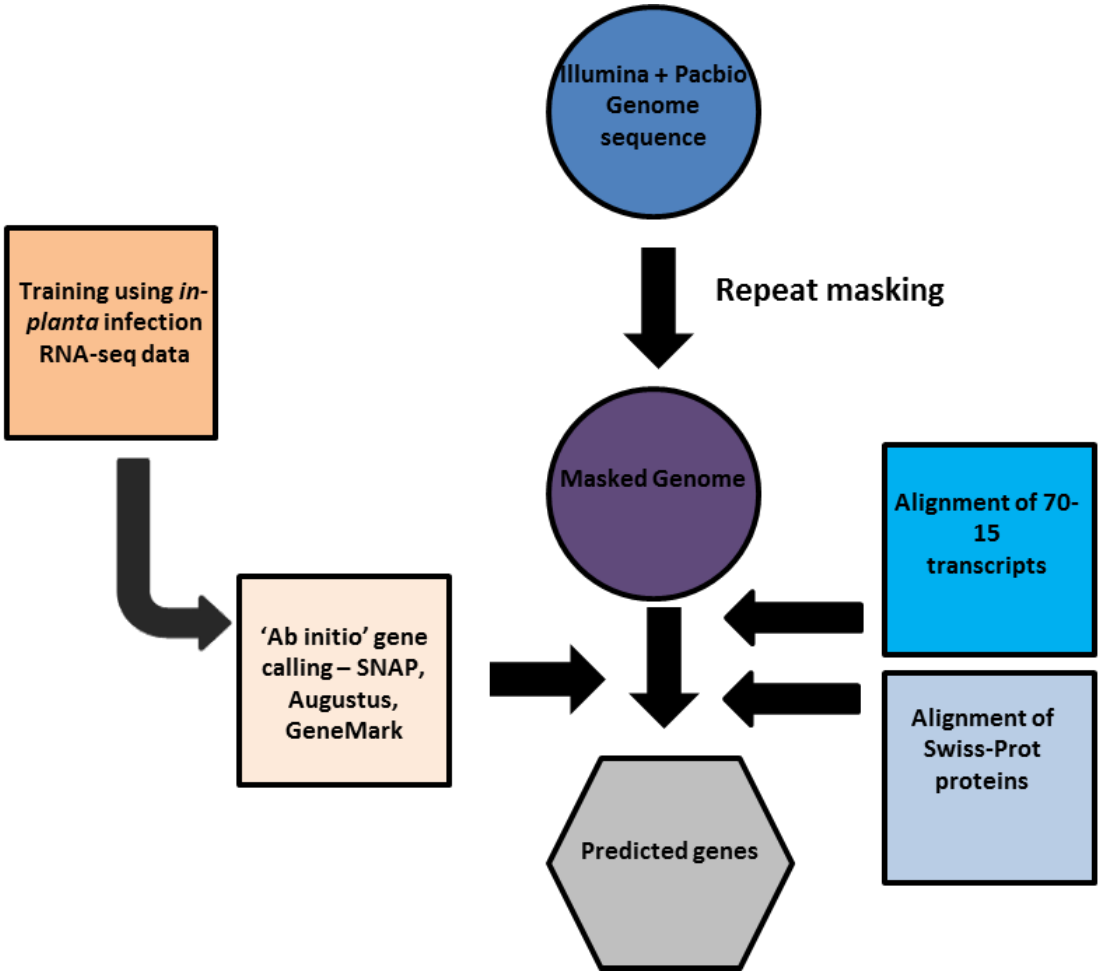

Figure 3.4 Schematic representation of gene calling pipeline using Maker.

(A) Maker makes use of other external executable programs including RepeatMasker (http://repeatmasker.org), alignment programs and incorporates several gene calling programs SNAP, Augustus and GeneMark. Maker was trained using RNA-seq data obtained from KE002 infection on susceptible rice line *Moukoto* normalised to KE002 mycelium grown on CM.

Table 3.2 Selected runs used for predict for putative effector protein-encoding genes

Run	Predicted genes	Up-regulated	Up-regulated with signal peptides[2]
Test_9	18896	1032	194
Tes_12	13306	931	180

[2] For test_9 run, RNA-seq data (junction file) was used for direct gene calling using Maker and output contained 18896 genes predicted and could predict all cloned effectors/AVRs. For test_12 (13306) run, Maker was employed to predict genes after RNA-seq training. This produced more accurate gene models but was unable to predict all the known effectors/AVRs (only 3 AVRs were predicted).

Table 3.3 Comparison of predicted genes from de novo assemblies of Guy11 and KE002 compiled using different technologies and gene calling programs

Isolate	Technology	Assembly size	Gene prediction	Number of genes
Guy11	Illumina	41004386	Augustus	10,534
Guy11	Illumina	41004386	Maker	13,083
Guy11	Pacbio	43304333	Maker	13,132
KE002	Illumina	41111481	Augustus	10,836
KE002	Illumina	41111481	Maker	12,716
KE002	Pacbio	45074264	Maker	13,306
70-15	Sanger	41027733	FGENESH	12,991

Table 3.4 Summary of predicted secreted proteins encoding genes in Guy11 and KE002 compared to the reference genome 70-15

Isolate	Technology	Predicted genes	Secreted	Effector predicted
Guy11	Pacbio	13132	1742	582
KE002	Pacbio	13306	1742	612
70-15	Sanger	12991	1762	621

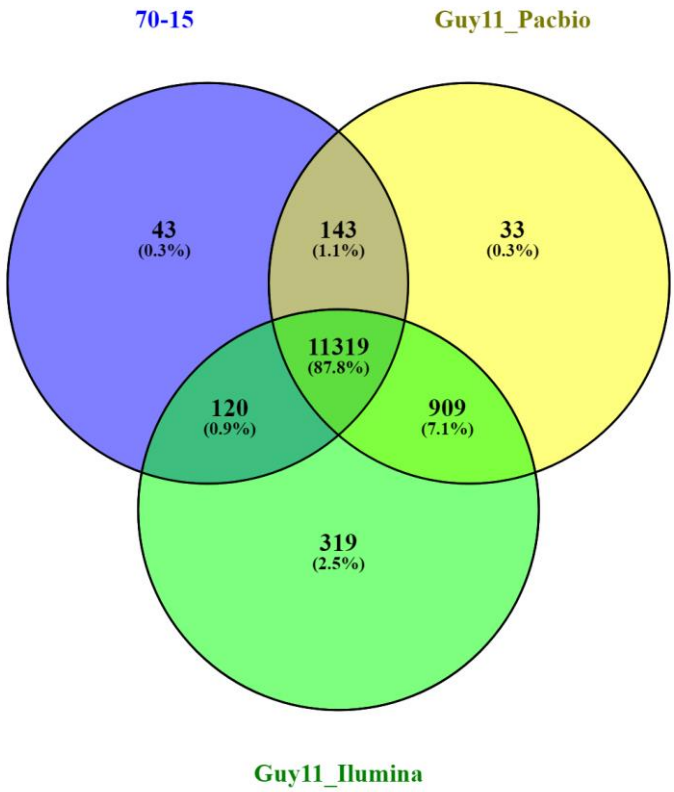

Figure 3.5 Venn diagram showing overlaps in the number of predicted genes from Guy11 *de novo* genome assemblies obtained using different technologies compared to the reference genome, 70-15

A total of 11,319 (87.8%) genes annotated in 70-15 genome assembly (blue) were also predicted in Guy11 Pacbio assembly (yellow) and Guy11 Illumina assembly (green). There is an overlap of 143 (1.1%) genes annotated in 70-15 that were predicted from Guy11 Pacbio assembly. An overlap of 120 (0.9%) genes annotated in 70-15 and predicted from Guy11 Illumina assembly. There is an overlap of 909 (7.1%) genes predicted in both Guy11 Pacbio and Illumina assemblies. The highest number of putative assembly-specific genes (319) predicted from Guy11 Illumina assembly (green) might be gene prediction artifacts, fragmentated predicted putative genes or contamination rather than extra predicted genes. Poorly assembled genomic regions containing repeats might lead to prediction of fragmented putative genes.

3.3.2.2 Identification of presence/absence polymorphism in predicted secreted proteins

To identify genes unique to Guy11, KE002 and the 70-15 genome sequences, presence/absence and polymorphism of predicted genes was analysed. FASTA files containing predicted gene sets from the reference genomes 70-15, Guy11 and KE002 long-reads assembled genomes were aligned using a standalone program *Proteinortho.* Out of 13,132 genes predicted in Guy11, 1742 contained putative signal peptide sequences characteristic of secreted protein-encoding genes. There were 74 genes encoding for secreted proteins identified as being unique to Guy11 and 232 predicted secreted proteins encoding genes were shared between Guy11 and 70-15, which confirmed the close relatedness between these strains. Out of 13,306 genes predicted in KE002, 1742 contained putative signal peptide sequences, among which 270 were classified as being unique to KE002 and missing in both 70-15 and Guy11. There were 102 predicted secreted protein encoding genes shared between KE002 and Guy11. However, BLAST analysis revealed that there was a significantly higher number of genes shared between the three isolates. There is a possibility that these genes were missed out during 70-15 genome annotation.

Selected genes encoding for secreted proteins were further annotated using *EffectorP* http://effectorp.csiro.au. This step was used to fast-track prioritisation of high-confidence effector candidates for functional characterisation as avirulence protein encoding genes. *EffectorP* uses the features of known fungal effectors to discriminate putative effectors from non-

effector protein encoding genes [182]. The annotation is performed based on sequence length, cysteine/serine or tryptophan content, molecular weight and net charge, giving it a sensitivity and specificity of up to 80% [182]. From this analysis, 582 secreted protein-encoding genes in Guy11 were annotated as putative effectors and only 33 of these were isolate-specific (Figure 3.6). As observed during the prediction of secreted protein-encoding genes, most of the predicted effector genes were shared between Guy11 and 70-15. 621 genes were annotated as effector protein-encoding genes in KE002 of which 139 were unique to KE002. KE002, is distantly related to either 70-15 or Guy11 and is hypothesised to contain more avirulence genes than Guy11 as shown in Figure 3.6.

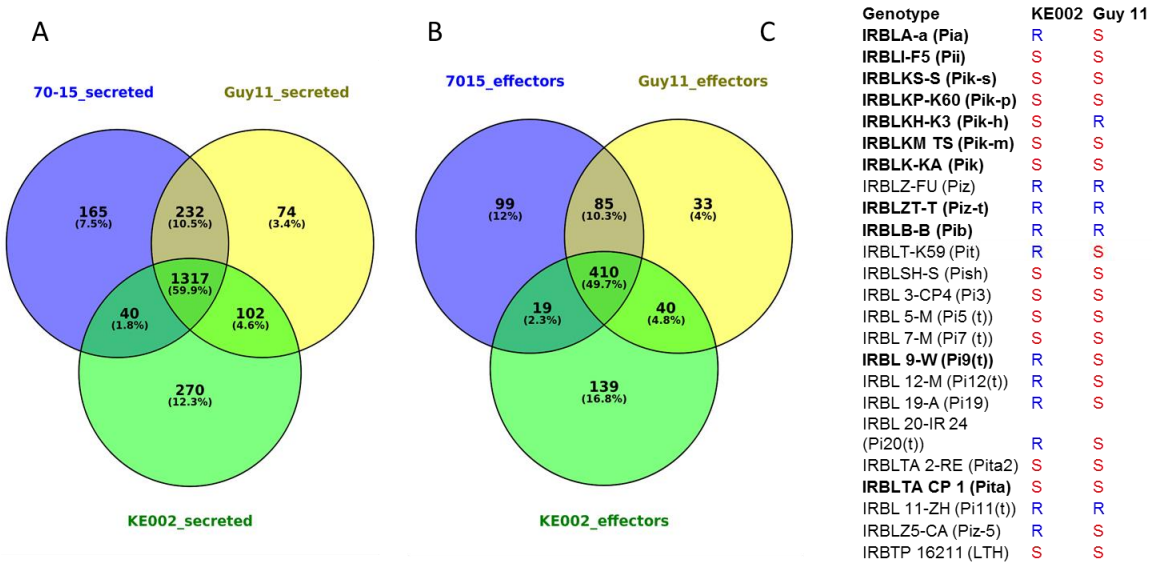

Figure 3.6 Venn diagram showing genes that are shared among different *M. oryzae* isolates.

(A) Variation between 70_15, Guy11 and KE002 predicted secreted protein-encoding genes. (B) Comparison between putative effectors (*EffectorP* annotated) 70-15, Guy11 and KE002. KE002 genome contains more isolate specific genes compared to 70-15 and Guy11 (closely related strains). (C) Virulence of Guy11 and KE002 on rice monogenic lines. Rice genotypes shown on the left and the two isolates on the right. Red (S) represents virulence and Blue (R) resistance. KE002 is avirulent to several rice monogenic lines.

3.3.2.3 Predicted *M. oryzae* effector genes contain putative transit sequences to different host-cell compartments

Different plant subcellular compartments contain different proteins that serve specialised biological functions [193, 194]. Such proteins are encoded in the host nucleus and require transit peptides in the N-terminus that assist in translocation from the cytosol into these organelles [193] [194]. Bacterial pathogens have been characterised to target different host cell compartments by secreting effectors that possess transit peptides directed to specific eukaryotic organelles [195-198]. Recently, the rust fungus, *Melampsora larici-populina* has been shown to target host cell chloroplast by mimicking host transit peptides [199]. However, the mechanism by which effectors, especially from *M. oryzae* enter plant organelles remains largely unknown. We used a bioinformatics approach to characterise putative effectors identified in 70-15, Guy11 and KE002. We employed an effector localisation prediction program called *Localizer* http://localizer.csiro.au/. This program has been trained to identify effector sub-cellular localisation using experimentally-verified plant proteins and pathogen effector protein-encoding genes [200].

Some effector proteins are thought to contain host cell transit signals separated from the signal peptide by a pro-domain and might be missed if analysis is carried out with the assumption that transit peptides occur immediately after signal peptides [200-202]. To accurately predict sub-cellular localisation of putative effectors, the predicted protein sequences including signal peptide sequences were submitted for analysis (Table 3.4). Out of the predicted effector repertoire, 14.2%, 13.9% and 13.1% for 70-15, Guy11 and

KE002 respectively were predicted to target the rice nucleus. *M. oryzae*'s Bas107 and a *P. infestans'* CRN8 are examples of pathogen secreted effectors that have been shown to localise to host nuclei [2, 197]. Our analysis revealed several effectors that were predicted to target the chloroplast, 5.5%, 5.5% and 5.2% for 70-15, Guy11 and KE002 respectively while the least number of effectors was predicted to target the mitochondria. *Localizer* can be used to enhance the process of effector characterisation. However, experimental verification of this analysis will be needed to validate these results and provide more insight to effector function.

Table 3.5 Properties of putative effector proteins in Guy11 and KE002 and 70-15 predicted using *Localizer*

No. of Effectors	70-15	Guy11	KE002
Number of proteins with CTP[3]	34 (5.5%)	32 (5.5%)	32 (5.2%)
Number of proteins with MTP[4]	7 (1.1%)	3 (0.5%)	5 (0.8%)
Number of proteins with NLS[5]	88 (14.2%)	81 (13.9%)	80 (13.1%)

[3] Chloroplast targeting protein

[4] Mitochondria targeting protein

[5] Nuclei localising protein

3.3.3 Single-Molecule Real-Time sequencing combined with Optical Mapping yields a nearly complete Guy11 genome

To further improve the contiguous assembly of the sequenced Guy11 genome, we employed optical mapping. This technique can be used to construct an ordered high resolution map of a genome from high molecular weight DNA (Figure 3.7). With this approach, restriction maps of Guy11 genome were generated to guide direct orientation of assembled contigs acquired from Pacbio and Illumina HiSeq 2500 sequencing. Following BioNano's Irys software workflow, Pacbio assembled contigs were aligned onto generated maps to obtain a hybrid assembly. Out of 52 NGS contigs, 29 successfully aligned to BioNano maps to give a total of 40.414 Mb constituted in 28 contigs (Table 3.5). Similarities between 70-15 and Guy11 was analysed using Mauve (http://gel.ahabs.wisc.edu/mauve). Mauve is a standalone program capable of aligning two or more genome sequences to determine either regional gain and loss or rearrangements [203].

The *M. oryzae* 70-15 was created by a genetic cross involving Guy11 and we hypothesised that these two genome sequences are closely related. Moreover, 11,462 *de novo* predicted genes from Guy11 clustered with genes annotated in 70-15 genomes further suggesting close relatedness of these two genomes [24] [44]. All major contigs from the Guy11 assembly aligned to the eight super contigs of 70-15 (Figure 3.8). Structural variations or reconstruction were observed in smaller contigs that might represent telomeric regions. This result suggests that the genomic composition of Guy11 and 70-15 are similar and these strains are closely related. Moreover this result is consistent with

Chiapello *et al* [44] study that showed only 2% of the 70-15 genome belongs to a weeping lovegrass pathogen.

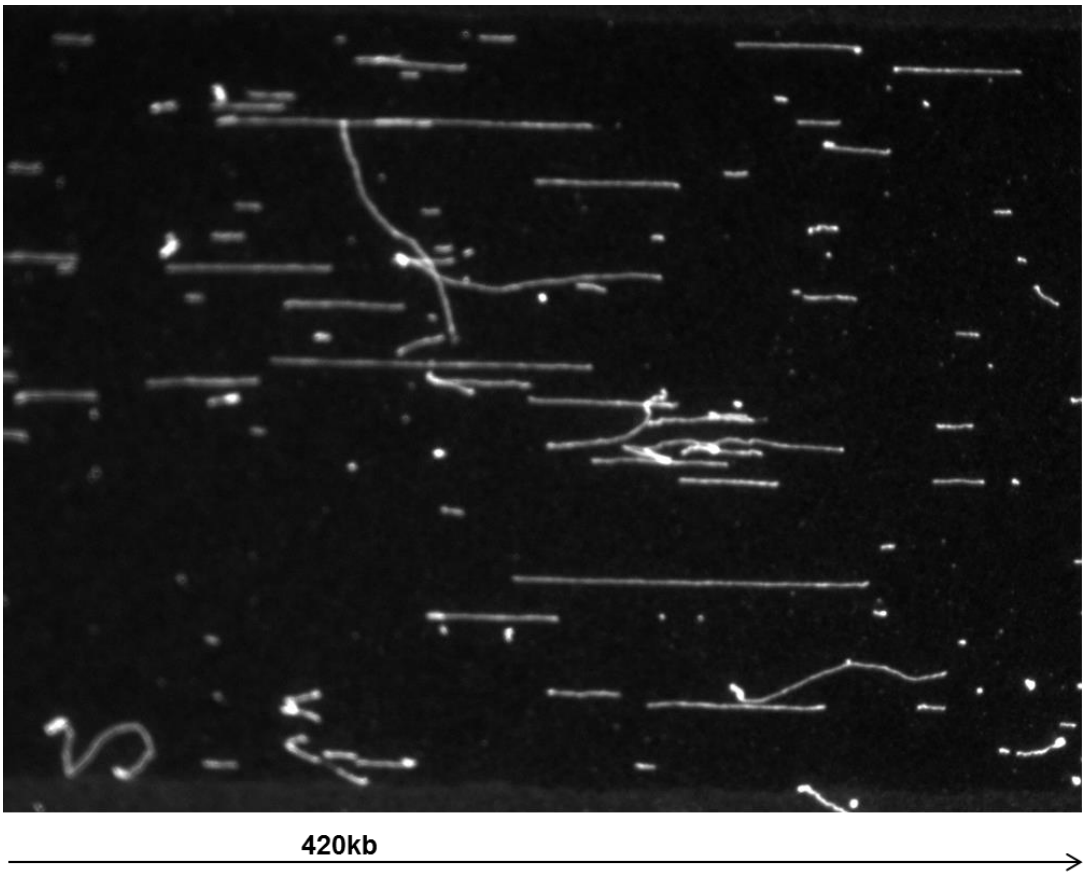

Figure 3.7 Representation of Guy11 DNA molecules immobilised and stretched onto an Opgen Argus Q-card.

DNA molecules were cut using *Kpn*1 stained using Argus Stain kit, and imaged by Argus imaging system at the Earlham Institute, Norwich, UK. The arrow represents the size of field of view which equate to 420 Kb strand length. Most of the selected molecules range from 100 – 150 Kb in size.

Table 3.6 Comparison of *de novo* assemblies of *M. oryzae* strain Guy11 using different technologies

Metric	Pacbio assembly	Pacbio + Optical map Hybrid
No. of contigs	52	28
Min contig	7,088	201,000
Max contig	4476857	4477000
N50 contig	2324512	2,404590
Total length	43304333	40414000
Reference length	**41027733**	**41027733**

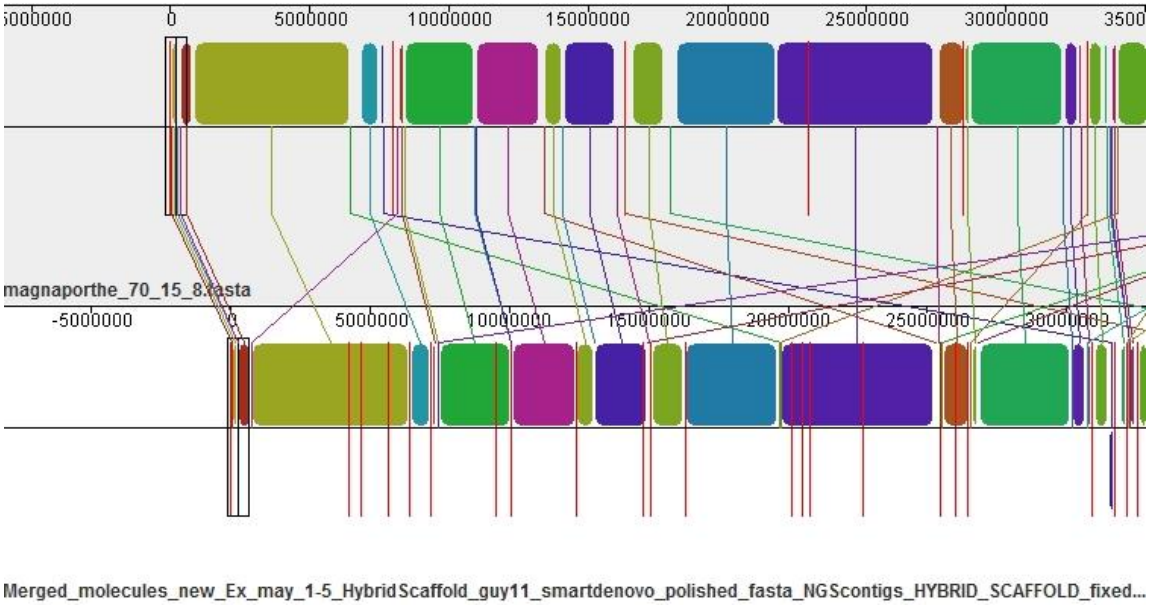

Figure 3.8 A pictorial representation of aligned hybrid assembly of Guy11 genome (bottom) and the reference genome (top).

Matching continuous regions are shown as solid coloured blocks while connecting lines indicate matching regions in both genomes. Alignment and reorientation of the contigs was generated using Mauve program.

3.3.4 Analysis of candidate effectors and putative host-targets expression during biotrophic invasion

3.3.4.1 Identification of genes involved biotrophic phase of rice infection by rice blast fungus

The transcription of effector encoding genes is thought to start when the rice blast fungus lands on the host leaf surface. Mostly, translation and secretion of effectors occurs *in planta* at a specific time of infection [139] [1]. The expression of genes encoding for secreted or predicted effectors proteins was analysed during infection of KE002 on a susceptible rice line *Moukoto*. By isolating RNA from rice leaf sheath infected with the rice blast fungus, Mosquera *et al* [1] were able to identify putative fungal and rice genes potentially involved in the biotrophic phase of a compatible reaction [1]. Gene expression was analysed at 36h post-infection [1]. At this stage of infection, invasive hyphae colonise multiple cells and begin to invade neighbouring cells infected tissue is therefore enriched with RNA from the invading fungus.

For this study, a different reproducible procedure was developed. First, rice leaves were inoculated instead of leaf sheath tissue. This approach was adapted in order to capture effector genes whose transcription begins prior to host cell penetration. Additionally, studies have shown that the onset of effector expression involved in the necrotic phase of infection occurs later, post-infection [110](Yan *et al* unpublished). In order to provide insight into the biotrophic and necrotic phases of infection, the infection process was therefore analysed from 24 to 72h post-infection. Using a $\log_2$ ratio of expression $\geq$ 1 threshold, 194 genes encoding secreted proteins showed up-regulation in three biological

replicates for at least one period of infection. Out of this group, 91 predicted putative effector encoding genes were found to be up-regulated compared to mycelial expression (Figure 3.9). Among these genes the biotrophy-associated secreted proteins, *BAS-4*, *BAS-107*, *MEP3* and *MEP12* were found to be highly up-regulated (Figure 3.3). Surprisingly, cloned effectors and avirulence genes *PWL2, BAS1, BAS2, BAS3, SLP1, AVR-Pizt, AVR-Pia* and *AVR-Pi9* in KE002 genome did not show up-regulation at this stage of rice infection.

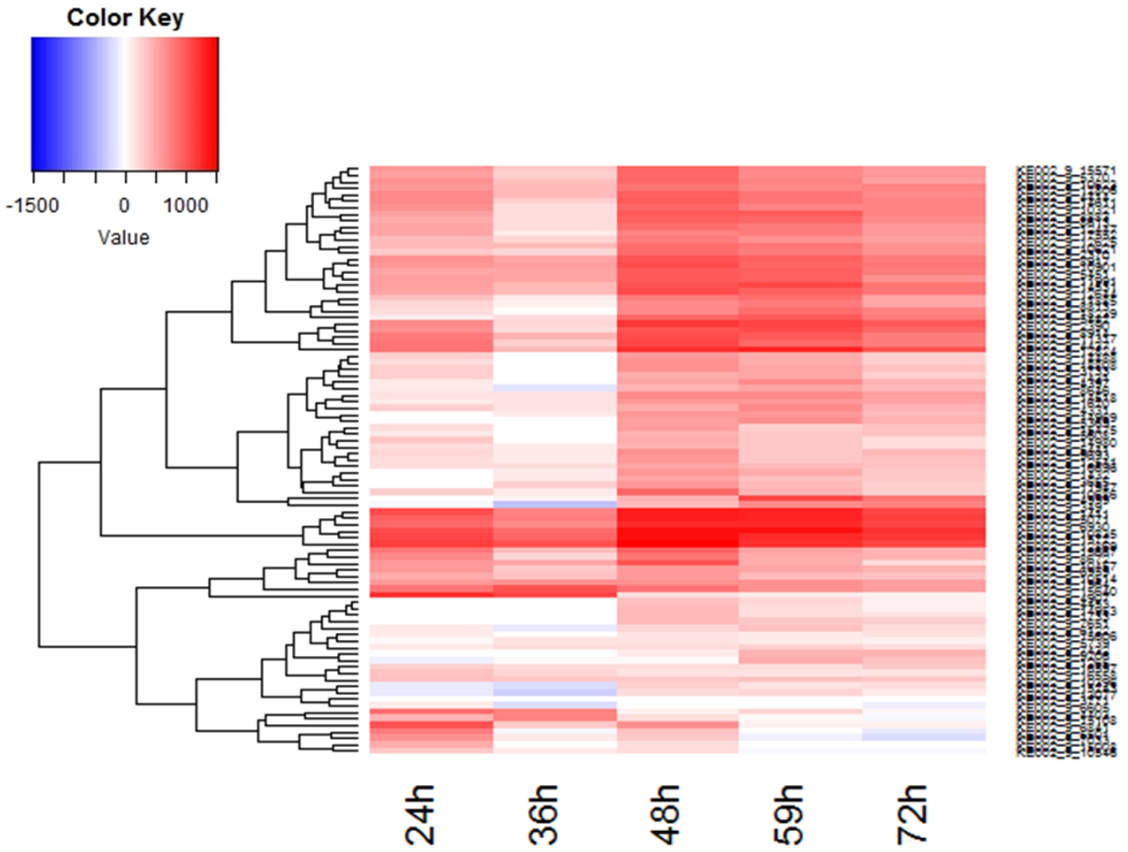

Figure 3.9 Heatmap showing transcript abundance of differentially expressed KE002 *EffectorP* annotated genes

Levels of gene expression are calculated using the logarithmic fold change during KE002 infection of *Moukoto* compared to KE002 mycelial gene expression (Blue= down-regulated Red= up-regulated).

3.3.4.2 Using Network Extracted Ontologies to predict genes associated with effector secretion

Despite availability of RNA-seq data at different stages of *M. oryze* infection process, a lot still needs to be done to understand specific pathways and genes involved at each stage of infection [204]. The function served by effectors at the infection stage is largely unknown [1, 2] [117, 118]. Understanding other well studied pathways involved during infection can be used to shade light on function of un-characterised effectors [204]. Network extracted ontologies (NeXOs) have been utilised to identify function of un-characterised genes and uncover unknown functional links [204]. During transcriptome data analysis, genes that play a central role in certain pathways might not show any differential expression [204]. For example, analysis of entities enriched with appressorium formation genes has helped to identify genes involved in appressorium formation that were previously not identified in differential expression studies [204]. I reasoned that understanding pathways related to effector secretion may highlight a novel potential function for a given effector. NeXOs was used to search for hierarchical structures resulting from KE002 infection on *Moukoto* RNA-seq data, to identify pathways that share the same expression patterns with known effectors and secreted proteins.

Gene expression ontology of *M. oryzae* was created from KE002 infection on *Moukoto* RNA-seq data, to determine the transcription relationship among genes involved in biotrophic growth using CliXO (version 0.3). Most effectors fell into large gene co-expression networks that contained multiple pathways up-regulated during infection and made it impossible to obtain

relevant information from such networks. However, one smaller differential gene co-expression network (regulome) contained 42 genes that included an effector effector encoding gene *BAS*4 and two other putative effector encoding genes, *MGG_08482* and *KE002_15475*. This, was further analysed (Figure 3.10). *MGG_08482* and *KE002_15475* showed a higher correlation with genes encoding for proteins with known functions including a sugar transporter, salicylate hydroxylase, cutinase, sucrose-6-phosphate hydrolase, gluconolactonase and a polyketide synthase.

These proteins might be involved in nutrient up-take or effector translocation mechanism during the biotrophic phase. However, it was not possible to infer the putative function of *MGG_08482* and *KE002_15475* from transcription correlation in this network. For example, cutinase is involved in the rice cuticle breakdown, a process but no studies have linked to the involvement of effectors. Polyketide synthases produce multiple secondary metabolites that exhibit different forms and play crucial role in plant-pathogen interactions [205]. AVR-CE1 is a characterised polyketide synthase [96]. The secondary metabolite synthesised by Ace1 is recognised by the cognate resistance protein *Pi33*, however the exact function served by this metabolite is not well known [96]. Some of these secondary metabolites might be involved in nutrient acquisition during fungal proliferation [206]. Furthermore, levels of long distance transported major carbon and nitrogen sources are known to increase at pre-symptomatic stage of infections [206]. At this stage of infection, the fungus has invaded a small percentage of plant and tissue and might explain the metabolic modulation and metabolite transport. *BAS4* expression did not show

transcription correlation with genes encoding for proteins of known function. Bas4 does not translocate into host cell which might explain this observation.

Of interest was the relationship between the effector repertoire and the host chloroplast. Several effector proteins in rust fungus are known to target the chloroplast [207], studying the association between accumulation of effectors and sugar transporters may explain why effectors might target the chloroplast in *M. oryzae*. In this study, the KE002 genome was predicted to have several effectors that were predicted to target the chloroplast (possess a chloroplast transit signal sequence in the C-terminal). The reference genome 70-15 contains more than 42 effector proteins predicted to target the chloroplast while in KE002 genome, 15 effector proteins predicted to contain this transit signal sequence were highly up-regulated (Figure 3.11). Further studies of subcellular localisation of these effectors will shade light on this subset of effectors and their role in sugar uptake or chloroplastic-related immune response suppression.

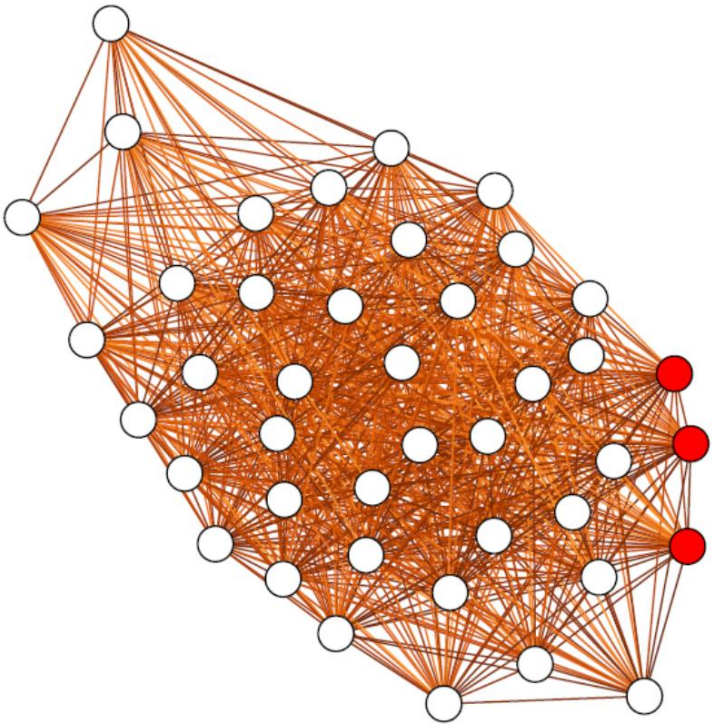

Figure 3.10 *M. oryzae* isolate KE002 transcription regulome enriched with co-regulated genes encoding for secreted proteins and effectors, during infection on a susceptible rice cultivar.

A generated regulome of genes co-expressed during infection of KE002 on a susceptible rice cultivar, *Moukoto*. Each node represents a single gene, while each connecting line represents the transcription relationship between two genes. The regulome contains genes with correlated transcription (correlation value R>0.85), and contains 42 genes co-expressed with *BAS4* and two putative effectors *MGG_08482* and *KE002_15475* shown as red nodes. *MGG_08482* and *KE002_15475* showed a higher correlation with well annotated genes including sugar transporter STL1, salicylate hydroxylase, cutinase, sucrose-6-phosphate hydrolase, gluconolactonase and a polyketide synthase. Genes co-expressed with effectors might encode for proteins involved in nutrient up-take or involved in effector translocation mechanism during the biotrophic phase. Moreover, some *M. oryzae* might be involved in nutrient up-take. Figure courtesy of Dr. Ryan Ames.

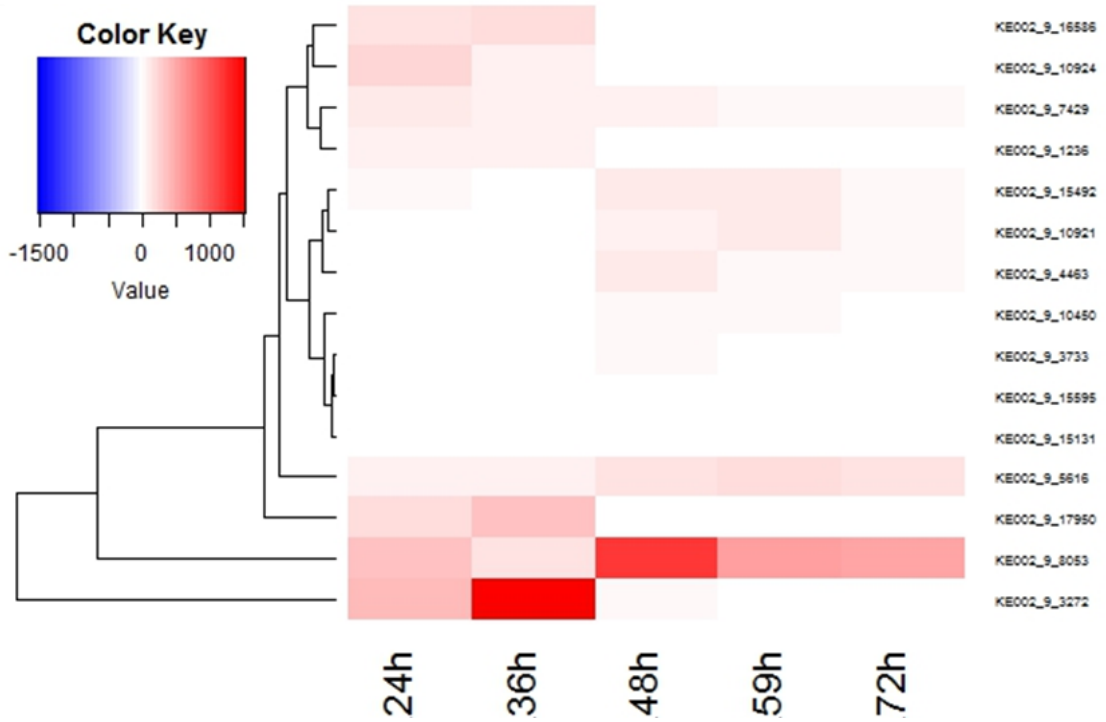

Figure 3.11 Heatmap showing relative levels of transcript abundance of KE002 genes predicted to encode for chloroplast targeting proteins.

Levels of expression are calculated using logarithmic fold change during KE002 infection of *Moukoto* compared to KE002 mycelial gene expression. Blue = down-regulated, red = up-regulated.

3.3.4.3 Analysis of gene expression patterns of putative effector-target genes in the rice genome

Nucleotide-binding and leucine repeat domain (NLR) immune receptors contain an NB-ARC domain (nucleotide-binding adaptor shared by APAF-1, R proteins and CED-4) and a C- terminal, leucine-rich repeat (LRR) domain (Figure 3.12). NLR architecture allows for flexibility in their structure and also enables insertion of additional domains. In rare cases, several integrated domains can occur in a single NLR [208]. This appears to allow them a broader range of pathogen recognition. Recognition of effectors can be achieved through direct interaction or indirect interaction (monitoring physiological changes caused by effectors), (Figure 3.12) [77]. Recent studies have provided growing evidence of direct interaction between effectors and NLRs [209].

For example, *M. oryzae* effector interaction with Pik-1, Rga5 and Pita [74, 85] have been well characterised [83, 84, 132, 210]. In case of Pik-1 and Rga5, there is an additional integrated domain that mediates pathogen recognition through binding of *M. oryzae* effectors AVR-Pik and AVR-Pia or AVR-CO39 respectively. The resistant gene *Pi-CO39/Pia* encodes a pair of NLR Rga4/Rga5 of which Rga5 directly interacts with either AVR-CO39 or AVR-Pia secreted by the rice blast fungus (Figure 3.12). Direct binding of Rga5 to *M. oryzae* secreted effectors AVR-CO39 or AVR-Pia causes activation of cell death and immune response signalling [83]. These effectors are recognised after binding to the ATX1 (RATX1) domain that is present in the C-terminus of Rga5 and is similar to a heavy-metal associated (HMA) domain from *Saccharomyces cerevisiae*. Another rice protein Pi21, is a HMA containing protein, and is a

susceptibility factor that slows the plant's defense against rice blast [86]. However, an allele of *pi21* in some Japonica rice lines confers resistance against blast characterised as a recessive resistance gene [86].

In indirect interactions, immune receptors will monitor changes on "guardees" – proteins that have a role in immunity or "decoys" if they mimic the host target, an event that culminates in activation of the NLR response [77]. A good example is the RPM1 interacting protein 4 (RIN4) that is kept in check by two *Arabidopsis* NLR, RPM1 and RPS2. Cleavage or phosphorylation resulting from interaction between RIN4 and bacterial effectors AVRRpt2 or AVRRpm1 is detected by RPS2 and RPM1 respectively [211-213]. Two effectors, PopP2 and AVRRps4, from the wilt pathogen *Ralstonia solanacearum* and *Pseudomonas syringae,* respectively, are recognised through the interaction or modification of WRKY DNA-binding domain of the RRS1 protein [214, 215]. Further studies have indicated that both effectors can interact with several WRKY transcription factors, suggesting that this is a major host effector target [214, 215].

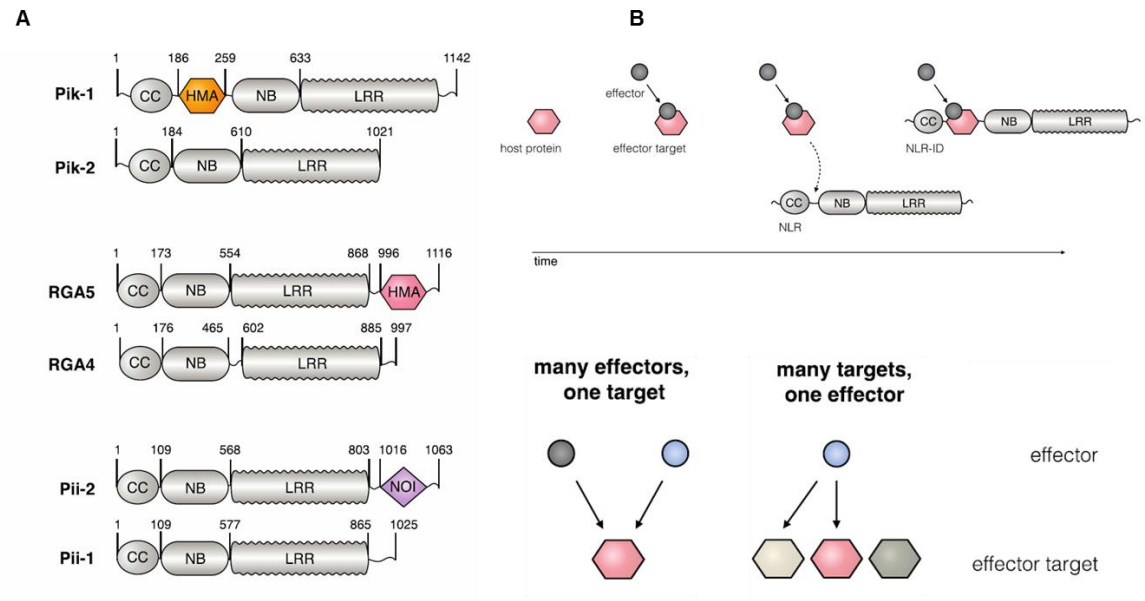

Figure 3.12 Rice NLRs contained integrated domains.

(A) Pik, Pia (Rga5/Rga4) with integrated HMA (hexagon) and Pii with AVRRpt cleavage/NOI (diamond) domains integrated in rice NLR architecture. (B) Top right demonstrates evolution of an NLR to integrate a targeted host protein and becomes a bait for effector recognition. Unrelated or related effectors can target same plant protein or one effector can be associated with multiple host proteins. This figure was adapted from Bialas *et al* [77].

Studies of integrated decoy domains in plants have revealed that several effector host targets are integrated into plant immune receptors as shown in Table 3.7 and Figure 3.12 [77]. Sarris *et al* [209] screened plant genome sequences for diverse sensors/decoys/integrated domains in more than 40 plant genome sequences including those of 19 crop plants. They identified integrated domains that occurred in more than one plant family *(*Table 3.7). For example, a Jacalin domain is fused to NLRs in the genomes of 6 out of the 8 *Poaceae* (grass family) and might be an effector-target conserved in the grass family [209]. This means that the fusion event might have occurred before the diversification of this family considering the family diverged 70-55 million years ago [209]. Most recently, it has been reported that overexpressing a Jacalin-related lectin protein OsJAC1 in barley, wheat and rice conferred resistance against important fungal pathogens such as *M. oryzae* and *Blumeria graminis* [216]. It is possible that in susceptible rice cultivars, the OsJAC1 or its paralogs will be bound by *M. oryzae* effectors and suppressed in order to promote infection. On the other hand, in rice cultivars with OsJAC1 domain fused to NLRs or acting as a 'guardee', the domain will act as a molecular sensor to recognise *M. oryzae* secreted effectors that target OsJAC1 via direct binding, and initiate immune response. In a different example, a fusion of NLR and Exo-70 occurs in some wheat cultivars and barley but not in rice where recognition of an Exo-70 targeting effector, AVR-Pii, happens via separate proteins [209]. This suggests that the fusion event might have occurred only recently.

Zinc Finger-BED (ZBED) protein is for example, another characterised integrated domain [208]. Together with WRKY and protein kinase domains, these three domains have been described as integrated in many NLRs and in

different plant species [208, 209]. Kroj *et al* [208] demonstrated that, overexpressing *ZBED* in rice plants led to reduced susceptibility to *M. oryzae* infection and inoculated plants had fewer lesions [208]. *Δzbed* mutant showed more susceptibility to rice blast. Similar to OsJAC1, it is possible that in susceptible rice cultivars, the ZBED proteins will be bound by *M. oryzae* effectors and suppressed in order to promote infection. On the other hand, in rice cultivars with ZBED domain fused to NLRs or acting as a 'guardee', the domain will act as a molecular sensor to recognise *M. oryzae* secreted ZBED-targeting effectors via direct binding, and initiate immune response.

From the integrated domain/decoy hypothesis, it is now thought that plant proteins with homology to integrated domains are targeted by pathogen secreted effector proteins to either promote susceptibility or resistance, (see Figure 3.12B) [208]. I analysed the expression profile of three genes (high-lighted in Table 3.7), encoding for rice proteins characterised as integrated domains and which have been associated with host susceptibility/resistance in three different studies. Moreover, the three integrated proteins were found to be fused to grass family NLRs including barley, wheat and rice [208, 209]. Overexpression of *ZBED* and *OsJAC1* in susceptible rice cultivars for example, leads to increased resistance to *M. oryzae* infection [208]. Rice HMA domain containing proteins have also been shown to directly bind *M. oryzae* effectors. The function of HMA domain containing proteins in plants has not been reported, but it is hypothesised that these proteins are bound by *M. oryzae* effectors to enhance disease susceptibility [209]. The MAX (*Magnaporthe* AVRs and ToxB like) effectors that target HMA-domain containing proteins, are estimated to be 5-10% of the effector repertoire of *M. oryzae* genome and 50%

of cloned avirulence genes belong to this expanded gene family [210]. Despite lacking protein sequencing similarity, these effectors possess a six-stranded beta sandwich stabilized by similarly positioned cysteine bonds [210].

I reasoned that, the susceptible cultivar *Moukoto* used in this study, lacks cognate R-genes specific to effector proteins secreted by KE002. However, this cultivar might possess effector-targeted host proteins that are not fused to NLRs or involved in pathogen recognition, but function as separate proteins in rice cell. During infection for example, up-regulated effector genes will encode for MAX proteins that bind HMA domain containing proteins to promote susceptibility. Additionally, some of the effectors will bind and inhibit ZBED containing proteins or OsJAC1 to supress immunity. I hypothesised that, in this susceptible cultivar, genes encoding for ZBED or OsJAC1 domain containing proteins, that have been shown to enhance resistance to *M. oryzae* will be down-regulated or not expressed, while genes encoding HMA domain containing proteins thought to enhance susceptibility will be highly up-regulated.

Understanding the transcription patterns of genes that encode for HMA, ZBED or OsJAC1 containing proteins during infection might provide an insight into their function in rice during a compartible reaction. This, thereby can contribute to effector target screening experiments using Yeast-two-hybrid analysis and co-immunoprecipitation in the future. In a co-immunoprecipitation screen to determine rice proteins targeted by effectors identified in this study like *MEP15*, *MEP13* or *MEP14*, if HMA, ZBED or OsJAC1 containing proteins produce high scores from mass spectrometry analysis, they can for example be

prioritised during validation. [208]. Ultimately, these integrated domains/decoys can also be used to clone new effectors and define host new processes targeted by effectors [208].

Table 3.7 The most prevalent integrated domains in NLRs of grass family. Table from Sarris *et al* [209]

Integrated domain	Family/Species	Description
Protein Kinase	*O. sativa*	Protein Kinase
DUF3542	*O. sativa*	Protein of unknown function
Protein tyrosine kinase	*O. sativa*	Protein tyrosine kinase
WRKY	*Poaceae*	WRKY DNA-binding domain
WD40	*O. sativa*	WD40 domain, G-beta repeat
Zf-BE	*O. sativa*	BED zinc finger
B3	*O. sativa*	B3 DNA-binding domain
DUF761	*O. sativa*	Cotton fiber-expressed protein
HMA	*O. sativa*	Heavy metal-associated domain
Thioredoxin	*O. sativa*	Thioredoxin
VQ	*O. sativa*	VQ motif
Zf-RVT	*O. sativa*	Zinc-binding in reverse transcriptase
C1_2	*O. sativa*	C1 domain
Jacalin	*O. sativa*	Jacalin-like lectin domain
FNIP	*Poaceae*	FNIP repeat
Kelch_1	*Poaceae*	Kelch motif
PP2C	*Poaceae*	Protein phosphatase 2C
Cleavage site for type III effectors	*O. sativa*	AVRRpt-cleavage
PP2	*Poaceae*	Phloem protein 2
UBN2_3	*Poaceae*	Gag-polypetide of LTR copiatype
PAH	*Poaceae*	Paired amphipathic helix repeat
PARP	*Poaceae*	Poly(ADP-ribose) polymerase catalytic domain

XH domain	*Poaceae*	XH domain
Zf-CCHC_4	*Poaceae*	Zinc knuckle
Zf-RING_2	*Poaceae*	Ringer finger domain
Glutaredoxin	*Poaceae*	Glutaredoxin
Abhydrolase_6	*Poaceae*	Alpha/beta hydrolase family

From this analysis, I identified seven genes encoding for HMA domain containing proteins that showed higher expression in un-infected control *Moukoto* leaf tissue but down-regulated in infected leaf tissue. However, two genes encoding for HMA domain containing proteins did show differential expression and were highly expressed in infected leaf tissue compared to control, un-infected leaf tissue from 24 h post-infection (Figure 3.13). Several MAX effectors, including *MEP3, MEP13* and *MEP15* discussed in chapter 4, were among the most highly expressed genes during KE002 infection of *Moukoto.* HMA domain containing proteins that are highly expressed might be involved in disease development, and may therefore be potential targets for *MEP3, MEP13* and *MEP15* in order to promotes susceptibility. This can be investigated more using yeast-two-hybrid.

Of interest was the expression of *ZBED*, the expression of which has been shown to have a correlation with partial resistance to rice blast infections [208]. A total of 26 genes encoding for ZBED domain-containing proteins isoforms showed differential regulation patterns during *M. oryzae* infection on rice cultivar *Moukoto*. Genes encoding for 8 isoforms showed increased transcription from 24 h and had peak expression from 48-72 h (Figure 3.14). At this stage of infection, the rice blast fungus has switched to necrotic growth and we could not correlate the late expression of *ZBED* and susceptibility to *M.*

oryzae infection. Overexpressing *ZBED* in a susceptible rice cultivar *Nipponbare,* has been reported to lead to reduced susceptibility in a previous study [208]. I conclude that from this study, secreted effectors by *M. oryzae* isolate KE002, might therefore bind and inhibit the activity of the highly expressed isoforms of ZBED domain-containing proteins to enhance susceptibility. The mechanism involved by ZBED domain-containing proteins in resistance against *M. oryzae* is still unknown [208]. Although no effectors have been reported to interact with this protein, in future, ZBED domain-containing proteins may be a potential target for effector characterisation studies.

I found 6 isoforms of Jacalin domain protein-encoding genes in *Moukoto* genome that show drastic down-regulation during infection but up-regulated in un-infected control rice leaf tissue (Figure 3.15). The expression data suggests that Jacalin-like-lectin transcription was down-regulated during infection leading to disease development. This is consistent with the studies suggesting that low expression of *OsJAC1* in barley, wheat and rice lead to increased susceptibility towards important fungal pathogens such as *M. oryzae* and *Blumeria graminis* [216]. It is possible that suppression of Jacalin-like-lectin transcription during infection by *M. oryzae* promotes infection. However we could not determine if this was due to interaction between *M. oryzae* secreted effector proteins and encoded OsJAC1 protein. No studies have been carried out on interaction between *M. oryzae* effector proteins and rice Jacalin-like-lectin proteins. This needs more investigation.

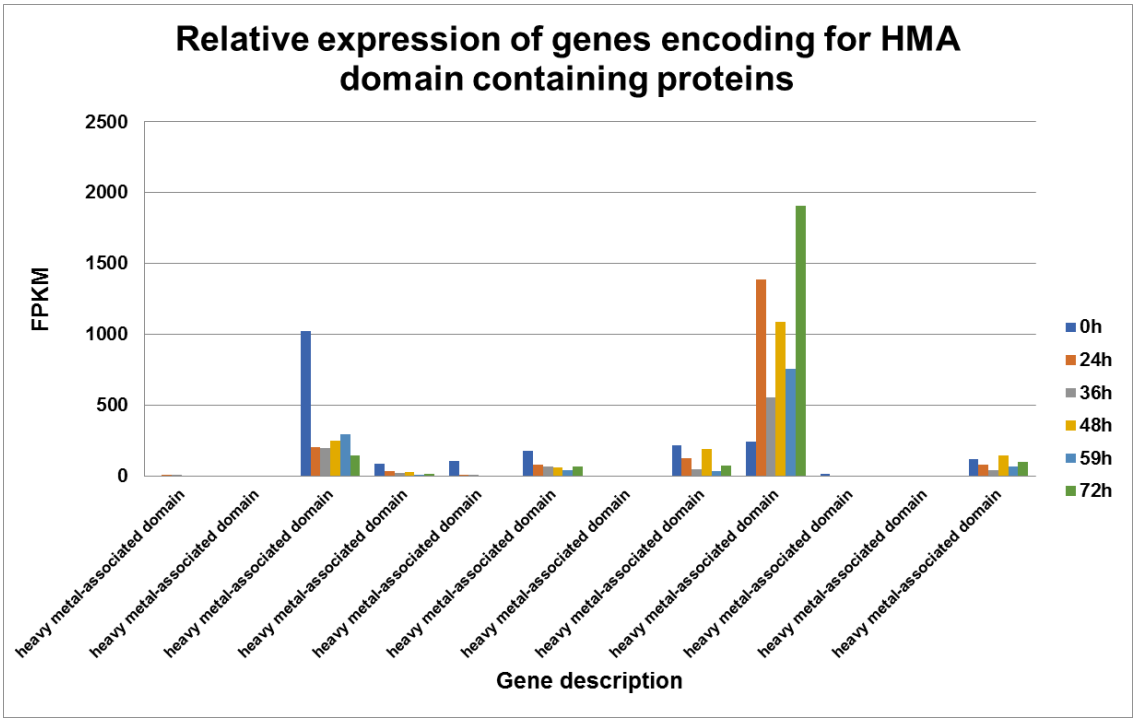

Figure 3.13 Expression at transcript-level of HMA domain containing proteins during infection of susceptible *Moukoto* rice line by KE002.

Fragments per Kilobase of transcript per million mapped reads (FPKM) values were generated using cuffdiff algorithm. FPKM values are shown on the Y-axis while HMA domains containing proteins encoding gene isoforms are shown on the X –axis. Expression at 24, 36, 48, 59 and 72 h post-infection are represented by different coloured bars.

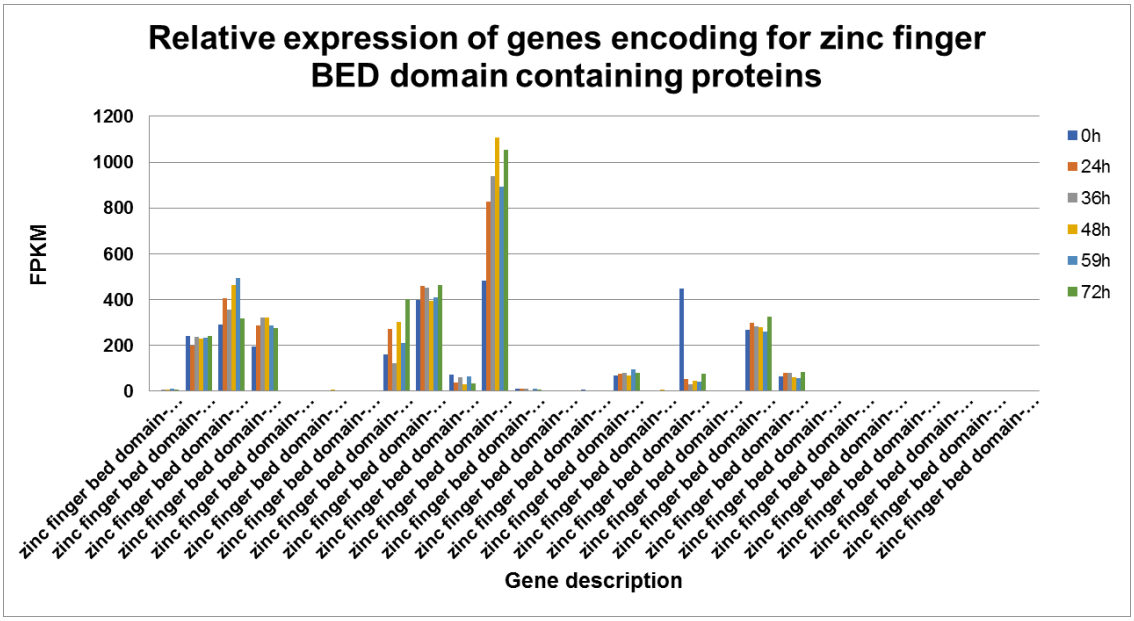

Figure 3.14 Expression at transcript-level of zinc-finger BED domain containing proteins during infection of susceptible *Moukoto* rice line by KE002.

Fragments per Kilobase of transcript per million mapped reads (FPKM) values were generated using cuffdiff algorithm. FPKM values are shown on the Y-axis while zinc-finger BED domain containing proteins encoding gene isoforms are shown on the X – axis. Expression at 24, 36, 48, 59 and 72 h post-infection are represented by different coloured bars.

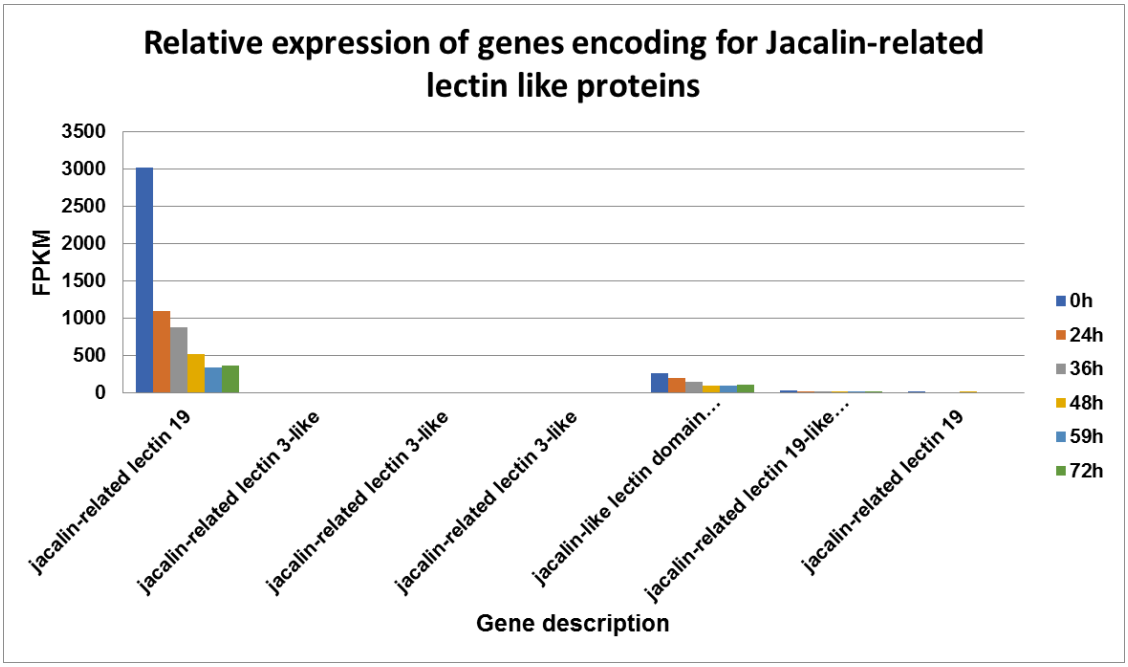

Figure 3.15 Expression at transcript-level of jacalin-related lectin proteins during infection of susceptible *Moukoto* rice line by KE002.

Fragments per Kilobase of transcript per million mapped reads (FPKM) values were generated using cuffdiff algorithm. FPKM values are shown on the Y-axis while jacalin-related lectin protein encoding gene isoforms are shown on the X –axis. Expression at 24, 36, 48, 59 and 72 h post-infection are represented by different coloured bars.

3.4 Discussion

To understand the relationship between effector repertoire of *M. oryzae* and their potential recognition as avirulence determinants in blast populations in Sub-Saharan Africa, we sequenced the genomes of 23 *M. oryzae* isolates from this region using Illumina. We also used Pacbio long-reads sequencing to further improve genome assemblies of Guy11 and KE002. When compared to assemblies generated from Illumina short-reads, the Pacbio assembly produced larger genomes, 43.3 Mb and 45 Mb for Guy11 and KE002 respectively, with improved N50 for both Guy11 (2.3 Mb) and KE002 (4.6 Mb). From these assemblies, we could predict more genes in long-read-generated genome assembly than in the short-read assembled genomes. Further analysis using BLASTn showed that extra genes that were predicted from long-read-assembled genomes are not actual additional genes but rather they are families undergoing duplication and gene expansion.

Two effector genes *PWL2* and *MEP13* are among genes in this study found to be undergoing genome expansion. This observation is consistent with several studies that have shown that effector genes are prone to genome translocation, duplication and deletion [217, 218]. We concluded that gene expansion events are common in *M. oryzae*, but cannot be well studied using data generated from short-read sequencing. Importantly, the observed increase genome size may be due successful sequencing and assembly of highly repeated regions that could not be assembled from short-read sequencing. Using optical mapping, we showed that the genome assembly of Guy11 could for example be further improved into a contiguous genome constituted in 28

contigs. Genome alignment to the reference genome 70-15 revealed high relatedness between these genomes. This is consistent with Chiapello *et al* [44] study that showed that only 2 % of 70-15 genome belongs to weeping lovegrass pathogen.

Unlike other well characterised protein families, most effectors lack functionally characterised domains or homologs in closely related eukaryotic or fungal species [188]. This makes it challenging to carry out *de novo* gene prediction for effector genes in fungi from genome sequences [182]. In this study, RNA-seq analysis of rice blast infections was incorporated into a gene prediction pipeline to improve prediction of effector/avirulence gene candidates. Using a combination of long-read sequencing and RNA-seq it was possible to accurately predict all cloned AVR genes occurring in KE002. Moreover, the number of predicted secreted protein-encoding genes was significantly increased when long-read sequencing was used. This study has also used an effector prediction program, *EffectorP*, to improve the effector annotation process. The program was able to differentiate putative effector genes from non-effector genes, making it a valuable tool in effector biology. Three novel effector genes identified using *EffectorP* are discussed in chapter 4.

Using RNA-seq, it was possible to gain new insights on fungal and rice gene expression during the biotrophic phase of invasion. First, all cloned avirulence genes were not highly expressed in this analysis. However, biotrophy associated protein-encoding genes like *BAS1, BAS4* and *BAS107* were highly expressed. Newly identified biotrophy effectors in our lab (Yan *et al* unpublished) *MEP3* (MGG_17249) and *MEP12* (MGG_10276) were for

example also highly upregulated. This is consistent with a study from Mosquera *et al* [1] that facilitated identification of biotrophy-associated secreted proteins 1 to 4. *MEP3* is a MAX domain-containing effector protein and was among two other reported MAX domain containing effectors (MGG_00043 and MGG_08482) that were upregulated during infection. Two MAX effector genes identified in this study *MEP13* and *MEP15* were also up-regulated during infection. The up-regulation of effectors thought to target similar host proteins is consistent with studies suggesting that effectors may carry out redundant functions during host colonisation [113, 219]. Effector redundancy is thought to be a result of the arms-race between the pathogen and its host [217, 220]. This phenomenon gives the pathogen robustness when faced with a rapidly changing environment [217, 219, 220].

It came as a surprise that most cloned effectors and avirulence genes *PWL2, BAS1, BAS2, BAS3, SLP1, AVR-Pizt, AVR-Pia* and *AVR-Pi9* in the KE002 genome did not show up-regulation during infection. These data suggests that this *M. oryzae* strain might regulate gene expression according to host signals. *Moukoto* is a highly susceptible rice cultivar and the pathogen might choose to reduce the expression of host immune suppressing genes. Another explanation maybe due to low abundance of infected rice cells at the time point of optimal effector expression (24-36 h). A mixture of fungal transcriptomes consisting of conidium, germ tube, appressorium and invasive hyphae might mask transcriptomic changes of lowly expressed effectors. From the rice genome, among the reported integrated domain proteins analysed, all had multiple isoforms that had variable expression patterns. Several cases of effector redundancy have been reported and can be explained by effectors that

target the same host proteins, for example MAX domain containing effectors that are abundant in the rice blast fungus genome as shown in Figure 3.12 [77, 210]. However, it is not clear if effector target proteins have evolved to gain genetic redundancy. Some commonly targeted host proteins might have several isoforms to counter rapidly evolving effectors or changes resulting from effector manipulation [77].

Chapter 4 Genome analysis of rice blast fungus *Magnaporthe oryzae* field isolates from Sub-Saharan Africa

4.1 Introduction

A number of genomes from plant pathogenic fungi have been sequenced since the *M. oryzae* genome was first sequenced in 2005 [24]. The genomes of non-disease causing fungal endophytes have also been sequenced including obligate biotrophs such as arbuscular mycorrhiza fungus. This, together with the increase in available genomes from bacteria, oomycetes and parasites has provided a good platform to study fungal pathogenesis, plant-pathogen interactions, evolution and co-evolution using informatics [221-223]. Additionally, the availability of full genome sequences, reference genomes, and gene/ protein databases has made identification on novel genes easier and quicker, especially for proteins that possess conserved domains [24, 89].

Fungal pathogens are known to secrete a large repertoire of secreted effector proteins which play a major role during fungal-plant interactions and are likely to influence fungi lifestyle [89]. According to Presti *et al* [89], fungal species have a varying number of secreted proteins in their genomes, depending on their lifestyles [89]. Pathogens with hemi-biotrophic lifestyles possess the highest number of secreted proteins in their genomes, which are understandably needed during both biotrophic and necrotic phases of infection [89]. During infection of *Arabidopsis thaliana* by the hemibiotroph *Colletotrichum higginsianum*, the biotrophic phase is predominantly characterised by up-regulated expression of genes encoding for secreted effector proteins, while the

necrotic phase is characterised by increased expression of secreted lytic enzymes and plant cell wall-degrading enzymes [92]. Obligate biotrophs, like hemibiotrophs, will have a high number of secreted effector proteins to maintain intimate interaction with their hosts [89]. Obligate and hemibiotrophic fungi insert their feeding structures in form of invasive hyphae or haustoria into host cells, or grow intracellularly and this requires the host immune system be suppressed [224]. Secreted effectors help to inhibit, modify, alter or modulate activities in host cells to the benefit of the proliferating pathogen [89].

Bioinformatics analysis on effector identity has suggested that they mostly lack or have weak similarity to characterised proteins. However, most recent studies on secondary structures of effector proteins have revealed unexpected similarity despite nucleotide sequence diversity. Well studied examples include conserved folds in MAX and WY domain containing effectors of fungi and oomycetes respectively [210]. Despite having the same host targets, these effectors possess plasticity that allows them to bind different proteins or evolve to acquire different activities [85]. *M. oryzae* possess multiple MAX domain-containing effectors [210]. These types of effectors may have redundant activities and several will probably target the same host pathway [77, 84]. This suggests that deletion or loss of function of one member of the family may not affect the rest [210]. Deletion of one or two MAX domain-containing effector genes for functional analysis might not show any virulence phenotype unless they are avirulence genes (Dr Xia Yan and N.J. Talbot, unpublished observations). The *Cladosporium fulvum* effector AVR2 and *P. infestans* effectors EPIC1 and EPIC2 are known to target host proteases even though these effectors lack any sequence similarity [225]. AVR2 has no known

similarity to other proteins while the two EPIC family effectors have similarity to the cystatin domain [113]. This means that, despite the lack of sequence similarity, these effectors seem to have evolved to target the same host proteases [114]. Early studies to identify effectors (avirulence genes) was through observing incompatible reactions elicited in a gene-for- gene interaction between the host and pathogen [125, 226]. One of the first characterised avirulence genes in *M. oryzae*, *PWL2* was cloned from a genetic cross between two *M. oryzae* parental strains that had varying virulence on weeping love grass [117, 118]. *PWL2* segregated among non-pathogenic progeny and a genetic complementation of pathogenic strains on weeping lovegrass transformed them into non-pathogenic strains [117, 118]. However, identification of effectors that lack cognate resistance (R) gene can be laborious and time consuming.

Resistance against rice blast is controlled by one or several resistance genes (R-genes) [133, 227]. While several major resistance genes have been cloned, a lot needs to be done to identify the most reliable and most effective R-gene. Furthermore, previously cloned R genes have been shown to be only effective in specific regions and cannot be deployed across the globe [228, 229]. The interaction between *M. oryzae* and its host R-genes has been associated with frequencies of resistance gene breakdown which renders deployment of one R-gene not a durable solution [228, 229]. The best strategy to deal with breakdown of host resistance is to pyramid several R-genes in local adapted rice cultivars. First an extensive analysis of a rice blast population against these resistance genes can be used to determine the most effective R-gene combination before deployment. The International Rice Research Institute (IRRI) has developed a panel of rice monogenic lines carrying 24 known R-

genes in the background of a susceptible Japonica line Lijiangxintaunheigu (LTH) [230] that can be used in such studies.

As part of this study, in order to determine the suitable R-genes to be introgressed in locally grown rice varieties cultivated in Sub-Saharan Africa (SSA), the virulence of isolates collected from this region was characterised [231]. Rice monogenic lines carrying known R-genes were screened for disease response against a rice blast population from Sub-Saharan Africa [231]. We have identified *Pi9* as a potentially suitable gene to be used for pyramiding to achieve durable resistance in this region [231]. In a new environment, pathogens evolve in order to successfully colonise the host and improve their fitness [228]. The same will happen if a new R-gene is introduced into a given population of the rice blast fungus [228]. These events are caused by high rate of mutations occurring in the fungal genome especially in its effector repertoire [228, 232]. In addition to understanding the virulence spectrum of selected rice blast pathogens against known resistance gene, a genome-wide study to understand the effector/avirulence repertoire of these isolates will help explain the relationship between occurrence of these genes and virulence [231]. More importantly, quicker and accurate identification of novel effectors/avirulence genes will help understand the dialogue between the fungus and its host and functionally characterise these effectors [231].

In this chapter, we set out to identify and characterise novel genes that encode for effector and avirulence proteins in rice blast isolates from Sub-Saharan Africa. To understand the relationship between the effector repertoire of *M. oryzae* and their potential recognition as avirulence determinants in this

population, the genomes of 23 rice blast isolates which had their virulence classified using rice monogenic lines differing in 24 major rice blast resistance genes, were sequenced. First, the correlation between virulence of these isolates on monogenic rice lines and the occurrence of known avirulence genes was analysed. We then determined presence/absence polymorphisms of predicted genes in all sequenced isolates. Secondly, a standalone program KSNP3 was used to infer relatedness of all sequenced isolates including selected control isolates. Thirdly, putative effector protein-encoding genes predicted in Chapter 3 were further analysed. Three of these were confirmed as biotrophy-associated effectors that accumulated in the BIC when analysed by live-cell imaging. To determine their role during biotrophic colonisation, targeted gene deletion and genetic complementation studies were carried out.

4.2 Materials and methods

4.2.1 General material and methods

For standard procedures used in this chapter see Chapter 2

4.2.2 BLAST analysis

BLAST searches were generated using BLAST 2.2.22 program. A FASTA formatted text file containing all cloned effectors/avirulence gene coding sequence nucleotide sequences was used to query a database represented by an isolate's genome sequences using a standalone BLASTn (e-value 1-5). Hits with identities ranging from 50-80%, 80-100% and 100% were defined as divergent, similar and identical respectively. Values below 50% were considered as not similar. Nucleotide sequences *PWL1* (U36923.1), *PWL2* (U26313.1), *PWL3* (1045533), *PWL4* (1045535), *AVR-Pita* (12642087), *AVR-CO39* (27450408), *AVR-ACE1* (47109413), *AVR-Piz-t* (194293523), *AVR-Pia* (237858322), *AVR-Pii* (237858324), *AVR-Pik* (237858326), *AVR-Pi9* (KM004023.1), *AVR-Pib* (KM887844.1) and *AVR-Pi54* (HF545677.2) were downloaded from NCBI.

4.2.3 Inoculation and disease score

Each genotype was represented by a single cell of a seed planting tray each with 5 growing seedlings. 3 weeks old plants were inoculated with fungal conidial suspension diluted to a final concentration of 5 $x10^4$ conidia mL-1 using spray method. Disease response and susceptibility was scored using the IRRI disease analysis scale of 0-9 standardised visual score where 0 represents lack

of any visible response, 1-3 representing hypersensitive response, 4-6 as varying degrees of susceptibility and 7-9 representing high susceptibility.

4.3 Results.

4.3.1 Virulence of Sub-Saharan Africa isolates on rice monogenic lines

A total of 122 *M. oryzae* isolates collected from Sub-Saharan Africa were screened against rice blast monogenic rice lines to predict for occurrence of virulence determinant genes in each isolate. The collection consisted of *M. oryzae* isolates from nine African countries, including Benin, Burkina Faso, Mali, Ghana, Kenya, Nigeria, Tanzania, Togo, and Uganda. Part of this work was carried out in collaboration with Professor James Correll's Laboratory at the University of Arkansas. From virulence spectrum we observed, 23 isolates were selected for sequencing using the Illumina platform. Isolates in this study were from different origins; 66% from West Africa, 34% from East Africa, 1 isolate from USA, 1 from Egypt and Guy11 from French Guyana was included as a control. Their virulence was assessed by inoculation on the IRRI-bred blast resistance lines (IRBL) or monogenic rice cultivars under controlled conditions. Rice genotypes used in this study are monogenic lines carrying 24 different R-genes in a susceptible background of the Japonica cultivar called Lijiangxituanheigu, see details in Table 4.1 [230].

Table 4.1 Description of rice cultivars used to screen for disease response against rice blast isolates

Rice genotype	Description	R-gene
IRBL 12-M	Monogenic Line	*Pi*12(t)
IRBL 19-A	Monogenic Line	*Pi*19
IRBL 1-CL	Monogenic Line	*Pi*1
IRBL20-IR24	Monogenic Line	*Pi*20(t)
IRBL 3-CP4	Monogenic Line	*Pi*3
IRBL 5-M	Monogenic Line	*Pi*5(t)
IRBL 7-M	Monogenic Line	*Pi*7(t)
IRBL 9-W	Monogenic Line	*Pi*9
IRBLA-a	Monogenic Line	*Pia*
IRBLB-B	Monogenic Line	*Pib*
IRBLI-F5	Monogenic Line	*Pii*
IRBLKH-K3	Monogenic Line	*Pik-h*
IRBLK-KA	Monogenic Line	*Pik*
IRBLKM TS	Monogenic Line	*Pik-m*
IRBLKP-K60	Monogenic Line	*Pik-p*
IRBLKS-F5	Monogenic Line	*Pik-s*
IRBLKS-S	Monogenic Line	*Pik-s*
IRBLSH-B	Monogenic Line	*Pish*
IRBLSH-S	Monogenic Line	*Pish*
IRBLTA 2-PI	Monogenic Line	*Pita*2
IRBLTA 2-RE	Monogenic Line	*Pita*2
IRBLTA CP 1	Monogenic Line	*Pita*

IRBLTA CT2	Monogenic Line	*Pita*	
IRBLT-K59	Monogenic Line	*Pit*	
IRBLZ5-CA(R)	Monogenic Line	*Piz*-5	
IRBLZ5-CA	Monogenic Line	*Piz*-5	
IRBLZ-FU	Monogenic Line	*Piz*	
Toride 1		*Piz-t*	
IRBTP16211/Lijiangxituanheigu (LTH)	Japonica variety		NA
75-1-127	*Pi9* donor line	*Pi9*	*Pi*9

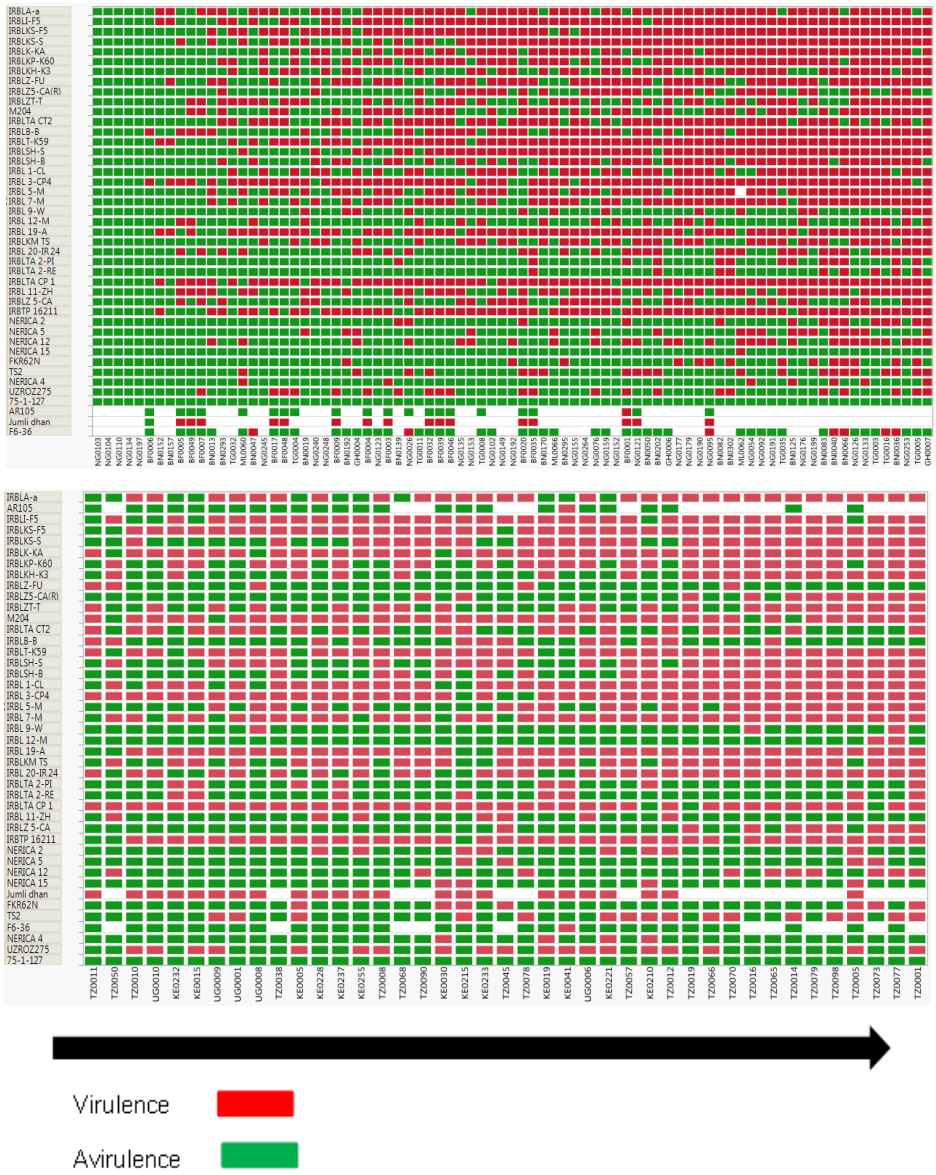

Figure 4.1 Virulence of Sub-Saharan Africa *M. oryzae* isolates on rice monogenic lines.

Rice genotypes shown on the Y-axis and field isolates on the X-axis. Red represents virulence and Green avirulence. Arrow indicates direction of increased virulence. Figure from Mutiga *et al* [231]. The upper panel shows virulence of West African isolates and the lower panel shows virulence of isolates from East Africa. NG0110 and NG0104 did show virulence on all the analysed rice genotypes.

Isolates

Rice Genotype	EG308	KE002	KE016	KE017	KE019	KE021	KE029	Guy 11
IRBLA-a (Pia)	S	R	R	R	R	R	R	S
IRBLI-F5 (Pii)	R	S	S	S	S	R	S	S
IRBLKS-S (Pik-s)	S	S	R	R	S	S	S	S
IRBLKP-K60 (Pik-p)	R	S	S	S	S	S	S	S
IRBLKH-K3 (Pik-h)	R	S	S	S	S	S	S	R
IRBLKM TS (Pik-m)	R	S	S	S	S	S	S	S
IRBLK-KA (Pik)	R	S	S	S	S	S	S	S
IRBLZ-FU (Piz)	R	R	R	R	S	R	R	R
IRBLZT-T (Piz-t)	R	R	R	R	R	R	R	R
IRBLB-B (Pib)	S	R	R	R	S	R	R	R
IRBLT-K59 (Pit)	R	R	S	R	R	R	S	S
IRBLSH-S (Pish)	S	S	R	S	R	R	S	S
IRBL 3-CP4 (Pi3)	R	S	S	S	S	S	S	S
IRBL 5-M (Pi5 (t))	R	S	R	S	S	S	R	S
IRBL 7-M (Pi7 (t))	R	S	S	S	S	S	S	S
IRBL 9-W (Pi9(t))	S	R	R	R	R	R	R	S
IRBL 12-M (Pi12(t))	R	R	S	R	S	R	R	S
IRBL 19-A (Pi19)	S	R	S	S	R	S	S	S
IRBL 20-IR 24 (Pi20(t))	R	R	R	R	R	R	S	S
IRBLTA 2-RE (Pita2)	S	S	S	S	S	S	S	S
IRBLTA CP 1 (Pita)	S	S	S	S	S	S	S	S
IRBL 11-ZH (Pi11(t))	R	R	S	S	R	R	R	R
IRBLZ5-CA (Piz-5)	R	R	R	R	R	R	R	S
IRBTP 16211 (LTH)	S	S	S	S	S	S	S	S

R Resistance

S Susceptible

Figure 4.2 Virulence of Sequenced isolates from Kenya, Egypt and Guy11 on rice monogenic lines.

Rice genotypes shown on the Y-axis and field isolates on the X-axis. Red represents virulence and Blue resistance. Guy11 is virulent on most tested monogenic rice lines. Figure from David Mwongera's thesis.

A total of 5 isolates from Nigeria were later found to be completely avirulent on all the tested monogenic lines as shown in Figure 4.1. Two of these strains NG0110 and NG0104 were selected for further analysis to determine the reasons for lack of pathogenicity or hypersensitive reactions in resistance reactions. To test if lack of virulence was related to the possession of avirulence genes, a PCR screen for presence or absence polymorphism was conducted. This screen identified presence of *AVR-Pik*, *AVR-Pi9*, *AVR-Pita* and *AVR-Piz-t* loci in NG0110 as shown in Figure 4.3. This ruled out the presence of avirulence genes as the reason for lack of pathogenicity on monogenic rice lines carrying *Pia* and *Pii* for example. Primers used for the experiment were provided by Dr. Bo Zhou (IRRI) are listed in Table 4.4. For more analysis, conidia from NG0110 and NG0104 were used to check for appressorium formation on hydrophobic surfaces or leaf sheath tissues.

Compared to the wild type strain Guy11, NG0110 and NG0104 were not able to form appressoria on either hydrophobic surfaces or on rice leaf sheath; this explained the lack of virulence on all tested rice cultivars as shown Figure 4.3. The isolates formed elongated germ tubes that did not differentiate into appressorium. All isolates in this study were collected from either rice leaf or neck lesions. To rule out the possibility of opportunistic infection, appressorium formation on hydrophobic surfaces was analysed with conidia from NG0110 mixed with that from a Guy11 expressing ToxA-GFP (Figure 4.3). Additionally, *Moukoto* rice leaf sheath was also co-inoculated with conidia from NG0110 mixed with that from a Guy11 expressing ToxA-GFP. Guy11 conidia could form appressoria and successfully colonise rice leaf sheath, unlike NG0110 that

formed elongated germ tube and was unable to form appressoria (data not shown).

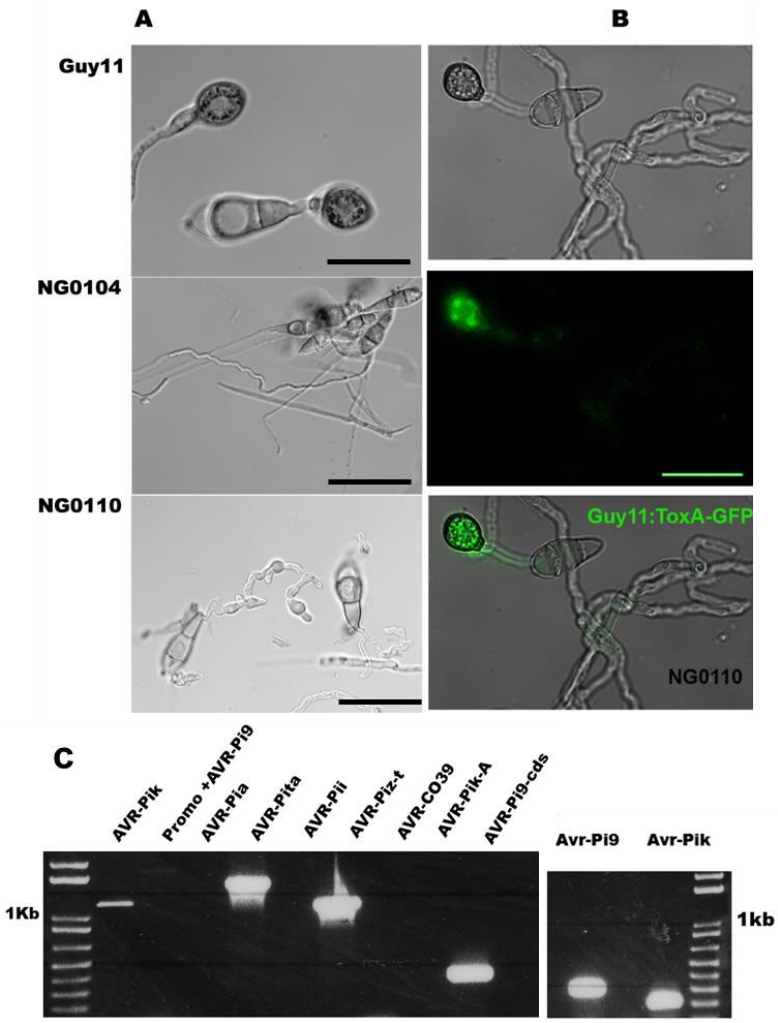

Figure 4.3 NG0110 lack of ability to form appressorium on a hydrophobic surface:

(A) Appressorium formation on a hydrophobic surface. The Guy11 was able to form appressoria while NG0110 and NG0104 did not form appressorium. (B) A mixture of conidial suspension containing Guy11 (expressing ToxA-GFP) and wild type NG0110. Guy11 ToxA-GFP could form appressoria while NG0110 could not. Scale bar represent 20 µm. (C) Presence absence PCR screen for different avirulence genes. Amplicons were separated by a 0.8% agarose gel electrophoresis. *AVR-Pi9*, *AVR-PikD*, *AVR-Pita*, *AVR-Piz*-t were amplified from NG0110 genomic DNA. *AVR-Pik* and *AVR-Pi9* were present in Guy11. 1 Kb plus size marker is shown on the left and right side of each gel image.

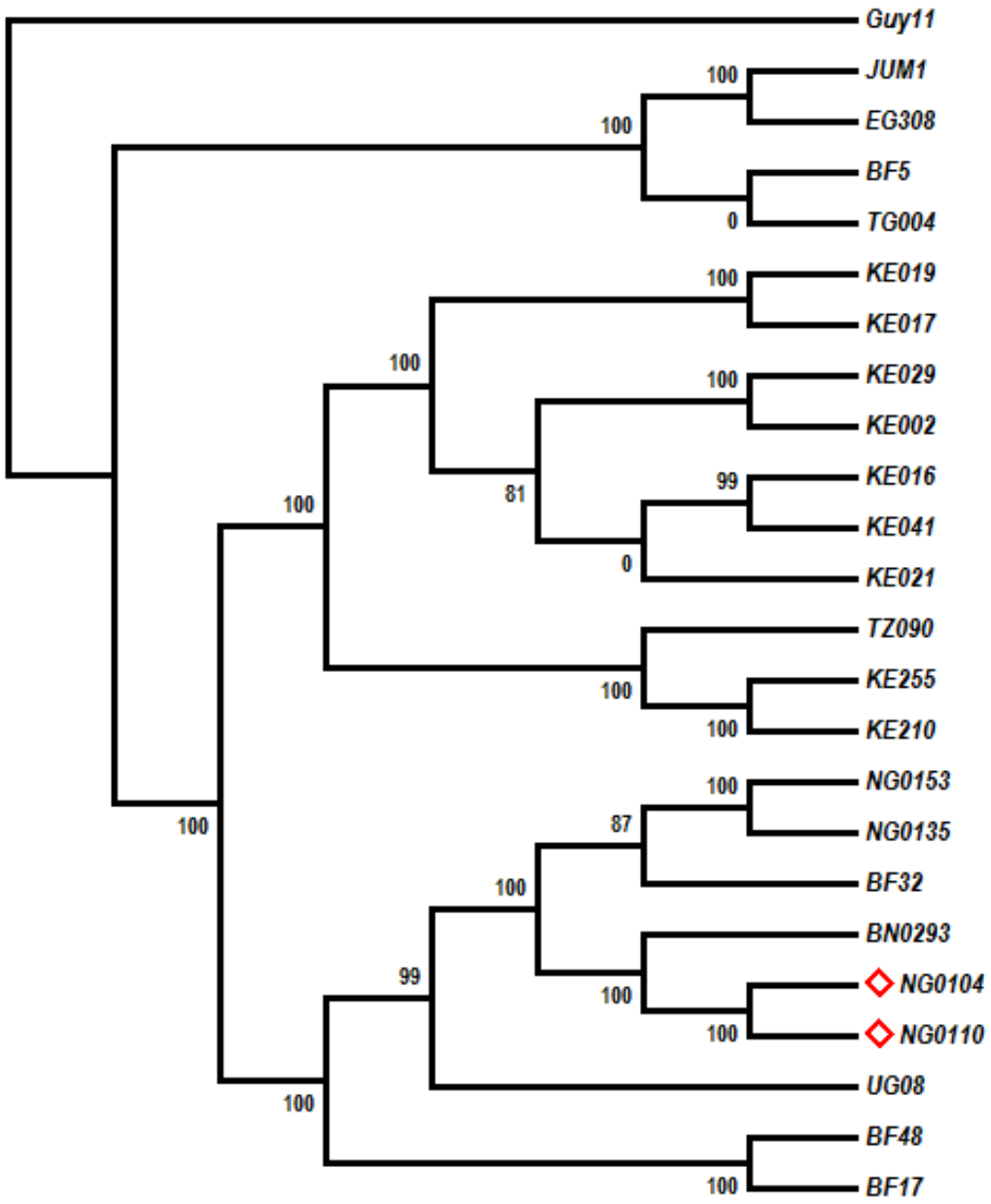

Figure 4.4 Allele count parsimony tree of *M. oryzae* field isolates from Sub-Saharan Africa.

Tree was generated using a standalone, whole genome multiple alignment/SNP call program called kSNP3. NG0110 and NG0104 are labelled with red diamonds.

These *M. oryzae* isolates were included in a subset of isolates that were selected for genome sequencing (Table 4.2). A maximum likelihood phylogenetic tree was generated to infer relatedness based on SNPs occurring in sequenced genomes. These two isolates clustered together and showed close relationship with BN0293 from Benin as shown in Figure 4.4. Gene calling program Augustus was used to predict genes from the genome sequences of these two isolates. Gene clustering program *Proteinortho* was used to align all predicted genes from NG0110 and NG0104 against genes predicted in Guy11 and the reference genome 70-15 transcripts. I reasoned that these two isolates may lack gene essential for appressorium formation or maturation. However, All genes reported to be involved in appressorium formation including *MAC1*, *PMK1*, *HOX7*, *WISH*, *PTH11*, *MGB1*, *RAC1*, *MST11*, *MST7*, *MST50*, *NIM1*, *NIME*, *CPKA/CPK2*, *MAGBA* and *OPDA* were found to be present in NG0110 and NG0104 and it was not possible to link non-appressorial forming phenotype to any clear genetic cause in terms of gene loss or mutation [33, 233-242]. The two isolates were disregarded from further avirulence gene mining process.

4.3.2 Whole genome sequencing infers genetic relatedness in rice blast isolates from Sub-Saharan Africa

To determine the relationship among sequenced *M. oryzae* isolates, a maximum likelihood tree was generated using kSNP3, a standalone program that aligns full genome sequences and identifies SNPs in either whole or a percentage of each of the aligned genomes. BF5 and BF48 TZ090 JUM1, BF17, BF32, BN0293, KE255, KE041, KE210, NG0135, NG0153, TG004 UG08 KE002, KE016, KE017, KE019, KE021, KE029 and EG308 were used in this analysis. A phylogenetic analysis was inferred by maximum likelihood based on SNPs present in 70% of each genome aligned against all 27 genomes sequences (n = 8 West Africa, n = 11 East Africa, n = 1 Egypt, n = 1 USA and n = 6 across the globe reference isolates) showed segregation into three major clades (Figure 4.5).

Isolates from West Africa segregated in one clade (including one East African isolate UG08) while those from East Africa clustered in a different clade. Isolates from Egypt and USA were segregated with Glhn3, P131, INA168, TH3, 76.3 and Guy11 from Asia and French Guyana (Figure 4.5**)**. There was evidence of close relatedness of sampled isolates in each region of Sub-Saharan Africa. There were major differences observed on virulence of isolates from each clade against rice monogenic lines. For this reason, I analysed, R-genes of which the cognate avirulence genes have been molecularly cloned were targeted (Figure 4.6).

Pi9 showed high level of resistance to the three clades, (n = 6 clade 1, n = 10 clade 2, n=0 clade 3). Other R-genes showed mixed resistance responses

against isolates from the three clades as follows *Pita* (n = 1 clade 1, n = 3 clade 2, n = 2 clade 3), *Pia* (n = 1 clade 1, n = 8 clade 2, n = 0 clade 3), *Pii* (n = 0 clade 1, n = 0 clade 2, n = 2 clade 3), *Pik* (n = 6 clade 1, n = 3 clade 2, n = 1 clade 3), *Pib* (n = 3 clade 1, n = 7 clade 2, n = 1 clade 3), *Piz-t* (n = 9 clade 1, n = 10 clade 2, n = 3 clade 3). Disease resistance from *Pik* alleles were as follows *Pik-s* (n = 1 clade 1, n = 0 clade 2, n = 0 clade 3), *Pik-p* (n = 4 clade 1, n = 0 clade 2, n = 1 clade 3), *Pik-h* (n = 5 clade 1, n = 1 clade 2, n = 1 clade 3), *Pik-m* (n = 6 clade 1, n = 0 clade 2, n = 1 clade 3).

Table 4.2 List of Isolates used in this study

Isolate	Country of origin	Description
JUM1	USA	Unknown
BF0005	Burkina Faso	Leaf blast
BF0017	Burkina Faso	Neck blast
BF0032	Burkina Faso	Leaf blast
BF0048	Burkina Faso	Neck blast
BN0293	Benin	Leaf
EG308	Egypt	Unknown
KE002	Kenya	Leaf
KE016	Kenya	Leaf
KE017	Kenya	Leaf
KE019	Kenya	Leaf
KE021	Kenya	Leaf
KE029	Kenya	Leaf
KE0041	Kenya	Leaf
KE0210	Kenya	Leaf
KE0255	Kenya	Leaf
NG0104	Nigeria	Leaf
NG0110	Nigeria	Leaf
NG0135	Nigeria	Leaf
NG0153	Nigeria	Leaf
TG0004	Togo	Leaf
TZ0090	Tanzania	Leaf
UG0008	Uganda	Unknown
Guy 11	French Guyana	Unknown

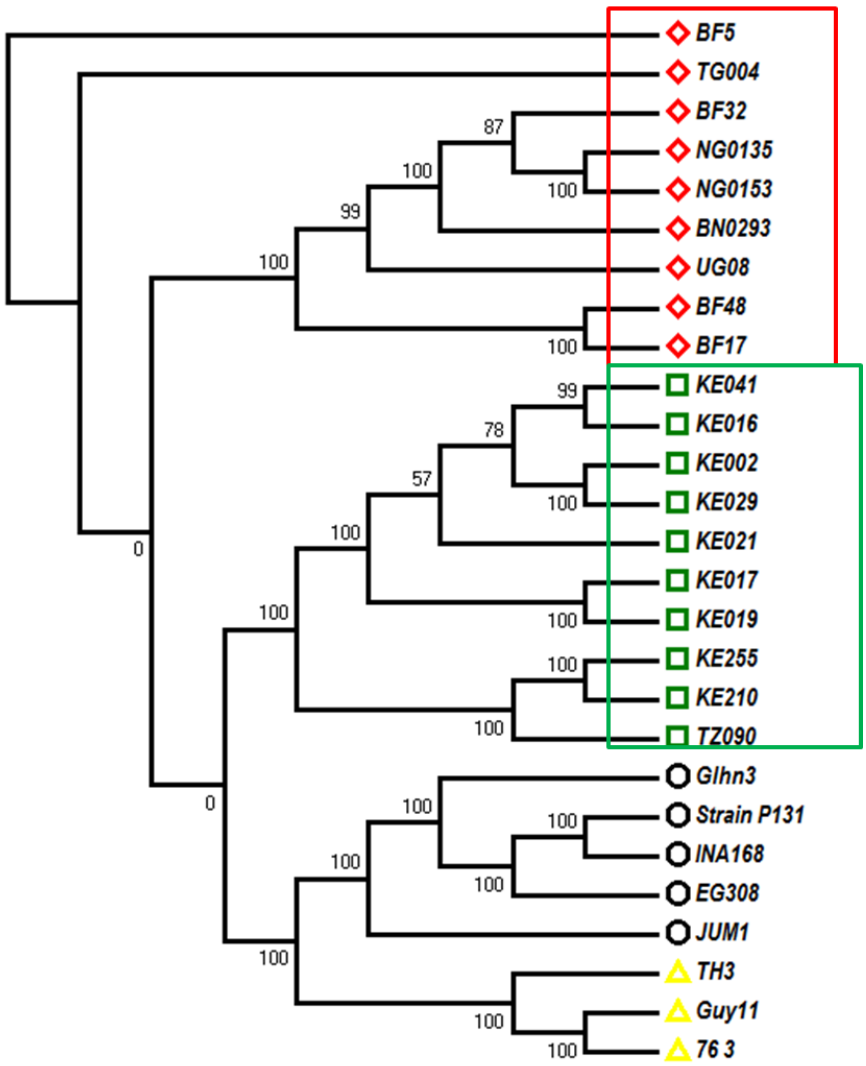

Figure 4.5 Maximum parsimony tree of *M. oryzae* field isolates from Sub-Saharan Africa in relation to selected reference isolates in the world

A maximum parsimony tree was generated using a standalone, whole genome multiple alignment/SNP call program called kSNP3. *M. oryzae* isolates were clustered into three major clades. All isolates from West Africa (shown by red diamond shapes) were clustered in clade 1 (labelled with red rectangle) while East African isolates (green squares) grouped in clade 2 (labelled with green rectangle). Reference isolates from Asia and other regions (black circles and yellow triangles) grouped in a different clade.

4.3.3 Genome analysis of virulence determinant-encoding factors in *M. oryzae* isolates from Sub-Saharan Africa.

4.3.3.1 Analysis of virulence determinant genes reveals most isolates harbour *AVR-Pi9* and *AVR-Piz-t*

The genome sequences of *M. oryzae* isolates were analysed for presence or absence polymorphism of all the known AVR/effectors as detailed in Figure 4.6. Strain NG0110 and NG0104 were not included in this analysis because of their inability to infect tested rice lines. As expected, *AVR-Pi9* occurs in all analysed isolates (n = 22) and correlated completely with incompatible reactions towards rice line 75-1-127, the *Pi9* donor. Apart from BF5, *AVR-Piz-t* was present in all analysed isolates (n = 21) which correlated to isolates lacking virulence on rice line *Toride*1, which is the *Piz-t* R-gene donor. *AVR-Pib* occurs in most isolates (n = 21) but was missing in TZ090 (n = 1). I observed sequence variation in the promoter region of *AVR-Pib* in isolates that clustered in clade 1 (n = 8), excluding TG0004. All of these *M. oryzae* isolates showed sequence variation in the promoter region which correlated with virulence on the IRBLB-B monogenic rice line carrying *Pib*. Isolates without this sequence variation produced an incompatible reaction on the *Pib* rice line. All isolates from Sub-Saharan Africa carry *PWL2* (n = 19) which was identical to the two copies annotated in 70-15 see Figure 4.4. Jum1 and EG308 had the virulence allele of *pwl2*.

AVR-Pia was present in most isolates from clade 2 (n = 8), rendering these isolates avirulent on the rice monogenic line carrying *Pia.* KE210 and TZ090 did not possess *AVR-Pia* and were pathogenic towards monogenic rice

lines carrying *Pia*. *AVR-Pik* alleles were present in all isolates in clade 1 (n =8) apart from TG004. With the exception of KE255, KE210 and TZ090, the *AVR-Pik* alleles were absent in isolates of clade 2. All analysed isolates in clade 3 by contrast carried *AVR-Pik* alleles. Correlation between presence/absence of *AVR-Pik* and virulence is discussed in Section 4.2.3.2. *AVR-Pita* was present in both clades, clade 1 (n = 5) clade 2 (n = 3) and clade 3 (n = 3) but there was no correlation with virulence on monogenic line carrying *Pita. AVR-Pita* is a very polymorphic avirulence gene, and this may explain the lack of correlation between presence of *AVR-Pita* and lack of virulence on *Pita.* Only 6 isolates showed incompatible reactions on the monogenic line carrying *Pita.* In most isolates in clade 2, *AVR-Pita* was absent (n = 7) and this correlated with virulence on rice monogenic line carrying R-gene *Pita*. *AVR-Pii was* absent in all isolates from Sub-Saharan Africa but was present in EG308 and JUM1 isolates from Egypt and USA respectively.

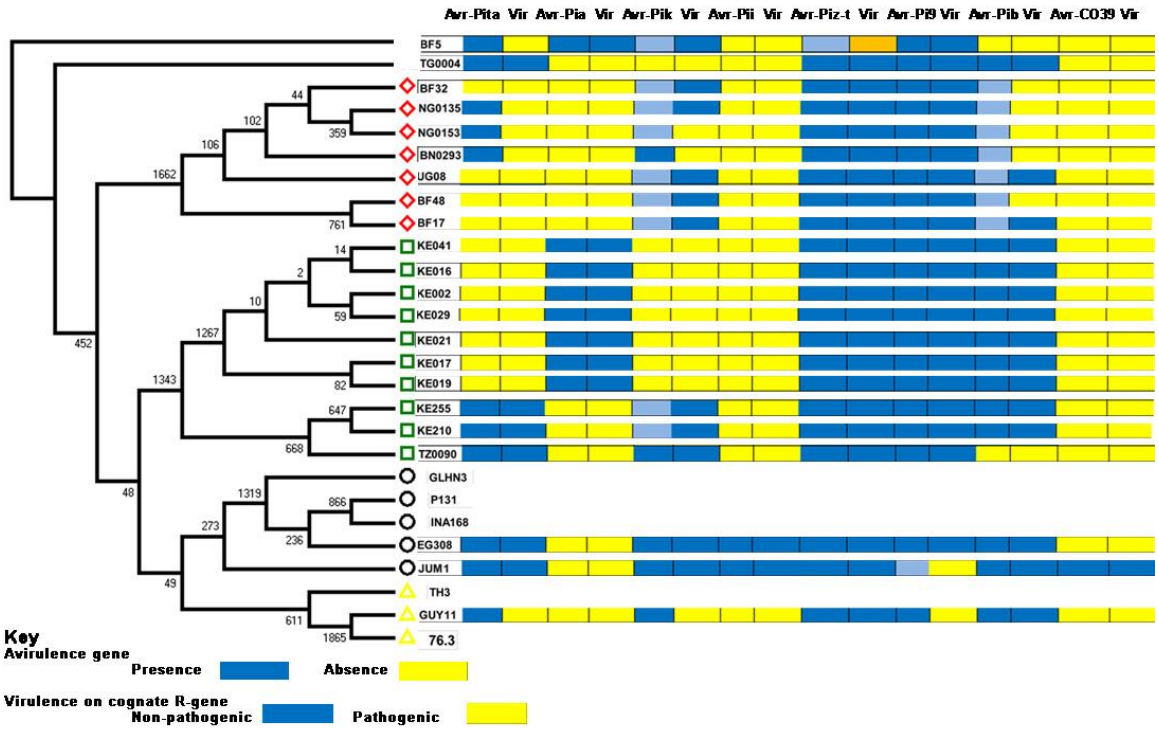

Figure 4.6 The avirulence genes repertoire of the sequenced *M. oryzae* isolates compared to virulence on rice monogenic lines.

Maximum parsimony tree of Sub-Saharan Africa *M. oryzae* isolates on the left. Presence or absence of each avirulence gene is represented by blue or yellow on the first column while the next column on the right represents virulence (Vir) on rice genotypes carrying cognate R-gene. For example, presence of *AVR-Pi9* (column 12) correlated with lack of virulence on rice line carrying *Pi9* (column 13). Yellow represents virulence and blue lack of virulence. Presence/absence of AVRs was determined using BLAST using BLAST 2.2.22.

4.3.3.2 Occurrence of *AVR-Pik* alleles suggest selection pressure against the rice gene *Pik*

The *AVR-Pik/km/kp* locus exhibits nucleotide changes that result in amino acid substitutions [46, 85]. On the other hand, the rice R gene *Pik* has multiple alleles that include *Pik-k, Pikm, Pikp, Piks* and *Pikh* which each exhibit different resistance specificities [85]. Selection polymorphism in *AVR-Pik* was reported by Yoshida *et al* [46] through the cloning of three alleles *AVR-Pik/km/kp*. The rice resistance protein Pik has also been shown to have multiple alleles in the form of Pik, Pikm, Pikp, Piks and Pikh which have varied recognition to the AVR-Pik alleles [85]. AVR-Pik exhibits polymorphism at amino acids 46-48 and 67, while polymorphism in Pik-1 is exhibited in the integrated HMA domain [74]. *Pikp* is resistant to strains that harbour *AVR-Pik*-D but is susceptible to those harbouring *AVR-Pik*-E, A and C [85]. *Pikm* is resistant to strains that possess *AVR-Pik*-D, -E and A but is susceptible to those harbouring *AVR-Pik*-C [85]. *Pik-k* is resistant to strains that possess *AVR-Pik*-D and *AVR-Pik*-E but is susceptible to those carrying *AVR-Pik*-A and *AVR-Pik*-C [85]. Piks has same resistance specificities identical to Pikp while Pikh has similar recognitions to Pikm [85]. AVR-Pik-D is recognised by almost all the Pik alleles including Pikp, Pikm, Piks and Pikh [85]. AVR-Pik-E is recognised by Pikm and Pikh [85].

Using BLAST 2.2.22, I analysed presence/absence polymorphism of these alleles in the genomes of the sequenced rice blast population (Figure 4.7). I observed that *AVR-Pik*-E was the most frequently occurring allele (n = 6) especially in clade 1. *AVR-Pik*-D (n = 3) and *AVR-Pik*-C (n = 3) showed a lower

frequency while *AVR-Pik*-A (n = 2) only occurred in clade 3. As expected Pikm (n = 6), Pikh (n = 5) and Pik-k (n =5) were the most effective of the alleles and could recognise AVR-Pik-D and AVR-Pik-E encoded by most analysed isolates. However, some of disease reactions produced were not clear cut and were slower in some cases, which may explain the ambiguous correlation between some of the alleles and disease reaction on specific rice line.

Virulence on rice genotypes with *Pik* alleles							Occurrence of *AVR-Pik* alleles				
	Pik-k	*Pik-s*	*Pik -h*	*Pik-m*	*Pik-p*		*AVR_PikA*	*AVR_PikB*	*AVR_PikC*	*AVR_PikD*	*AVR_PikE*
BF5	R	S	R	R	R	BF5					
TG0004	S	S	S	S	S	TG0004					
BF32	R	S	R	R	S	BF32					
NG0135	S	S	S	S	S	NG0135					
NG0153	S	S	S	S	R	NG0153					
BN0293	S	S	R	R	S	BN0293					
UG08	R	R	R	R	S	UG08					
BF48	R	S	S	R	R	BF48					
BF17	R	S	R	R	R	BF17					
KE041	S	S	S	S	S	KE041					
KE016	S	S	S	S	S	KE016					
KE029	S	S	S	S	S	KE029					
KE021	S	S	S	S	S	KE021					
KE002	S	S	S	S	S	KE002					
KE017	S	S	S	S	S	KE017					
KE019	S	S	S	S	S	KE019					
KE255	S	S	R	S	S	KE255					
KE210	S	S	S	S	S	KE210					
TZ0090	S	S	S	S	S	TZ0090					
EG308	R	S	R	R	R	EG308					
JUM1	S	S	S	S	S	JUM1					
GUY11	R	S	S	S	S	GUY11					

Presence

Absence

Figure 4.7 The presence of distinct *AVR-Pik* alleles compared to the virulence on rice monogenic lines containing *Pik* alleles.

The presence/absence polymorphism of *AVR-Pik* was determined by BLASTn using BLAST 2.2.22. Blue boxes represent presence of specific *AVR-Pik* allele, light blue represents occurrence of nucleotide substitution of the known allele and yellow refers to absence of any of the alleles. R represents resistance reaction and S represents susceptibility reaction of specific R-gene.

4.3.4 Three novel secreted proteins exhibit BIC during rice blast infection

M. oryzae effectors are secreted by invasive hyphae and either translocate into the biotrophic interfacial complex (BIC) or localise in the apoplast [2]. This criterion can be used to characterise *M. oryzae* effectors [1]. Apart from AVR-CE1, which is a polyketide synthase that localises in the appressorium, all cloned *M oryzae* avirulence proteins have been shown to localise to the BIC before translocating into the cytoplasm [57]. Rice sub-cellular localisation of identified secreted proteins encoding genes was investigated as shown in Figure 4.8. To first identify putative effector genes, gene prediction process was carried out using program called Maker as explained in 3.3.2. Different bioinformatics tools were the used to improve gene prediction; annotation and characterisation as detailed in see Section 3.3.2.2. The coding sequencing of selected genes (Table 4.3) with their native promoters and expressing GFP at the C-terminal were cloned into fungal transformation vector, pCB1532.

Successful constructs were confirmed by sequencing and transformed into either KE002 or Guy11 backgrounds. Successful transformants were then used to inoculate a susceptible *Moukoto* rice leaf sheath and sections of leaf sheath observed by epifluorescence microscope after 24-48 h. Three of the predicted genes were found to encode for putative effector proteins and were named *Magnaporthe* oryzae effector proteins (MEP). Mep13, Mep14 and Mep15 were expressed and localised into the BIC as shown in Figure 4.8. To understand virulence of these two effectors, a *Δmep13* null mutant was

generated in KE002 background using the split marker method. For *MEP14*, gene complementation was carried out.

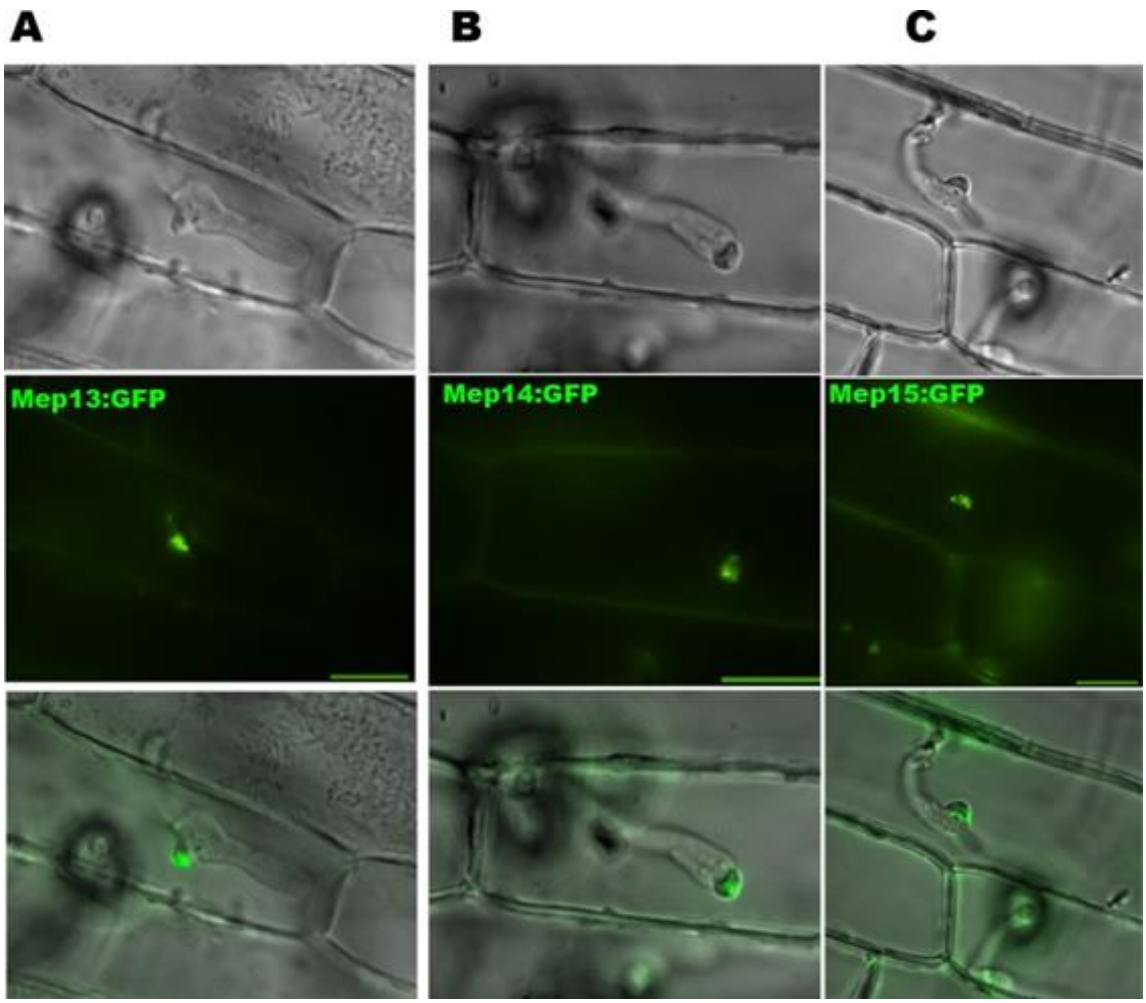

Figure 4.8 Expression and localisation of Mep13, Mep15 and Mep14 into the BIC.

Micrographs obtained on conventional epifluorescence show leaf sheath tissue of susceptible rice cultivar *Moukoto* invaded by KE002 strains expressing GFP-fused putative effectors. (A) Mep13-GFP, (B) Mep14-GFP and (C) Mep15-GFP 25 hpi. Scale bar represent 10 µm.

Table 4.3 List of uncharacterised predicted effectors that show *in planta* up-regulation during rice blast infection.

Gene name	Length	Cysteine content	BLAST2GO annotation	EffectorP Prediction	Remarks[6]
KE002Y1_contigs.g9192.t1	162	4	Protein elicitor protein	Effector	
KE002Y1_contigs.g1117.t1	104	6	hypothetical Y34	Effector	MAX
KE002_13250	82	6	hypothetical MGCH7	Effector	MAX
Mep15	97	2	hypothetical Y34	Effector	
KE002_7508	115	0	hypothetical 7bg7.17	Effector	
KE002_4999	85	12	hypothetical Y34	Effector	
KE002_5442	120	0	hypothetical Y34	Effector	
KE002_6605	143	3	hypothetical Y34	Effector	
Mep14	103	8	hypothetical Y34	Effector	
KE002_8917	123	5	NA	Effector	
KE002_10623	228	0	hypothetical 7bg7.17	Effector	
KE002_10686	103	0	hypothetical 7bg7.17	Effector	Nucleus targeting
KE002_14611	40	1	hypothetical Y34	Effector	
KE002_15475	122	2	hypothetical Y34	Effector	Nucleus targeting
Mep13	121	4	NA	Effector	MAX

[6] Selected genes encoding for secreted proteins were further annotated using *Localizer* or by Thomas Kroj to determine presence of MAX-domain.

4.3.4.1 *MEP13*

The *MEP13* gene was up-regulated and highly expressed at 48 hpi in invasive hyphae during KE002 infection on *Moukoto* (Figure 4.9). *MEP13* encodes a hypothetical protein with 121 amino acids. No paralogs were found in *M. oryzae* version 8 and no orthologs occurred in other organisms. This gene was not annotated in the 70-15 genome and was thought to be absent. However, BLASTn search in different *M. oryzae* isolates proved that it was conserved in most isolates, including 70-15, probably in unassembled reads. BLASTn search in the Guy11 genome sequence generated using Pacbio, produced three hits for this gene representing three copies in the genome. This suggested that this gene has undergone gene duplication and genome expansion. BLASTn search in KE002 genome produced a single hit, equivalent to one copy and subsequent gene deletion was carried out in the KE002 background. The encoded protein did not show similarity to any known proteins. Mep13 has four conserved cysteines residues and was predicted to be a MAX domain containing effector (courtesy of Thomas Kroj INRA, BGPI, Biology and Genetics of plant-pathogen interactions, Montpellier). The protein predominantly localises in the BIC (Figure 4.8).

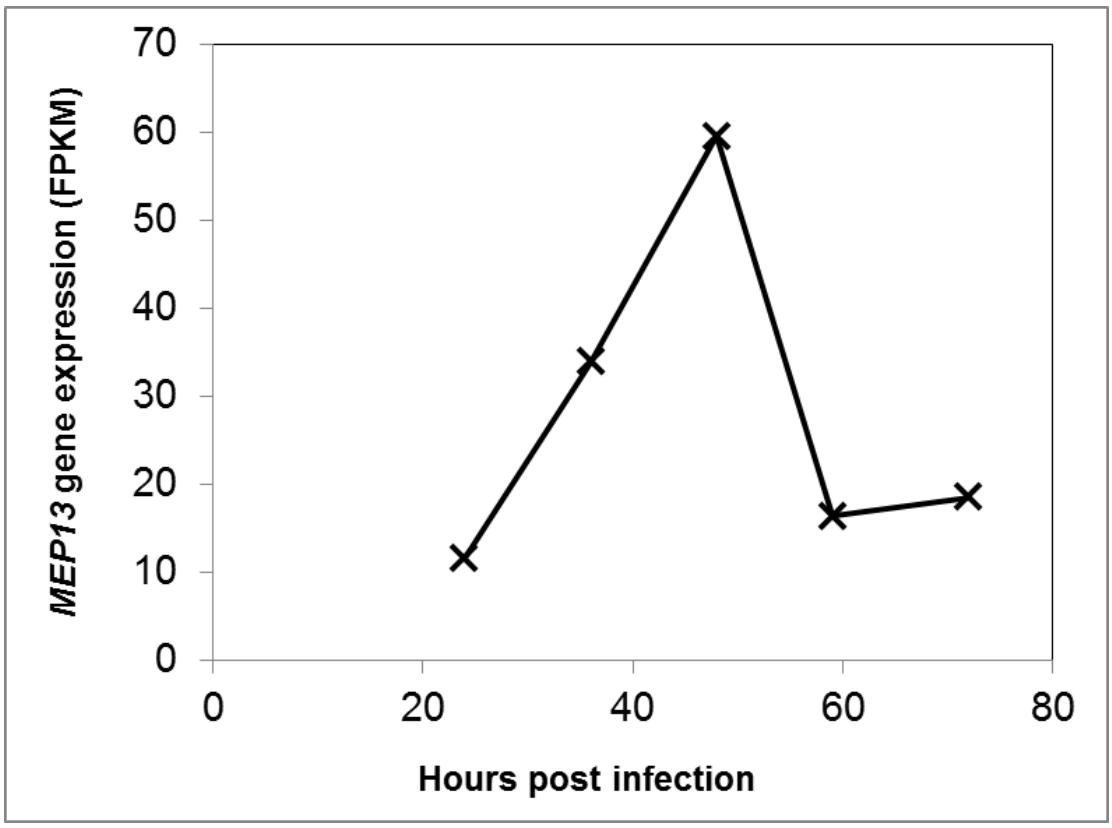

Figure 4.9 Expression at transcript-level of *MEP13* during infection of susceptible *Moukoto* rice line by *M. oryzae* isolate KE002.

Maximum gene expression of *MEP13* was observed at 48 hpi. No expression of *MEP13* was observed in *M. oryzae* mycelium (0 h). FPKM values are represented on the Y-axis while h post-infection at 24, 36, 48, 59 and 72 shown on the X -axis. *M. oryzae* mycelium transcriptome were used as negative control. Fragments per Kilobase of transcript per million mapped reads (FPKM) values were generated using cuffdiff algorithm.

4.3.4.2 *MEP14*

This gene was up-regulated and highly expressed mostly at 24 and 48 hpi in invasive hyphae during KE002 infection of *Moukoto* (Figure 4.10). *MEP14* encodes a putative-secreted protein with 103 amino acids, including eight cysteine residues. BLASTn searches in sequence *M. oryzae* isolates showed presence in all Sub-Saharan Africa isolates except BF5. The gene was also absent in Guy11 and JUM1. Microscopy of the secreted Mep14-GFP protein showed fluorescence in the BIC of invasive hyphae (Figure 4.8). As observed in the transcriptomic data, the maximum fluorescence was observed at 24-25 h then at 48 h, immediately after penetration and in newly formed invasive hyphae of invaded neighbouring cells. Attempts to generate null mutants in KE002 have not been successful. To functionally characterise this gene, complementation analysis was carried out by expressing the gene in Guy11.

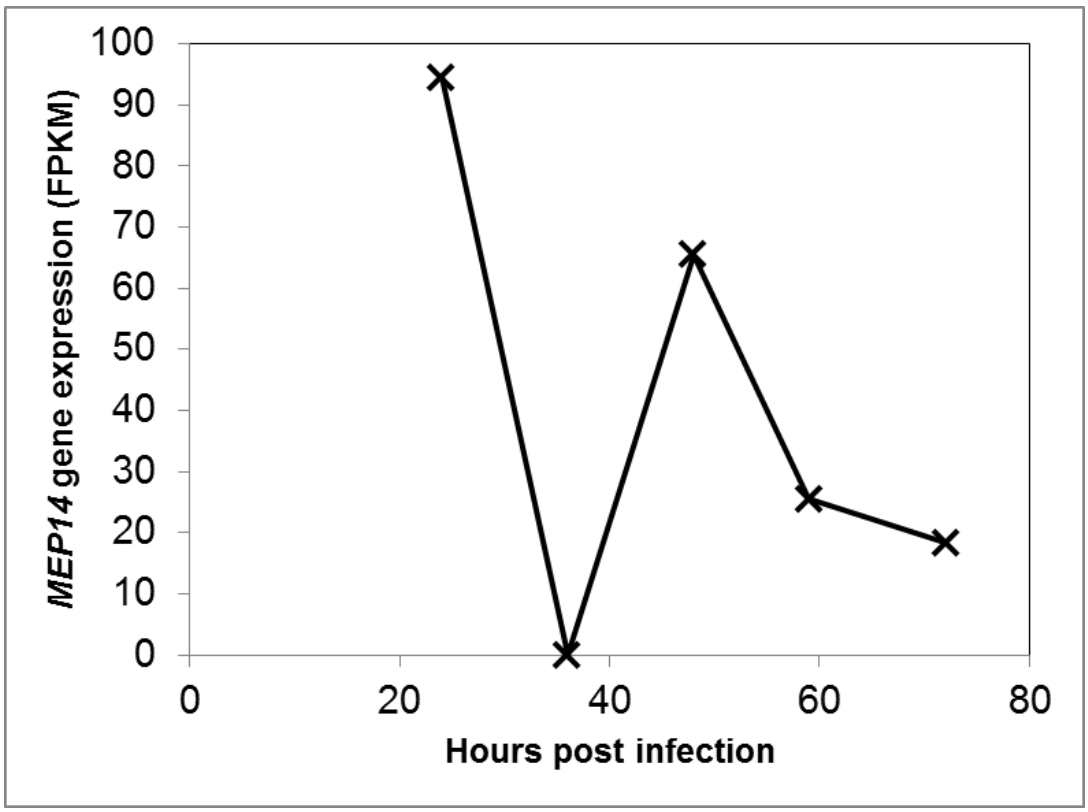

Figure 4.10 Figure 4.10 Expression at transcript-level of *MEP14* during infection of susceptible *Moukoto* rice line by KE002.

Maximum gene expression of *MEP14* was observed at 24 and 48 hpi. There was no expression in *M. oryzae* mycelium (0 h). FPKM values are represented on the Y-axis while h post-infection at 24, 36, 48, 59 and 72 shown on the X -axis. *M. oryzae* mycelium transcriptome were used as negative control. Fragments per Kilobase of transcript per million mapped reads (FPKM) values were generated using cuffdiff algorithm.

4.3.4.3 *MEP15*

This gene encodes a putative-secreted protein with 97 amino acids containing two cysteine residues. The gene was highly up-regulated from 36 - 48hpi in the invasive hyphae during KE002 infection on susceptible rice line *Moukoto* (Figure 4.11). BLASTn search in the genomes of different *M. oryzae* isolates showed presence in all Sub-Saharan Africa isolates but with a nucleotide substitution leading to amino acid change in BF5, TG004. The gene is absent in JUM1, EG308, Y34 and P131. Microscopy of the secreted Mep15-GFP protein showed strong fluorescence in the BIC of invasive hyphae (Figure 4.8). Mep15 was also predicted as a MAX domain containing effector (courtesy of Thomas Kroj). Gene replacement mutant or complementation for this gene has not yet been carried out.

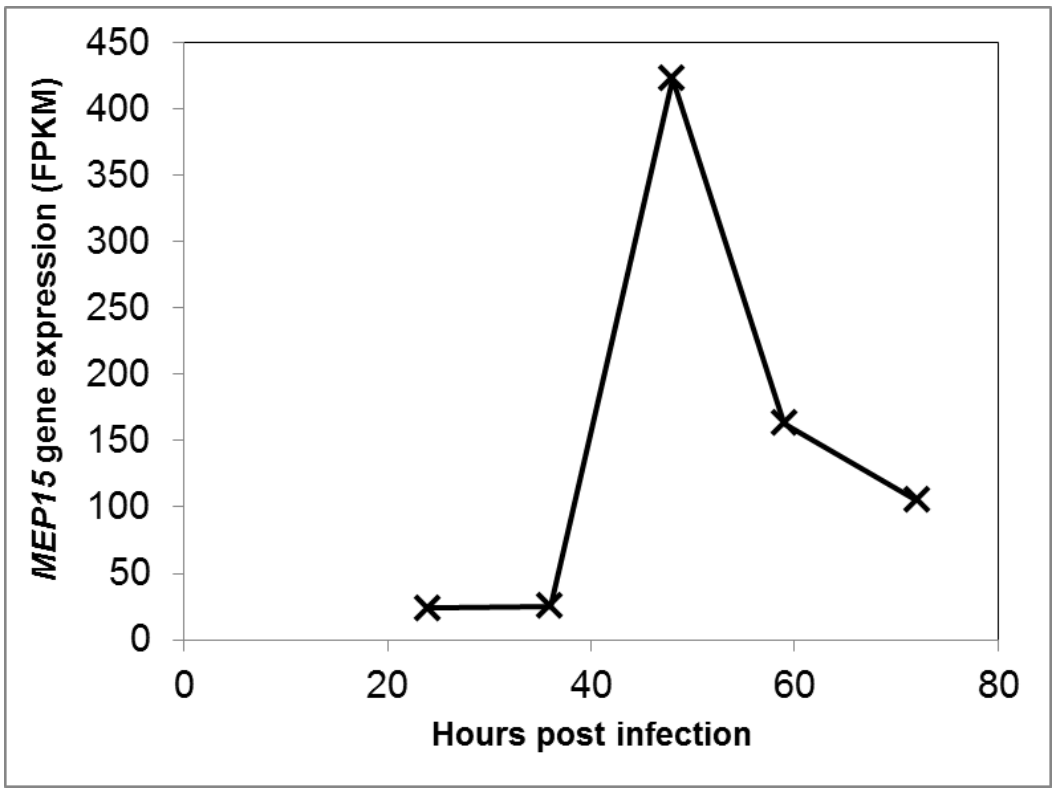

Figure 4.11 Expression at transcript-level of *MEP15* during infection of susceptible *Moukoto* rice line by KE002

Maximum expression of *MEP15* was observed at 48 hpi. There was no expression in *M. oryzae* mycelium (0 h). FPKM values are represented on the Y-axis while h post-infection at 24, 36, 48, 59 and 72 shown on the X -axis. *M. oryzae* mycelium transcriptome were used as negative control. Fragments per Kilobase of transcript per million mapped reads (FPKM) values were generated using cuffdiff algorithm.

4.3.5 Targeted deletion of *MEP13* does not have a phenotypic characteristic

To determine the role of *MEP13* during biotrophic invasive growth, I carried out gene replacement using PCR-based split marker deletion method (Figure 4.12) (Kershaw and Talbot, 2009). In first round of PCR, sequences up-stream and downstream of *MEP13* coding sequence were PCR amplified using primers designed to include overhangs that are complementary to a fragment of hygromycin phosphotransferase resistance gene cassette *HPH*. The reverse primer for the up-stream sequence provides regions complementary to up-stream half of hygromycin cassette, while the forward primer used to amplify the down-stream half had an extension complementary to the other half of hygromycin cassette.

In the second round of PCR, two fused fragments were generated. The up-stream and down-stream fragments of *MEP13* coding sequence were fused together with up-stream half (HY) and down-stream half (YG) of hygromycin cassette at regions with overhang sequences generated in first round PCR. Regions surrounding the coding sequence allow for homologous combination in fungal genome leading to targeted deletion (replacement by the hygromycin cassette). The two PCR products each containing overlapping fragments of hygromycin cassette flanked with sequences homologous to targeted gene for deletion were used to transform *M. oryzae* protoplast and the transformed protoplasts grown on plates overlaid with media containing hygromycin as elaborated in Section 2.6. Primers used for targeted gene deletion of *MEP13* are listed in Table 4.4.

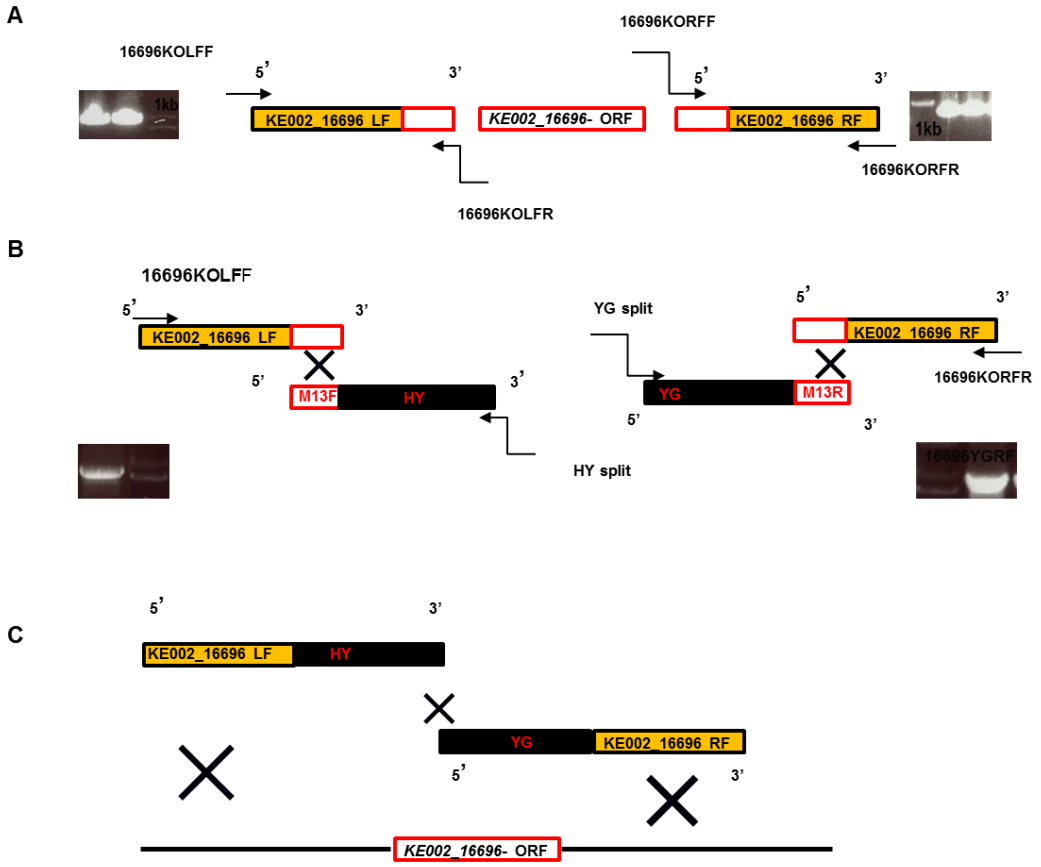

Figure 4.12 Schematic representation of PCR-based split-marker deletion method used for targeted deletion of *MEP13*.

(A) In the first-round PCR, 849bp sequence upstream and 1210bp downstream of *MEP13* coding sequence were amplified using primers indicated in Table 4.4. (B) In the second-round PCR, *MEP13* LF was fused to one half of hygromycin resistance gene cassette (HY) using a reverse primer that generated overhangs complementary to sequences in the YG fragment. *MEP13* RF was fused to the other half of Hygromycin cassette (YG) using a forward primer that generated overhangs complementary to sequences in the HY. (C) Second round PCR products were used for fungal transformation. The flanking regions used for homologous recombination, resulting in replacement of *MEP13* coding sequence with hygromycin.

4.3.5.1 Analysis of *Δmep13* putative transformants

Hygromycin resistant transformants of *M. oryzae* were sub-cultured on cellophane discs for DNA isolation. Genomic DNA from the transformants and wild type KE002 was digested with *Bam*H1 and *Hind*III followed by Southern blot analysis. Two different restriction enzyme digest were carried out. Both involved digestion of genomic DNA using either a combination of *Bam*H1 and *Hind*III or a single digest with *Hind*III. Digested genomic DNA was fractionated using gel electrophoresis before being transferred onto two different Hybond-N membranes (Amersham). The membrane containing fragments digested with *Hind*III was probed using *MEP13* coding sequence probe *(*Figure 4.13).

The membrane containing fragments digested with *Bam*H1 and *Hind*III was probed using *MEP13* left flank probe (Figure 4.13). The two probes were generated using DIG-labelling PCR. The coding sequence probe was hybridised to a 4.1 kb fragment in separated wild type KE002 genomic DNA and transformants that had ectopic insertion. The probe did not hybridise to any fragments in transformants 5, 8 and 14. In the second blot, the left flank probe hybridised to a 1.88kb fragment in separated wild type KE002 genomic DNA and transformants with ectopic insertion but hybridised to a 5.1kb fragment in transformants 8 and 14, with successful deletion of *MEP13*. T8 and T14 were identified as *Δmep13* null mutant and used for subsequent phenotypic analysis.

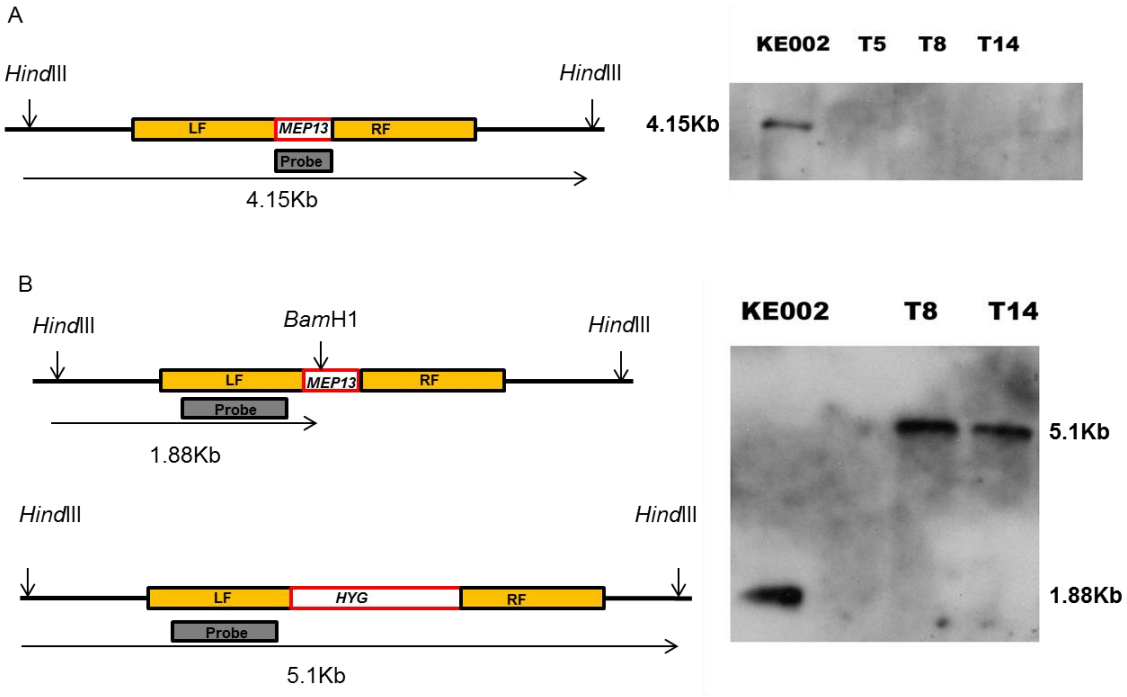

Figure 4.13 Figure 4.13 Southern blot analysis of selected *Δmep13 putative* mutants.

(A) Southern blot analysis showing *Hind*III restriction digest probed with *MEP13* coding sequence probe. The probe hybridised fragmented to a 4.1 Kb fragment in wild type KE002, but did not hybridise to T5, T8 or T14 genomic DNA. (B) Southern blot analysis showing *Bam*H1 and *Hind*III restriction digest probed using the *MEP13* left flank probe. The probe hybridised to a 1.88 Kb fragment in wild type KE002 but 5.1 Kb fragments in T8 and T14.

4.3.5.2 Phenotypic and pathogenicity assay of *Δmep13* mutant

Vegetative growth and colony morphology of *M. oryzae Δmep13* mutants were analysed. The KE002 and T14 strain (*Δmep13* mutant) were grown on CM plates and observed 10 days post sub-culturing as shown in Figure 4.14. T14 displayed vegetative growth like the wild type KE002 which had normal dark concentric rings and light growing edges. This suggested that *MEP13* might not be involved in vegetative growth of the rice blast fungus on CM.

To investigate the role of *MEP13* in rice blast disease, T14 was tested for gain of virulence on selected rice monogenic lines. I reasoned that if *MEP13* is an avirulence gene, T14 (*Δmep13* mutant) would gain virulence on rice monogenic lines containing *Pit, Piz, Piz-5, Pi1, Pi11 (t), Pi12 (t), Pi19 (t)* and *Pi20*, resistant to wild type KE002. *Δmep13* deletion mutants however showed no gain of virulence on any of the analysed monogenic lines that were resistant to KE002 5-6 days after infection. Moreover, *MEP13* does not have pathogenic defects on rice infection because *Δmep13* mutant was virulent on several rice lines as shown in Figure 4.14. However, loss of *MEP13* might have other minor fitness defects that are not easily visible and they will be studied in the future.

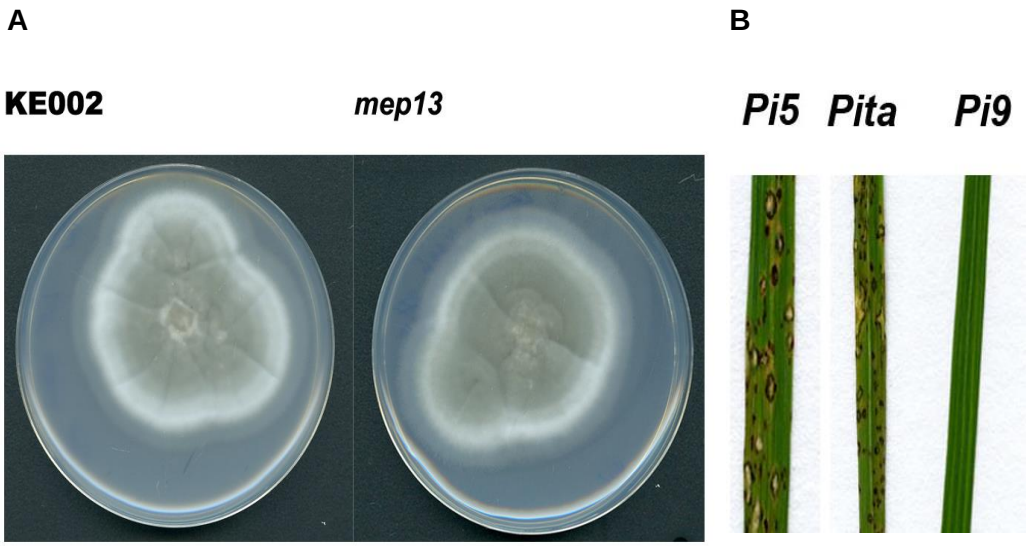

Figure 4.14 Colony morphology and compatibility assay of *Δmep13* mutant.

(A) Wild type KE002 and *Δmep13* were inoculated on CM plates and incubated at 25 °C. Images were obtained after 10 days. B) Transformant T14 spores were used to spray 3 weeks old monogenic line plants and images were obtained 6 days' post-infection. Disease reactions from rice monogenic lines carrying *Pi5, Pita* and *Pi9* inoculated with *Δmep13*. Images were obtained using Epson Expression 1680 Pro scanner. The mutant was tested three times with three replicates for each rice genotype. For each genotype, observations were consistent in 3/3 infections.

4.3.6 *MEP14* gene complementation in Guy11 and analysis for virulence phenotypic characteristic

To determine the role of *MEP14* as a virulence determinant gene, and to understand its function during biotrophic invasive growth, genetic complementation was carried out. *MEP14* was transformed into Guy11, a pathogenic strain that lacks this gene. *MEP14* coding sequence including a 1.2 Kb sequence containing its native promoter and 1 Kb terminator sequence were cloned into a fungal transformation vector Pcb1532. Successful construct was confirmed by sequencing then transformed into isolated Guy11 protoplast as elaborated in Section 2.6. Resistant colonies were selected for genomic DNA extraction and screened for successful integration of the *MEP14* locus using PCR.

4.3.6.1 Analysis of *MEP14* gene complementation putative transformants

In successful transformants, a PCR product of similar size (*MEP14* coding sequence) like that from KE002 genomic DNA was obtained in transformants but not from wild type Guy11. These transformants were further analysed using Southern blot analysis. Wild type KE002, Guy11 and selected transformants genomic DNA was digested using *Xba*1 and *Hind*III. Digested genomic DNA was fractionated by gel electrophoresis before being transferred onto a Hybond-N membrane. The membrane was probed using *MEP14* coding sequence probe generated by DIG-labelling PCR to detect successful insertion of *MEP14*. The coding sequence probe hybridised to a 2.5 kb fragment in the wild type KE002 and all selected transformants but not in wild type Guy11 *(*Figure 4.15). However, only transformants that had one copy integrated were

selected for phenotyping. Transformants T5, T6 and T8 were selected for further study following successful complementation with *MEP14*.

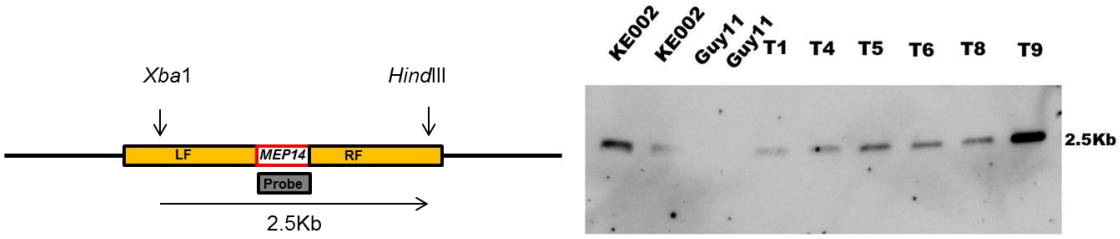

Figure 4.15 Southern blot analysis of selected *MEP14* genetic complementation transformants.

Southern blot analysis showing *Xba*1 and *Hind*III restriction digest of wild type KE002, Guy11 and transformants fragmented genomic DNA probed with *MEP14* coding sequence probe. The probe did not hybridise to fragmented wild type Guy11 genomic DNA but hybridised to fragmented wild type KE002 and all transformants. Transformants with single hybridised band were selected.

4.3.6.2 Phenotypic and pathogenicity assay of *MEP14* genetic complements

Vegetative growth and colony morphology of *MEP14* genetic complements were analysed. Guy11 and transformant T6 (complemented with *MEP14*) were grown on CM plates and observed 10 days post sub-culturing. T6 display vegetative growth like Guy11 and had normal dark concentric rings and light growing edges (Figure 4.16). *MEP14* did not alter the vegetative growth of Guy11 on CM. I reasoned that if *MEP14* is an avirulence gene, genetic complements would lose virulence on rice monogenic lines infected by wild type Guy11. Virulence of these transformants was analysed on rice monogenic lines *Pish, Pit, Pita-2, Piz-5, Pi3, Pi5 (t), Pi7 (t), Pi11 (t), Pi12 (t), Pi19 (t),* and *Pi20* that are susceptible to Guy11. 3-4 weeks old monogenic rice plants were inoculated with conidia collected from 8-12 days old cultures of Guy11 complemented with *MEP14* (T6). Disease symptoms were analysed 5 -7 days post-infection. T6 was still able to cause disease in the analysed rice cultivars and did not show any defects as shown in Figure 4.16. However, it is possible that *MEP14* might be involved in biotrophic growth but does not have visible virulence effects. *MEP14* was highly up-regulated immediately after penetration and in newly formed invasive hyphae of invaded neighbouring cells which suggests that this effector might play an important role during new cell penetration. Null mutant in KE002 background will give more insight to the function of *MEP14.*

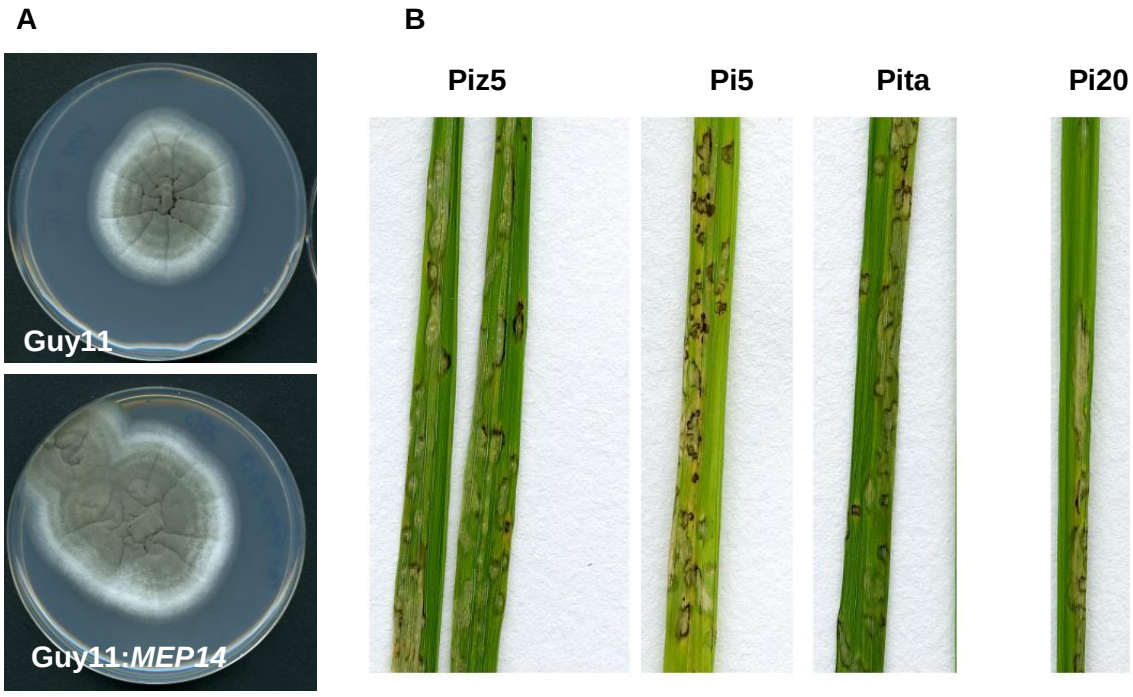

Figure 4.16 Colony morphology and compatibility assay of *MEP14* genetic complementation transformant.

(A) Guy11 and transformant T6 (complemented with *MEP14*) were inoculated on CM plates and incubated at 25 °C. (B) T6 spores were used to spray 3 weeks old monogenic line plants and images obtained 6 days' post-infection. Images were obtained after 10 days using Epson Expression 1680 Pro scanner. The genetic complement was tested two times with three replicates for each rice genotype. For each genotype, observations were consistent in 2/2 infections.

Table 4.4 List of oligonucleotide primers used in Chapter 4

Primer	Oligo sequence (5' to 3')
16696ORFF	ATGCACCCCGAGAACCTTTTCGCC
16696ORFR	CTAGATTCTGACATTCGGGAACCT
6959ORFF	ATGCGCAGCTCTCTCATCACCCTC
6959ORFR	TTACTCGGCGCAACCGACAAATCC
KE26959F	CGCGGTGGCGGCCGCTCTAGATCTTTTCTTCTGGCAATACGCGGA
KE26959R	CTCGGCGCAACCGACAAATCCAAC
KE26959FGFP	GTTGGATTTGTCGGTTGCGCCGAGATGGTGAGCAAGGGCGAGGAGCTG
KE216696F	CGCGGTGGCGGCCGCTCTAGAGGGGAAACCCTATGGGGCTTGTAG
KE216696R	GATTCTGACATTCGGGAACCTGGC
KE216696FGFP	GCCAGGTTCCCGAATGTCAGAATCATGGTGAGCAAGGGCGAGGAGCTG
1434PF	CGCGGTGGCGGCCGCTCTAGATGCTGCAGTGTTTTGATGGCCAGG
1434PR	CTTGCTGTAGGTCGTAACTTTGTG
1434FGFP	CACAAAGTTACGACCTACAGCAAGATGGTGAGCAAGGGCGAGGAGCTG
Kpn1Trpc-R	CTATAGGGCGAATTGGGTACCAGTGGAGATGTGGAGTGGGC
16696KOLFF	TGAAGACTATCCAGACCCCCCAAT
16696KOLFR	TCGTGACTGGGAAAACCCTGGCGGCAGAGGCTGGGAAAAGCGCCGTG
16696KORFF	TCCTGTGTGAAATTGTTATCCGCTCGTGCGACACCCCCCCAGAGTATC
16696KORFR	GTCGTTCTGGGTCGGCCTGAGCAT
KE26959F2	TCTTTTCTTCTGGCAATACGCGGA
KE26959termR	CTATAGGGCGAATTGGGTACCAACACCAGCGTTTCGAATGCCCAA

4.4 Discussion

In this chapter, association genetics was employed to identify novel avirulence genes from 23 sequenced *M. oryzae* isolates mainly from Sub-Saharan Africa. First, extensive pathotype analysis was carried out on monogenic rice lines carrying 24 rice blast resistance genes; *Pia, Pib, Pik, Pik-h, Pi-km, Pik-p, Pik-s, Pish, Pit, Pita, Pita-2, Piz, Piz-5, Piz-t, Pi1, Pii, Pi3, Pi5 (t), Pi7 (t), Pi9, Pi11 (t), Pi12 (t), Pi19 (t),* and *Pi20*. Genome analysis for presence of cloned avirulence genes showed that *AVR-Pi9* and *AVR-Piz-t* are the most frequently occurring avirulence genes among all sequenced isolates. These isolates produce non-compatible reactions on rice lines carrying *Pi9* and *Piz-t,* respectively. Among the *AVR-Pik* alleles, *AVR-Pik*-D and *AVR-Pik*-E occurred in most *M. oryzae* isolates suggesting selection pressure from corresponding *Pik* alleles. An arms-race co-evolution study on *Pik* and *AVR-Pik* alleles has suggested that *AVR-Pik*-E is derived from *AVR-Pik*-D [85]. Moreover, a phylogenetic analysis revealed that *AVR-Pik*-D is the ancestral allele from which *AVR-Pik*, *AVR-Pik*-A, *AVR-Pik*-B, *AVR-Pik*-C and *AVR-Pik*-E were derived [85]. This study suggested that the most ancestral *AVR-Pik* allele will be recognised by most *Pik* alleles and has undergone less selection [85].

From our analysis, most *M. oryzae* isolates from West Africa carried either *AVR-Pik*-D or *AVR-Pik*-E of which were recognised by most *Pik* alleles except *Pik-s* [85]. Deployment of *Pik* alleles that recognises *AVR-Pik*-D might have resulted in selection of *AVR-Pik*-E to evade recognition, especially in West Africa. Two *M. oryzae* isolates from East Africa (KE210 and KE255), EG308 (Egypt) and JUM1 (USA) showed more divergence and carry *AVR-Pik*-C and

AVR-Pik-A, respectively, which suggested more selective pressure on cultivation of rice line with more recent *Pik* alleles. Genetic analysis of rice cultivars grown where these *M. oryzae* strains were isolated can offer more information regarding selection pressure imposed.

Phylogenetic analysis suggested that *M. oryzae* isolates in this study fell into three major clades that differed in virulence determinant genes and virulence spectra. Although there was no major differences in virulence, isolates from clade 1 (West Africa) and clade 3 (mostly Asia) showed more virulence that those in clade 2 (East Africa). It is possible that there have been more events of evolution in isolates from West Africa and Asia (clade 3) than in East Africa. Asia and West Africa have longer rice growing history than East Africa, especially Kenya [231, 232]. Cultivation of rice cultivars of different genotypes might contribute to differences in virulence towards certain monogenic lines and result in varied effector/avirulence gene repertoire in a given rice blast population [231, 232]. Isolates from West Africa were virulent on both *Pia* and *Pib,* which was not the case with isolates from East Africa. All isolates from West Africa lacked *AVR-Pia* and had variable sequences in the *AVR-Pib* promoter region which could explain this observation. Lack of *AVR-Pia* and occurrence of a virulence *AVR-Pib* also suggests possibilities of selection pressure in West Africa compared to East Africa.

This study shows that a combination of screening a rice blast population and comparative genomics can be a powerful tool for pathogen surveillance and breeding for resistance. Most isolates analysed from each region of Sub-Saharan Africa shared genetic relatedness which includes the avirulence gene

repertoire. With a few exceptions, isolates from each clade showed similar disease reactions on analysed rice genotypes. This can be used to provide representation of a regional population and assist in selecting suitable R genes for deployment. For example, most isolates were non-pathogenic on *Pi9* which correlated with presence of *AVR-Pi9* in rice blast population. This means that *Pi9* is a suitable candidate to be used in combination with other R genes to breed for resistance in these two regions of Africa.

The presence of specific R genes will shape the genetic architecture of a *M. oryzae* population in a given geographic region [124]. To gain an understanding of genetic variability in a blast population, avirulence determinant genes can be used to monitor frequencies of resistance gene breakdown and the possibility of a host jump [231]. An extensive analysis of blast population and genome wide-analysis for the presence/absence of cloned and novel avirulence genes will assist in planning for effective rice blast control [231]. Improvement in genome sequencing technologies and bioinformatics tools will also help increase the capacity to identify more avirulence genes.

The majority of avirulence genes in *M. oryzae* have been identified through map-based cloning [124]. With recent developments in understanding the evolution of avirulence genes, the drastic divergence in the gene locus or surrounding regions, is a major challenge in map-based cloning. This method falls short in detection of a point mutation or TE insertion for example that may lead to gain of virulence. Recent studies have employed whole genome sequencing to identify several avirulence genes [46]. A comparative genomics study using a field isolate and the laboratory isolate, 70-15, was used to identify

three avirulence genes *AVR-Pia*, *AVR-Pii* and *AVR-Pik* alleles [46]. In a different study, from sequencing an avirulent *M. oryzae* isolate, several candidate gene products were screened for interaction with *Pi54* to identify *AVR-Pi54* [243]. Whole genome sequencing of two closely related isolates from a sequential planting experiment enabled a quick identification of *AVR-Pi9* [42]. Both isolate carried *AVR-Pi9* gene but in the virulent isolate, gene function was disrupted by a Mg-SINE element insertion within *AVR-Pi9* locus [42].

Initial attempts to identify associations between presence and absence polymorphism of predicted genes and virulence on rice monogenic lines were not successful. The lack of correlation was thought to be due to incorrectly predicted genes or genes omitted during gene calling. This was related to incompleteness of most genome sequences used in this study and improper assembly of isolate specific region. Use of long-read assembled genome sequences made it possible to predict more genes and improve annotation of more secreted protein-encoding genes in the *M. oryzae* genomes generated. As detailed in Chapter 3, effector prediction program *EffectorP* was used to further improve the annotation process and was able differentiate effector genes from non-effector genes. *EffectorP* could accurately predict 3 effector proteins encoding genes, *MEP15*, *MEP13* and *MEP14* which were expressed and localised into the BIC, typical of all cloned avirulence proteins apart from Ace1 [96].

A *Δmep13* mutant did not show gain of virulence when analysed on selected rice monogenic lines (resistant to KE002). Moreover, *Δmep13* mutant was virulent on a susceptible rice line *Moukoto* and did not show any

pathogenic defects. *MEP14* complementation analysis did not show any fitness or virulence defects when analysed on selected rice monogenic lines and had no pathogenicity defects towards infecting rice plants. However this experiment will be followed up by QRT-PCR to confirm expression of *MEP14* in generated genetic complements. Maximum expression of fluorescently labelled Mep14 was observed at 24-25 h and 48 h, immediately after penetration and in newly formed invasive hyphae invading neighbouring cells. This suggests that Mep14 is a biotrophy-associated protein that might play an important function in newly invaded rice cells.

In Chapter 5 two *Δpwl2* mutants generated using CRISPR/Cas9 genome editing in a Guy11 background drastically gained virulence towards weeping lovegrass. However, these mutants retained pathogenicity on a susceptible rice cultivar, *Moukoto*. Consistent with other reports, most *M. oryzae* effectors might serve redundant functions that are complemented with several sets of other effectors [1, 95]. A study by Saitoh *et al* [95] for example, showed that disruption of more than 78 putative secreted proteins failed to produce any fitness, growth or pathogenicity defects. In a different study, gene replacement of two effectors *BAS2* and *BAS3* failed to show any mycelial growth, sporulation or pathogenicity phenotype [1]. It is possible that these effectors might have other minor fitness defects that are not easily quantifiable that need to be further studied.

Functionally redundant effectors from the same or phylogenetically unrelated pathogen can target the same host proteins [113, 219, 244]. In such cases the targeted protein might belong to a conserved pathway that pathogens

need to modulate for successful host colonisation [10, 57, 245]. In *M. oryzae* AVR-Pik, AVR-Pia and AVR-CO39 are examples of effectors that have evolved to target related host HMA-domain containing proteins thereby promoting infection [74, 84, 210]. It is possible that effector redundancy resulted from an arms-race like selection between the pathogens and their hosts and is advantageous to the pathogen in a changing environment [77, 220]. This means that pathogens carrying several effectors that counter the host immunity pathway will survive changes in new environment [77, 220].

This could explain why most studies on fungal effector null mutants do not show any visible phenotype in terms of fitness and pathogenicity. We propose identification of more effectors with similar host cell localisation patterns or conserved protein structures, as a way of elucidating effector function. This can be followed up by deletion of multiple effectors modulating the same cellular process which might produce drastic phenotypic changes. Alternatively, in vivo competition assays can be carried out to determine putative fitness defects of an effector mutant in presence of a wild virulent strain of *M. oryzae*.

Chapter 5 Functional characterisation of *PWL2*, a *Magnaporthe oryzae* host-range determinant gene

5.1 Introduction

Pathogenic fungi and oomycetes secrete effector proteins that modulate host physiology and cell signalling to sustain pathogen colonisation [10, 57, 245-247]. Understanding the exact function of these fungal effectors is key to elucidating the molecular basis of plant-microbe interactions and boosting the chances of managing crop disease [10, 207]. The study of effector function has become an important research theme in solving problems related to crop diseases. Improved technologies in whole genome sequencing have enabled identification of hundreds of genes that putatively encode for effector proteins and some have already been molecularly cloned and characterised [10]. However, important questions remain such as how fungal effectors translocate, localise and function in plant cells after being secreted into either the apoplast or host cytoplasm [98, 221, 245]. Studying sub-cellular localisation of effectors in host cell compartments will help to answer these questions and enable the host proteins with which effectors interact to be identified [207, 246, 247]. Fungal and oomycete effectors can target different subcellular compartments, including the nucleus, plasma membrane, ER and cytosol [97, 248, 249].

In bacteria, for example, *Pseudomonas syringae* effectors have been shown to target different host compartments, including chloroplast and mitochondria through sub-cellular localisation studies [97, 196, 248]. In general, effectors can have enzymatic activities, target nucleic acids or a range of host proteins. These targeted proteins can either be modulated to benefit the

pathogen or act as helper proteins that facilitate effectors trafficking or maturation [102, 250, 251]. Recent studies have employed use of *Agrobacterium tumefaciens* transient expression in a model plant, *Nicotiana benthamiana,* to observe sub-cellular localisation of fluorescently-tagged effectors [113, 252, 253]. This model plant has large and easily transformable abaxial epidermal cells which makes it suitable for subcellular localisation studies. Additionally, *N. benthamiana* is a host to different bacterial, fungal and oomycetes pathogens [196, 197]. Physiologically important host targets can also be identified using a technique called co-immunoprecipitation (coIP). Pathogen and host protein complexes are pulled out using antibodies specific to either the effectors or effector-tagged proteins and then analysed using liquid chromatography-tandem spectrometry [254, 255]. Yeast two-hybrid screen can also be used to identify effectors-interacting proteins [105]. Using this approach three host-targets for AVR-Piz-t were identified, which was essential in determining the role played by this effector during host colonisation [105].

Weeping lovegrass (*Eragrostis curvula*), is valuable as a forage grass for livestock and is sometimes grown for this purpose [256]. This species of grass is highly susceptible to *M oryzae,* and ability of the pathogen to infect this host is controlled by a single gene *PWL2* [118] [117]. *PWL2* was first identified in a genetic cross between two *M. oryzae* strains, a parental strain 4224-7-8, which could infect weeping lovegrass and lacked *PWL2*, and another strain, 6043, which was non-pathogenic and possessed *PWL2* [118]. In the genetic cross (strain 4360), each of the five tetrads had four ascospore progenies that were pathogenic to weeping lovegrass and four that were non-pathogenic [118].

Therefore, tetrad analysis showed that the ability to infect weeping lovegrass was due to a single genetic locus.

In some cases, spontaneous mutant strains lacking *PWL2* were obtained from non-pathogenic strains, and these, could also infect weeping love grass [118]. It was for example possible to obtain spontaneous mutants from non-pathogenic parental strain 6043 at a lower frequency than its non-pathogenic progenies. However, It was not possible to obtain spontaneous mutant from Guy11 (which is the 6043 parental strain) [118]. When rice blast pathogens lacking *PWL2* were transformed with the gene, virulence towards weeping lovegrass was lost but the strains still retained pathogenicity towards barley and rice [118]. These strains did not have any fitness defects but had lost virulence towards a specific host. Most field isolates of the rice blast fungus were shown to possess one or two copies of the *PWL2* gene [117, 118]. It is not clear what role the gene plays in rice blast disease, however, the gene may have an important a role because it occurs in a high percentage of rice blast isolates collected globally [118].

The *PWL2* gene belongs to a family of genes that includes three additional putative effector-encoding genes *PWL1*, *PWL3* and *PWL4* [117], characteristic of a gene family expansion. At the amino acid level, Pwl2 shows 75 % similarity to Pwl1, 51 % similarity to Pwl3 and 57 % similarity to Pwl4. Pwl1 shows 46 % and 50 % identity to the Pwl3 and Pwl4 proteins, respectively. There is 72 % identity between Pwl3 and Pwl4 amino acid sequences [117]. At the nucleotide sequence level, the homology of *PWL1* and *PWL2* starts from 70bp upstream of the start codon through to the stop codon with 78 % identity

overall [117]. *PWL1* was first identified in progeny from crosses between a weeping lovegrass pathogen and a goosegrass pathogen of *M. oryzae* but was cloned from a Finger millet pathogen. *PWL1* controls pathogenicity against weeping love grass. The other genes in this family, *PWL3* and *PWL4* appear to be non-functional, although *PWL4* regains its avirulence function when driven by *PWL1* and *PWL2* promoters [117]. *PWL1* encodes for a slightly larger protein, comprising 147 amino acids (aa) and a molecular mass of 16.2kDa. Pwl3 and Pwl4 proteins are smaller with molecular mass of 14.9kDa (137 aa) and 15.0kDa (138 aa), respectively [117].

PWL2 encodes for a hydrophilic, secreted protein containing 145 amino acids with a molecular weight of 16.16kDa [117, 118]. A guanine to adenine substitution in *PWL2* leading to an amino acid change from aspartic acid to asparagine at residue 90, causes loss of recognition by cognate R gene [118] . The amino acid sequence in the wild type *PWL2* gene product, usually DKS, is altered to NKS, which is a signal-sequence for N-linked glycosylation [118, 257]. In 2010, Schreider *et al* carried out a study to characterise the importance of this allelic variation on Pwl2 protein structure [118]. By expressing and purifying the Pwl2 protein and using spectroscopic techniques to evaluate the structure of the protein, they found that alteration of residue 90 from D90 to N90 (Figure 5.2) caused the protein to be intrinsically disordered, compared to the predicted structure of wild type Pwl2 [118, 257]. The study related this mis-folding to altered posttranslational modification, which leads to lack of recognition by the host R-gene culminating in host immune evasion.

Comparative genomic analysis of two *M. oryzae* field isolates, P131 and Y34, and the reference strain 70-15, identified several genes disrupted by transposon elements (TEs) in all the three strains [43]. Most of these genes encode for hypothetical proteins with unknown functions [43]. In the laboratory strain, 70-15, the avirulence gene *PWL2* was surrounded by TEs within 1 kb of the opening reading frame coding sequence, which explains duplication and expansion events [43]. In comparison to *M. oryzae* isolates, P131 and Y34, this genome had more duplicated genes which included *PWL2* that had two copies annotated in the 70-15 genome. The expansion, or loss of members in gene families, will either be specific to one species or spread across the fungal kingdom [85, 117, 130, 131, 167]. Gene family expansion and diversification can lead to host specification, as observed in Irish potato famine pathogen *Phytophthora infestans* and its sister species *Phytophthora mirabilis*, responsible for infecting *Mirabilis jalapa* [131]. Through diversification of a cystatin-like effector EPIC1, the oomycete could target cysteine proteases from a different host which is thought to have facilitated host jump [131]. However, in most cases, the reason for gene expansion is not understood.

During biotrophic growth, Pwl2, like other AVR proteins and several biotrophy-associated secreted (Bas) proteins is secreted into the biotrophic interfacial complex (BIC) before translocating into the host cytoplasm [1-3]. Accumulation of effectors in the BIC has been routinely used to characterise effectors and study the biotrophic phase of cell invasion by the *M. oryzae* [1, 2]. However, the fluorescence of translocated effectors into cytoplasm is weak

compared to the signal from the BIC. Two different methods have been used to image fluorescence of effector proteins in rice cells. First, imaging a rice cell containing proliferating invasive hyphae secreting fluorescent effectors that are been subjected to plasmolysis can be used to observe effector translocation and accumulation in the host cell cytoplasm [3].

Secondly, expressing effector proteins tagged with a nucleus localisation sequence (NLS) can be used to concentrate the translocated effector into the rice nucleus so that it can be visualised, an indication of movement through the host cytoplasm after uptake to the rice cell [2, 3]. Confocal images of Pwl2 fused to a red fluorescent protein and nucleus localisation signal showed the protein initially accumulating in the BIC before translocating to the nuclei [3]. Using this method, Pwl2 has been shown to translocate into un-invaded neighbouring rice cells ahead of invasive hyphae (Figure 5.1) [3]. The BIC is described as a membrane-rich structure that first appears at the tip of the penetration peg and remains sub-apical to the first invasive hyphal cell when the hyphae transforms to a bulbous growth form [2, 3]. The mechanism by which effectors translocate from the BIC, into host cytoplasm is still under investigation.

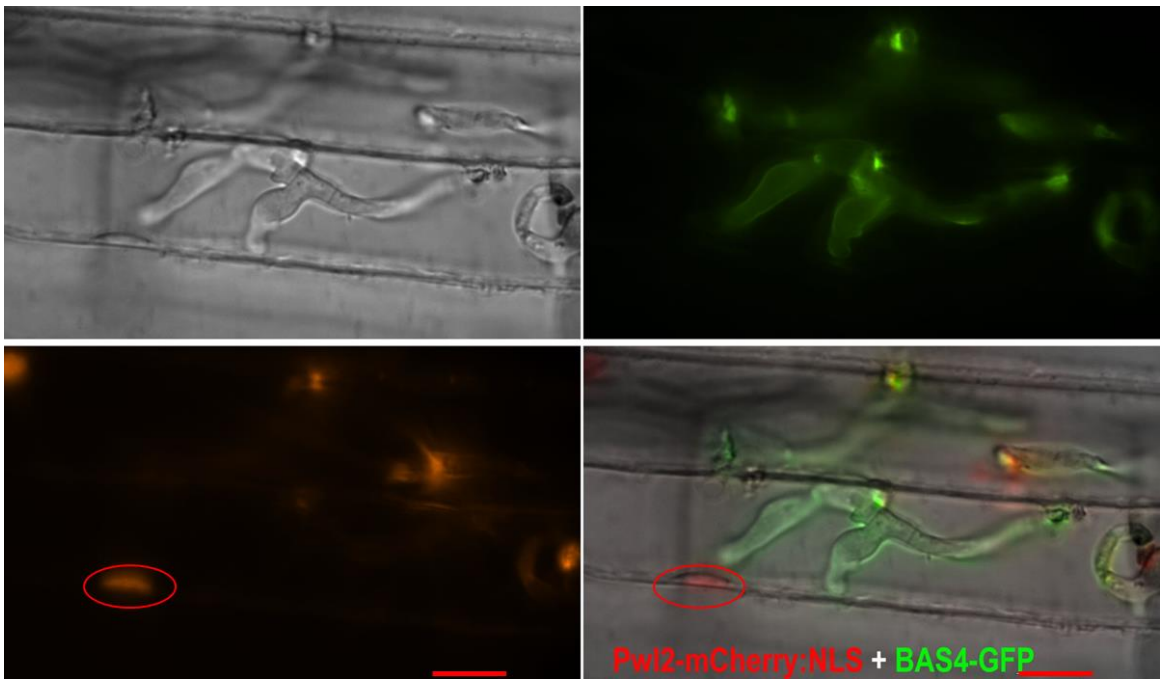

Figure 5.1 Expression of Pwl2-mCherry-NLS and Bas4-GFP at 25 hpi.

The biotrophic phase of growth *M. oryzae* is characterised by the presence of a biotrophic interfacial complex (BIC) in invaded cells, which is believed to be the point of delivery of effectors. Rice tissue infected with *M. oryzae* strain expressing Pwl2-RFP-NLS and Bas4-GFP. Pwl2-RFP (red fluorescence) localises to the BIC in invaded rice cell before translocating into the cytoplasm. The RFP signal was also observed both in the BIC and nucleus of invaded rice cell indicated by red oval shapes. Bas4-GFP (green fluorescence) localises in the space between the fungal cell wall and the host plasma membrane. Scale bar represent 10 µm. Images from this study.

Using *PWL2,* I set out to understand how effectors are expressed *in planta*, secreted into the BIC, and delivered into the host cell. By use of QRT-PCR and laser confocal microscopy I present results here that suggest the expression of *PWL2* starts before the fungus penetrates the host cell, but translation of the protein appear to occur only once the host cell is invaded. I also show that Pwl2-mCherry moves into neighbouring cells as early as 25 h after infection, 1 or 2 h before the invasive hypha invade these cells. More importantly, I have used the CRISPR/Cas9 genome editing approach to generate knockout and *PWL2* null mutants in a Guy11 background that has three genomic copies of *PWL2*.

5.2 Material and methods

5.2.1 General material and methods

For standard procedures used in this chapter see Chapter 2

5.2.2 BLAST analysis

BLAST searches were generated using the BLAST 2.2.22 program [258]. A FASTA formatted text file containing *PWL1-4* coding region nucleotide sequences was used to query a database represented by the isolate genome sequence using a standalone BLASTn program at a probability value of (e-value 1-5).

5.2.3 Laser Confocal microscopy

Laser Confocal microscopy imaging was carried out using a Leica, TCS SP8 motorised inverted laser confocal microscope at X a 63 objective with oil immersion. Lasers were set as follows: GFP and RFP tagged proteins excitation was set 488 and 561 nm laser diodes and the emitted fluorescence detected using 495-550 and 570-620 nm, respectively. The auto-fluorescence from chlorophyll was detected at 650-740 nm.

5.2.4 *In vitro* Cas9-sgRNA RNP synthesis

A sgRNA was designed for CRISPR-Cas9 genome editing using an online tool E-CRISP http://www.e-crisp.org/E-CRISP/. A 20-nucleotide sequence was selected at the *PWL2* locus (not including the PAM NGG-sequence). At the 5' end of this sequence, a T7 promoter sequence was appended and 14 nucleotide overlap sequence added to the 3' end. To synthesise the sgRNA, an EnGen 2 x sgRNA Reaction Mix kit was used following manufacturer's instruction. The synthesis reaction mix was composed of 20 µL of 2 x reaction buffer, 2 µL of 5 pmol/ul target specific oligo and 4 µL enzyme mix then adjusted to a final volume of 40 µL. The enzyme mix contains DNA polymerase that synthesises a dsDNA used as a template from which RNA polymerase then synthesises a guide RNA. This mixture was incubated at 37°C for 1 h before adding 60 µL of nuclease free water and DNAse followed by extra 30 min incubation. To remove any salts, proteins and unincorporated nucleotides, the synthesised sgRNA was purified using Zymo Research RNA clean and concentrator-25 kit. A 208 µL aliquot of RNA binding buffer and 312 µL of 100 % ethanol were added to the synthesised sgRNA and mixed gently. The mixture was then transferred to a spin column and centrifuged for 30 sec. A 400 µL aliquot of RNA preparation buffer was then added to the column and centrifuged for 30 sec. The column was washed twice using RNA wash buffer before eluting in 50 µL nuclease-free water.

The RNP complex formation was made as follows; 6 µg of Cas9, 1.5 µg of sgRNA, 1 x Cas9 Nuclease reaction buffer were mixed and the mixture made up to a final volume of 4 µL with milliQ water. The mixture was incubated at

room temperature for 10 min before fungal transformation into Guy11 protoplast. The RNP complex together with the donor template was mixed with Guy11 protoplasts re-suspended in 150 µl of STC to a concentration of 1 X 10^6, and incubated at room temperature for 25 min. A 1 mL aliquot of PTC buffer was added, mixed by gentle inversion, and incubated at room temperature for 25 min. The mixture was added into molten (45°C) osmotically stable CM (OCM), mixed gently and poured into sterile petri dishes. The plates were incubated at 24°C in the dark for at least 16 h before overlaying with molten complete media (CM) agar containing Hygromycin B to a final concentration of 200 µg mL^{-1} per plate (see Section 2.6).

5.3 Results

5.3.1 Identification, phylogenetic and sequence analysis of *PWL2* in *M. oryzae* isolates

The genome sequences of 91 *M. oryzae* field isolates were analysed for presence or absence polymorphisms in the *PWL* gene family. Nucleotide sequences for *PWL1* (U36923.1), *PWL2* (U26313.1), *PWL3* (1045533) and *PWL4* (1045535) were downloaded from the NCBI database and used to query a database represented by the genome of each isolate, using the Stand-alone Basic Local Alignment Search Tool (BLAST) program (e-value 1-5) [258]. Sequences with identities ranging from 50-80%, 80-100% and 100% were defined as divergent, similar or identical, respectively. Values below 50% were considered to be dissimilar. Wheat and weeping lovegrass pathogens of *M. oryzae* shared a similar effector repertoire with respect to the *PWL* gene family, in that these isolates do not have either *PWL1* or *PWL2* but carry *PWL4*, a member of the gene family that lacks a functional promoter. However, these host species are not closely related and might have diverged further as the pathogens have evolved [259]. Moreover, Wheat and weeping lovegrass pathogens of *M. oryzae* showed close relatedness when analysed phylogenetically as shown in Figure 5.3.There is a possibility that wheat infecting pathogens can also infect weeping lovegrass pathogens.

Rice infecting isolates did not possess *PWL4* whilst 70% possess *PWL3*. All isolates from Sub-Saharan Africa carry *PWL2* (Figure 5.4). In total, more than 95% of rice-infecting isolates analysed in this study carried one or more copies of *PWL2* (Figure 5.3), and strikingly, there was frequent occurrence of a

distinct *pwl2* allele, causing gain of virulence towards weeping lovegrass [118]. This observation was investigated by PCR to confirm the presence/absence polymorphism of *PWL2* in selected isolates across the globe. PCR primers (Table 5.3) were designed to amplify a DNA fragment equivalent to the *PWL2* coding sequence (438bp) and selected isolates produced a single band representing 438bp of *PWL2* (Table 5.1). The amplified PCR products were purified and submitted for DNA sequencing. The resulting sequences were aligned to the wild type *PWL2* sequence using Manager (http://www.scied.com/pr_cmbas.htm).

All isolates from Sub-Saharan Africa carried the wild type *PWL2* except TG004, a strain isolated from Togo. Most isolates from China carried the allelic *pwl2*. Pm1 (Pearl millet pathogen of *M. oryzae*) and Pgky (*Lolium* pathogen of *M. oryzae*) both carried *PWL2*. *PWL1* was cloned from a Finger millet (*Eleusine coracana*) pathogen and I expected Pm1 to have *PWL1* rather than *PWL2* [117, 118]. The sequence amplified from Pm1 showed 100 % identity when aligned to *PWL2* and not *PWL1*. Isolates with *pwl2* allelic variation were mostly from Asia as shown in Figure 5.5, with one Egyptian isolate (EG308) and USA isolate (Jum1) also carrying this allele as shown in Figure 5.4. This observation might be due to selection pressure resulting from field pathogen-host interaction between these isolates and weeping lovegrass, or rice cultivars carrying cognate R-gene [59]. In this type of antagonistic co-evolution, pathogens are under selection pressure to achieve successful interaction while the plants are selected for resistance. Effector gene deletion or allelic variation may result in failure of host detection by resistant cultivars leading to virulence strains [19].

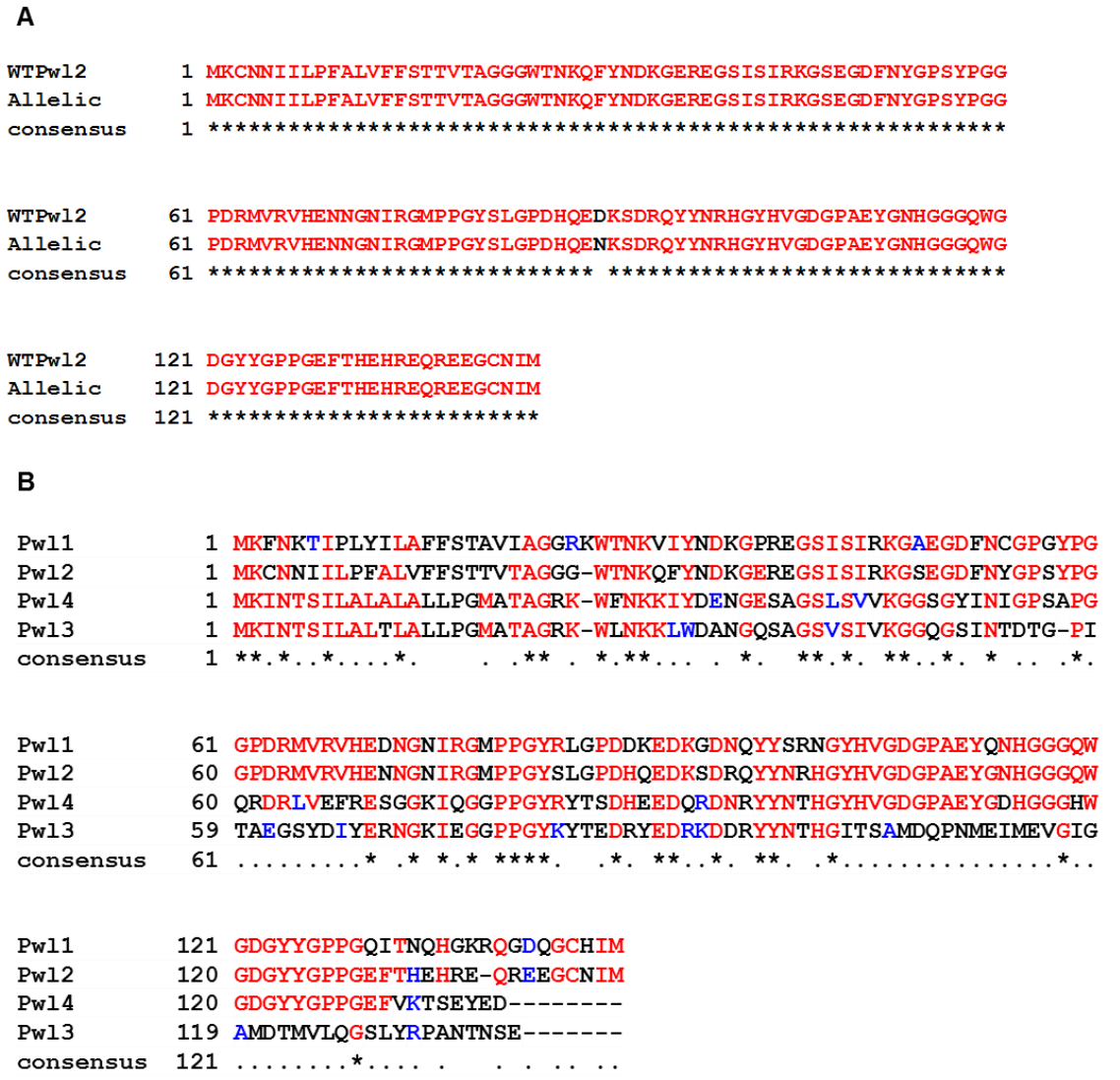

Figure 5.2 Multiple amino acid sequence alignment of *PWL* gene family.

(A) Amino acid sequence of the commonly occurring *pwl2* allele aligned against the wild type *PWL2* from the reference strain 70-15. Residue 90 is altered from D (aspartic acid) to N (asparagine). **(B)** The amino acid sequence of Pwl2 aligned against the Pwl gene family. Regions of identity are indicated by asterisks. Alignments were generated by CLUSTALW.

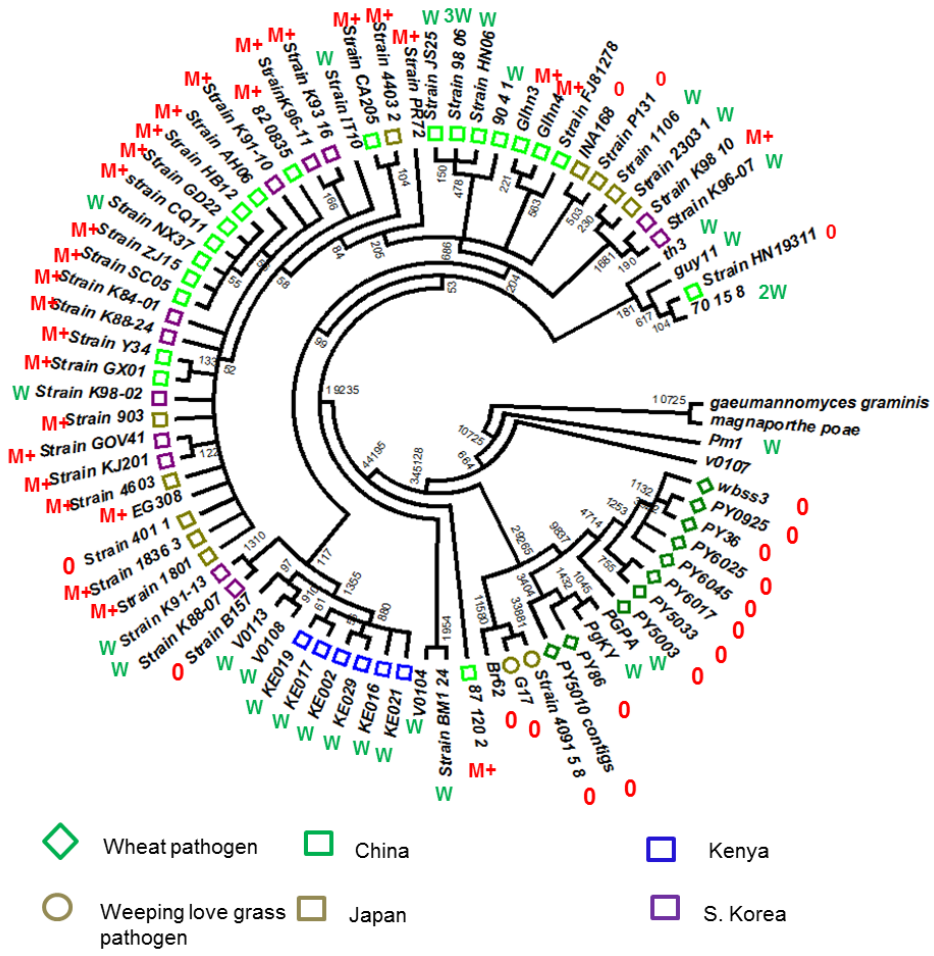

Figure 5.3 Relationship between the frequently occurring allelic variation of *PWL* gene family and *M. oryzae* isolates divergence based on comparative genome analysis.

Maximum parsimony tree of *M. oryzae* strains generated using a standalone, whole genome multiple alignment/SNP calling program, kSNP3. The tree is rooted to two *M. oryzae* related species labelled in yellow (*Magnaporthe poae*) and blue (*Gaeumannomyces graminis*): Presence (wildtype or critical mutation) and the absence of *PWL2* is indicated with **W** representing wild type, **M+** allelic *pwl2* and **0** absence of *PWL2*. Weeping lovegrass pathogens are labelled with pickle green circles while wheat blast isolates are indicated in green diamond shapes. Rice pathogens are labelled with different coloured squares per country of origin; green for China, dark green for Japan and purple for South Korea, blue for Kenya.

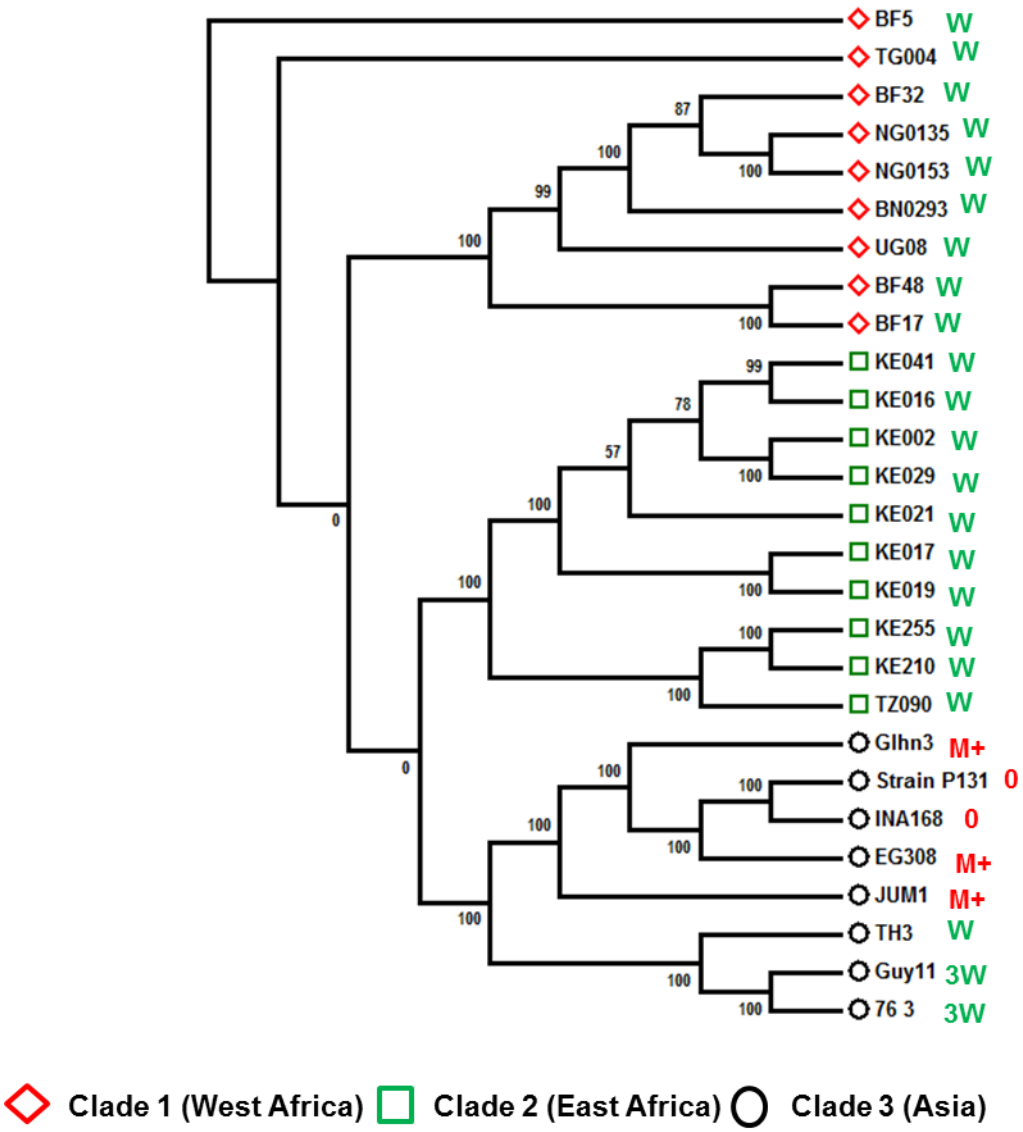

Figure 5.4 *M. oryzae* isolates from sub-Saharan Africa possess *PWL2*.

A Maximum parsimony phylogenetic tree was generated using a standalone, whole genome multiple alignment/SNP calling program, kSNP3. Presence (wild type and allelic copy)/absence of *PWL2* are indicated with **W** for wild type, **M+** for allelic *pwl2* and **0** for absence. Isolates are labelled with red diamonds for clade 1 (isolates from West Africa), green squares for clade 2 (East Africa) and black circles for clade 3 (Asia).

5.3.2 Gene expansion and occurrence of allelic *pwl2* in strains across the globe suggests constant selection pressure

To understand the correlation between the occurrence of *PWL2* in different *M. oryzae* isolates and virulence towards weeping lovegrass, I selected 6 isolates for infection assay. Isolates were selected based on results from a BLASTn search analysis. The wild type Guy11 which carries a wild type *PWL2* was used as a negative control, strain Glhn3 from China that carries an allelic variable *pwl2* was selected to confirm correlation between amino acid substitution and loss of function. Two isolates that lack *PWL2*, INA168 (rice pathogen) and G17 (an aggressive weeping lovegrass pathogen) both from Japan were also selected. Two G17 strains into which *PWL2*-GFP had been introduced, were also used in this experiment. Conidia from these strains were spray-inoculated onto 3 weeks old weeping lovegrass plants, which were then incubated for 7 days at 24°C and 85% humidity until disease symptoms appeared.

The pathogenic interaction was scored as fully colonised and dead leaf tissue after 5-7 days. As expected, weeping lovegrass plants inoculated by Guy11 were not susceptible to blast disease whilst those inoculated with G17 showed severe symptoms (Figure 5.6). Plants inoculated with Glhn3 and Ina168 developed disease lesions but these were not as severe as those observed after G17 infection. This result confirmed *PWL2* as a functional AVR and indicates the presence of a cognate R gene in weeping love grass as reported by Sweigard *et al* [118].

Surprisingly, the two G17 isolates complemented with *PWL2*-GFP construct were still pathogenic on weeping lovegrass. The inability of *PWL2*-GFP to complement wild type Pwl2 phenotype suggests that the GFP fusion protein might interfere with specificity in the C-terminal terminal region of the protein. Pwl2 mature protein is 16.16 kDa in size, while GFP is 27 kDa in size. However, QRT-PCR will be carried to confirm the expression of *PWL2* in generated G17:*PWL2*-GFP strains. The ability of a field isolate Glhn3 that carries the *pwl2* allele to infect weeping lovegrass suggests it may have encountered selection pressure in the field. However, at this point it is not clear whether the R gene imposing such selection pressure occurs only in weeping lovegrass, or if a similar gene is present in some rice cultivars as well. Rice infecting isolates used in this study were isolated from rice leaves, necks or panicles rather than from wild grasses, and there is no information of interactions of the isolates with weeping lovegrass. This may suggest that either some rice cultivars possess the R gene cognate to *PWL2* or that some rice pathogens might infect weeping love grass and other wild grasses. As reported by Sweigard and co-workers [118], there was no phenotypic difference between isolates that lack *PWL2* and those expressing the gene. Two isolates P131, and Ina168, that lack *PWL2* did not have any observable fitness or phenotypic defects and were fully pathogenic on the susceptible rice cultivar *Moukoto*.

Initial efforts to obtain null mutants in either Guy11 or another laboratory isolate Th3 backgrounds were not successful. The 70-15 reference genome assembly was annotated as having two copies of *PWL2*, so I set out to identify the frequency of gene duplication and gene family expansion with regard to *PWL2* in different isolates of *M. oryzae* by Southern blot analysis. The aim

being to identify isolates with one copy of *PWL2* for targeted deletion and characterisation of the null mutant. Genomic DNA of selected strains isolated worldwide was digested using different combination of restriction enzymes; *Afl*II, *Afl*II/*Pst*1, *Pst*1 or *Bam*H1/*Pst*1, fractionated by gel electrophoresis, before being transferred to a Hybond-N membrane. The membrane was probed using *PWL2* coding sequence probe generated by DIG-labelling PCR (Figure 5.7).

In Guy11, the probe hybridised to two restriction fragments of 3.5 Kb and 10 Kb, or 3.5 Kb and 6 Kb, when digested with either *Afl*II in a single digest or a double digest involving *Afl*II and *Pst*1, respectively. In a single digest with *Pst*1, the *PWL2* coding sequence probe hybridised to a single restriction fragment of 6.5 kb. In Th3, the probe hybridised to two restriction fragments of 6 Kb and 6.5 Kb representing two copies of the *PWL2* locus when digested with *Afl*II/*Pst*1. In the Kenyan isolate KE002 and Burkina Faso isolate BF48 the probe hybridised to, three and four restriction fragments respectively, an indication of a larger expansion of the *PWL2* gene in these backgrounds. From BLASTn searches in their genomes, the two *M. oryzae* isolates only showed one copy of *PWL2* and the same was observed in Guy11 and Th3. Moreover, PCR screens could not detect the gene duplication and expansion events. All genome assemblies used in this study were a result of short read sequencing and this probably explains why only one copy of *PWL2* could be detected in these genome sequences using BLASTn searches. In a Chinese isolate 0-137 genomic DNA, *PWL2* probe hybridised to two restriction fragments representing two copies of *PWL2.* The Chinese isolate Y34 showed a single hybridising fragment suggesting a single copy, but this was related to low DNA quality, rather the occurrence of one copy of *PWL2*. No hybridisation occurred for isolates, Ina168 and G17,

confirming the lack of *PWL2* in these isolates as expected from their genome sequences.

This results show that *PWL2* gene duplication and expansion has occurred in most *M. oryzae* isolates but it is difficult to determine accurately by Southern blot analysis. *PWL2* gene has repeated sequences both upstream and downstream, which may be involved in genome and chromosome translocation [117]. Moreover, the analysis of genomes assembled from short read sequencing technologies, indicated that *PWL2* was present as a single copy due to identical sequences upstream and downstream of the coding sequence. This made it impossible to detect the copy number from a BLASTn analysis. Therefore, in order to obtain improved genome assemblies of Guy11 and KE002, single-molecule real-time (SMRT) Pacbio DNA sequencing was used. Long reads obtained using this technology, were used to assemble the genome into fewer contigs to improve genome continuity (Table 3.1).

BLASTn searches for *PWL2* were then conducted in newly assembled genome sequences and surprisingly, three copies of *PWL2* were detected in Guy11 and five copies of *PWL2* in the KE002 genome. In Guy11, there was a duplication of *PWL2* on a single contig, whilst the third copy involved a translocation event resulting in divergence of an allelic *pwl2* copy in a separate contig (Figure 5.5). In the KE002 genome, *PWL2* occurred in four different contigs, with two *PWL2* copies on the same contig and the other three on individual contigs. This provides evidence that *PWL2* is located onto four different loci in KE002 (Figure 5.5). Improved genome assembly therefore

enabled identification of an accurate determination of *PWL2* copy number than was possible using short read sequencing or by southern blot analysis.

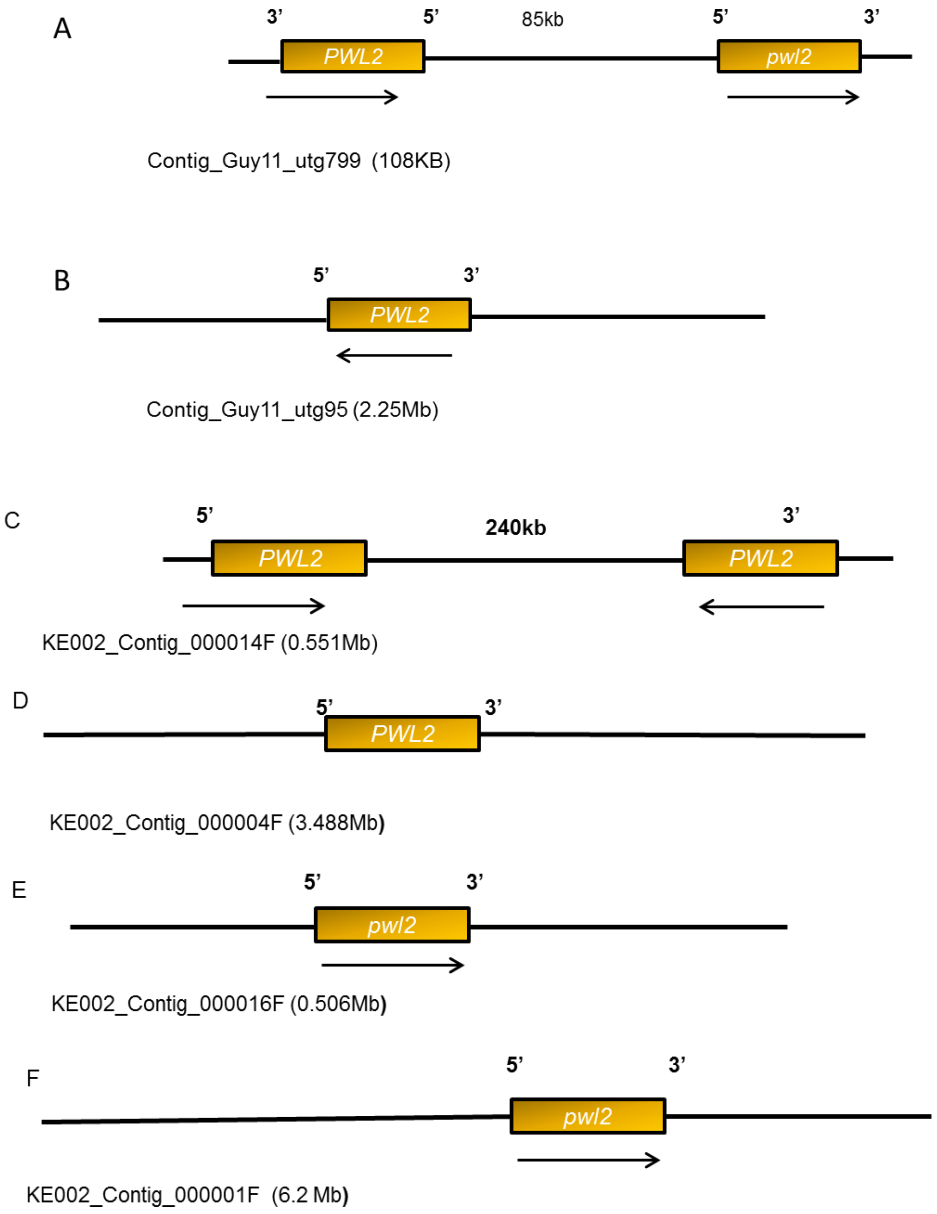

Figure 5.5 Schematic representation of the genetic map around *PWL2.*

A and B represents the estimated chromosomal location of *PWL2* loci on different assembled contigs of Guy11 genome. In A, the separation (85 Kb) distance between the two copies. (C and F) Show estimated the location of *PWL2* on different assembled contigs of KE002 genome. In C there is a separation of 240 Kb distance between the two copies. *PWL2* coding region is indicated as a solid yellow box. Arrows indicate the direction of contig sequence.

Table 5.1 *M. oryzae* isolates used for PCR screen for occurrence of *PWL2*[7]

Isolate	Host	Country of origin	Presence/absence of *PWL2*
KE002	Rice	Kenya	W
UG08	Rice	Uganda	W
TZ090	Rice	Tanzania	W
BF48	Rice	Burkina Faso	W
NG0153	Rice	Nigeria	W
BN0293	Rice	Benin	W
TG004	Rice	Togo	M
JUM1	Rice	U.S.A	M+
GUY11	Rice	French Guyana	W
Th3	Rice	Thailand	W
Glhn3	Rice	China	W
87-120.2	Rice	China	W
82.083.5	Rice	China	M+
O-137	Rice	China	M+
EG308	Rice	Egypt	M+
V0113	Rice	U.S.A	W

[7] Presence (wild type and allelic copy)/absence of *PWL2* are indicated on each strain with **W** representing wild type, **M+** critical **M** for different un-characterised mutation.

Y34	Rice	China	**M+**
PGKY	*Lolium*	U.S.A	**M**
PM1	Pearl Millet	U.S.A	**M**

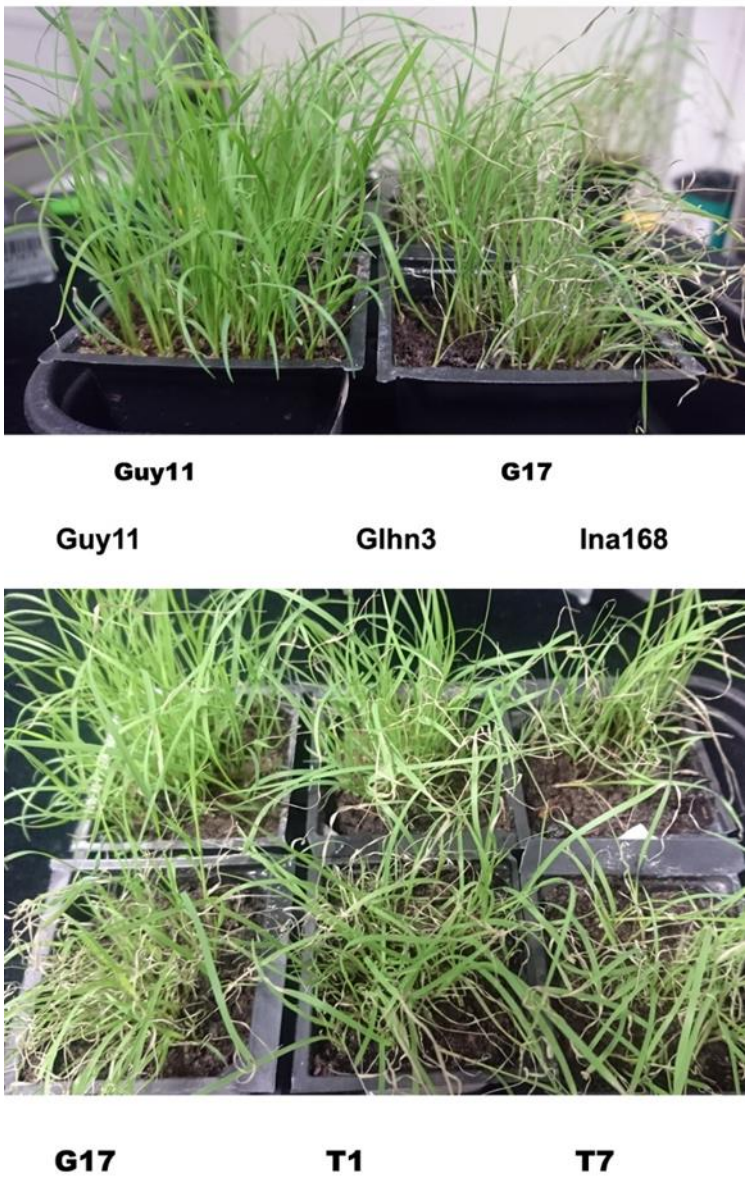

Figure 5.6 Weeping lovegrass infected with different *M. oryzae* isolates 7 days' post-infection.

Un-infected weeping lovegrass (*Eragrostis curvula*) shown at top left were inoculated with Guy11 containing *PWL2*. Plants shown on the top right panel were infected by a pathogenic strain of *M. oryzae* G17, and produced symptoms of infection (brown and were shrivelled leaves). Plants inoculated by Ina168, Glhn3 and G17:*PWL2*-GFP transformants (as labelled) produced less severe lesions as indicated. Two pots of weeping lovegrass seedlings were sprayed with different *M. oryzae* strains in three replicates. Observations were consistent in 3/3 infection replicates.

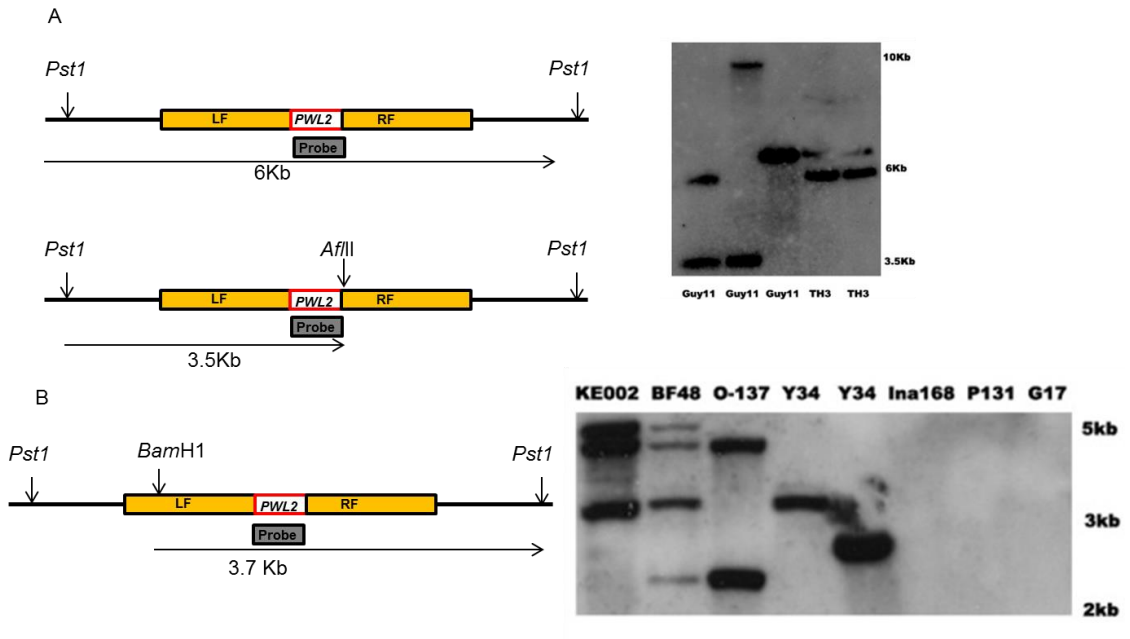

Figure 5.7 Southern blot analysis for presence of *PWL2* in selected *M. oryzae* isolates.

(A) Genomic DNA from Guy11 and Th3 was digested either using *Afl*II in a single digest or a double digest involving *Afl*II and Pst1 and probed using the 438 bp *PWL2* gene fragment. (B) Selected *M. oryzae* genomic DNA digested using *Bam*H1/*Pst*1 and probed using the *PWL2* 438 bp probe. The order of digested genomic DNA from different isolates is indicated on the lanes. Size estimates are from the *PWL2* loci annotated 70-15 genome.

5.3.3 Amino acid substitution in the allelic Pwl2 does not affect expression and accumulation to the BIC

The occurrence of gene deletions, point mutations, chromosomal rearrangements and transposable elements insertions in avirulence genes can benefit the pathogen in terms of evading recognition from a host plant [3, 42]. I hypothesised that amino acid substitutions observed in *PWL2* may be directed towards preventing detection or preventing expression of the protein thereby leading to a gain of virulence. To investigate this, I made two constructs with *PWL2* (wild type) and a chimeric allelic *pwl2* to express a green fluorescent protein (GFP) as a C-terminal gene fusion. The chimeric allelic *pwl2* was generated using PCR site-directed mutagenesis and the plasmid construct sequenced to confirm nucleotide substitution. Primers used to generate this mutation are listed in Table 5.3. The wild type *PWL2* and *pwl2* allele under control of their native promoters were cloned into a fungal transformation vector, pCB1532, including GFP (*PWL2*promoter; *PWL2cds;* GFP). The constructs were transformed into a *M. oryzae* strain, Ina168, which lacks *PWL2* and transformants selected on sulfonylurea. Positive transformants were used to inoculate rice leaf sheaths which were observed by epifluorescence microscopy after 30 h.

These two strains were analysed for host cell expression and localisation of Pwl2. The strains with wild type Pwl2 localised in the BIC as expected, and the same was observed in the strains with allelic Pwl2. The intensity of fluorescence and time of expression were all similar in strains expressing either wild type *PWL2* or allelic *pwl2*. This result suggests that expression, localisation

into the BIC and translocation of *PWL2* into the host cytoplasm is not interfered with by this mutation. I conclude that, due to selection pressure, an amino acid substitution at Pwl2 (N_{90}) might interfere with recognition of Pwl2 mature protein by its cognate R gene but the protein will still translocate into the host cell and serve a yet unknown function.

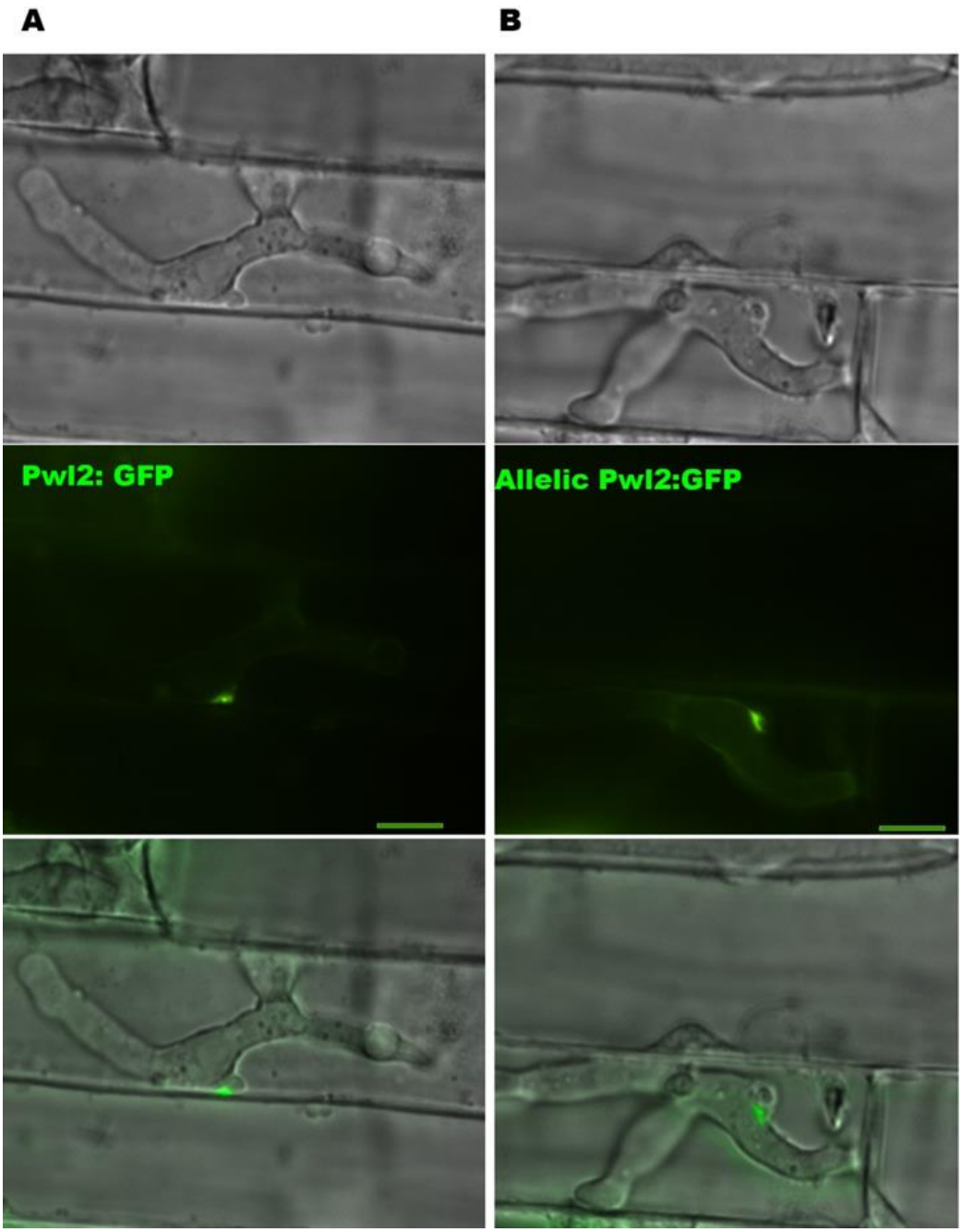

Figure 5.8 Expression and localisation of allelic Pwl2 in the BIC.

A susceptible *Moukoto* rice leaf sheath infected with *M. oryzae* strains, Ina168 expressing Pwl2-GFP and Ina168 strain expressing allelic Pwl2-GFP at 30 h post-infections (A) Shows localisation of the wild type Pwl2 into the BIC. In (B) show localisation into the BIC for the allelic Pwl2. Scale bar represent 10 µm.

5.3.4 Using the expression of *PWL2-GFP* to investigate the biotrophic stage of *M. oryzae* infection

5.3.4.1 Two-stage transcription and translation of *PWL2* pre-and post-penetration, and cellular localisation

The environmental signal that triggers expression of fungal effectors is thought to be the point at which when *M. oryzae* conidia encounter the leaf surface and the fungus penetrates host tissue using the appressorium. Expression of *PWL2* was not observed when germinating conidia of the *PWL2-GFP* strain was observed on a hydrophobic cover slip (data not shown). Expression of *M. oryzae* effectors occurs after rice cell penetration and no expression is observed in fungal mycelium. This means recognition of the rice leaf surface appears to be essential for activation of effector transcription. Analysis of leaf-sheath tissue infected with *M. oryzae* at 36 h post-infection was previously used to identify *in planta* expressed genes during invasive hyphae proliferation which resulted in identification of four effector gene [1].

Previous studies have shown that expression of most effectors, including *BAS4*, *AVR-Pita* and *MEP3* commences at 24 h, a point at which the fungus has penetrated the rice cell (Yan and N.J. Talbot, unpublished observations). In this chapter, the expression profile of *PWL2* was used to investigate the relationship between infection related development and transcription of effector encoding genes. QRT-PCR was carried out to quantify *PWL2* expression at different time points of infection. I reasoned that it would prove difficult to detect expression of *PWL2* before the *M. oryzae* has penetrated the host cell (before 24 h). To overcome this limitation, a leaf drop infection assay was used to

improve the fungal: plant material ratio. Leaves of a susceptible rice cultivar CO39 leaves were inoculated with freshly harvested Guy11 conidia and infected material collected at different time points (16, 17, 18, 20, 22 and 24 h).

Samples were collected from inoculated leaf drops and 100 mg infected plant material used to extract RNA (see Section 2.7.1). The cDNA synthesised from each sample was used to perform Real-time PCR. Samples from *M. oryzae* mycelium were used as a control. Primers were designed to amplify a region spanning 200 bp of *PWL2* coding region (see Table 5.3). Serial dilution of synthesised cDNA was used to test for primer efficiency and the *M. oryzae* housekeeping actin encoding gene (MGG_03982) was amplified as a normalisation signal [1]. Following amplification, generated CT values were normalised against the house keeping actin encoding gene (MGG_03982) and fold change determined using the formula $2^{\Delta\Delta CT}$, where $\Delta\Delta$Ct = ((CtGOI - Ct Actin in GOI) – (CtNC - Ct Actin in NC)) and GOI is the gene of interest and NC is the negative control of mycelial cDNA. The expression of *PWL2* could be detected as early as 17 h post-infection (hpi) which is more than 5 h before host cell penetration, which estimated to occur between 23-25 h (Figure 5.9).

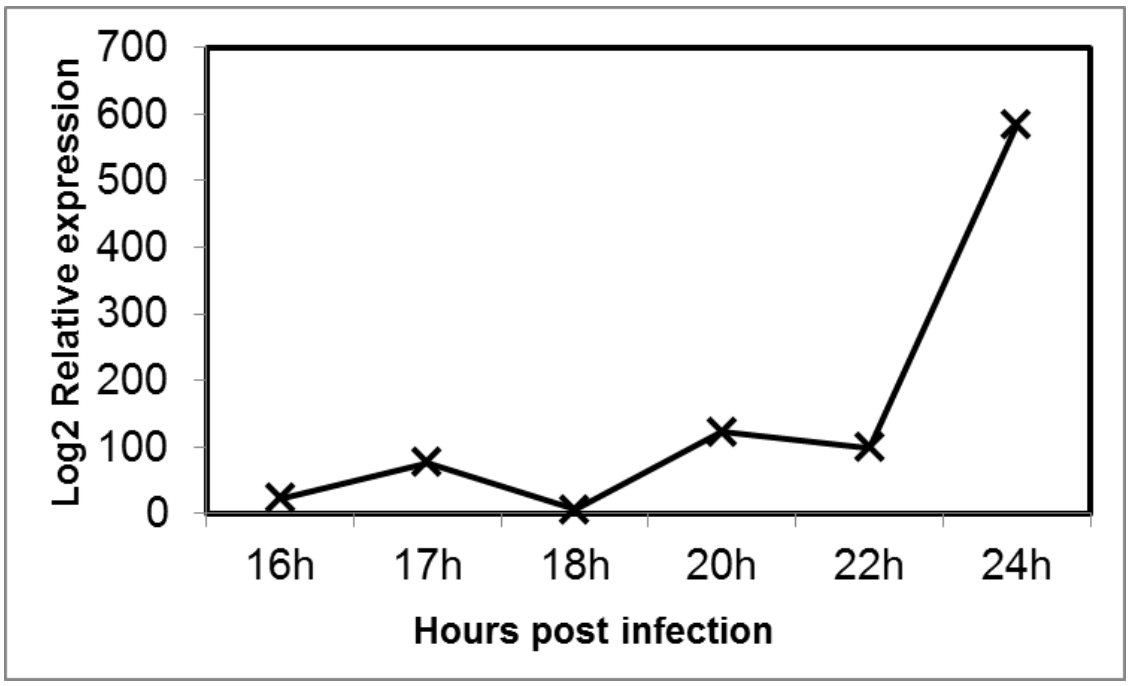

Figure 5.9 Relative expression of *PWL2* during biotrophic growth of Guy11.

Representation of log_2 relative expression levels of *PWL2* in rice leaves infected by *M. oryzae* isolate Guy11 and leaf material sampled at 16h, 17h, 18h, 20h, 22h, 24h post-infection (hpi). Expression of *PWL2* starts at 17 hpi with maximum expression observed at 24 hpi. Normalisation of data was carried out using housekeeping actin encoding gene. RNA from fungal mycelium was used as a control. Log_2 relative expression is represented on the Y-axis while h post-infection at 16h, 17h, 18h, 20h, 22h, 24h is shown on the X –axis.

The effector, Pwl2 has been shown to move into neighbouring rice cells ahead of *M. oryzae* invasive hyphae [2]. I hypothesised that this effector is essential for invasive hyphae proliferation and maybe secreted into the first invaded host cells before the penetration peg. To investigate into this, a Guy11 strain expressing *PWL2-mCherry*, fused to a nucleus localisation sequence (NLS) at the C-terminus of the fusion protein and also expressing *BAS4-GFP*, was used to inoculate leaf sheath of a susceptible rice cultivar, *Moukoto*. The NLS directs accumulation of the fusion protein to the nucleus [2]. Infected rice leaf sheath was examined by epifluorescence microscopy between 24 - 30 h post-infection. Accumulation of Pwl2 in the nucleus was observed as early as 25 hpi which is 1 to 2 h after the penetration event (Figure 5.10). At 26 h post-infection, partial accumulation of Pwl2-mCherryNLS into neighbouring cells was observed. However, no fluorescence was observed in non-invaded plant cells prior to the formation of either a penetration peg or invasive hyphae. Non-invaded cells did not show fluorescence of Pwl2-GFP in either conidium, appressorium, host cell cytoplasm or nuclei. Therefore, expression of *PWL2* starts immediately upon contact with the leaf surface, but translation into mature Pwl2 protein appears to occur in host cells. It also suggests that the invasive hyphae, and specifically the BIC, are responsible for effector Pwl2 secretion and delivery in host cells.

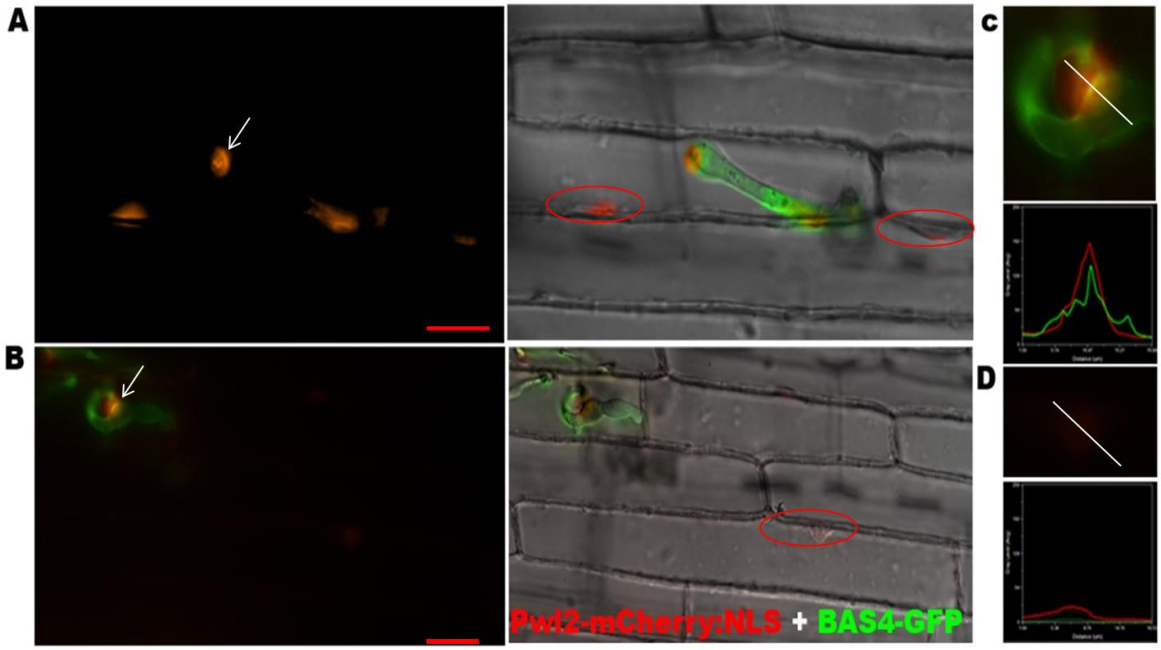

Figure 5.10 Pwl2-mCherry-NLS used to predict time of effector translocation in cytoplasm.

M. oryzae strain expressing *BAS4-GFP* and *PWL2-mCherry* visualised in leaf sheath of a susceptible rice line *Moukoto*. Arrows indicate the position of the BIC and the oval shapes indicate rice nuclei. (A) Bright Pwl2-mCherry fluorescence occurring in the nucleus of invaded and un-invaded cell at 25 hpi. (B) Faint fluorescence occurred in the nucleus of neighbouring uninvaded cells 25 hpi. (C) Fluorescence intensity scan for the BIC in B (D) Fluorescence intensity scan for the nucleus of un-invaded cell. Scale bar represent 10 µm.

5.3.4.2 Expression of Pwl2 into colonised rice cells provides an insight into direction and mechanism of effector delivery

The BIC has been described as a plant-derived structure located outside the fungal cell wall. This was demonstrated using a fungal strain expressing Pwl2-mRFP and the *M. oryzae* plasma membrane ATPase Pma1-GFP [2]. Fluorescence of Pwl2 did not co-localise with A Pma1-GFP but co-localised with the rice plasma membrane marker LTi6B-GFP when the strain was visualised in a transgenic rice line (Figure 5.11) [2]. In a living rice cells, invasive hyphae are enclosed by an extra-invasive hyphal membrane (EIHM) which remains intact before the invaded rice cell dies followed by rupture and collapse of the rice plasma membrane [93]. Recent studies have also shown that newly invaded cells remain viable until *M. oryzae* leaves the first invaded cell and moves into neighbouring cells [93]. It is not clear whether contents of the BIC are exclusively fungal secreted proteins or if some of the content is derived from the host cytoplasm. It is also possible for example that Pwl2 is secreted into rice cells and then sequestered into the BIC from the rice cell, rather being secreted into the BIC from fungal invasive hypha.

To attempt to answer this question, leaf sheath infection on a susceptible rice cultivar *Moukoto* was carried using two different strains – Ina168 expressing Pwl2-GFP and Guy11 expressing Pwl2-mRFP and observed by laser confocal microscopy for cells invaded by both strains 25-36 h post-infection. At 30 hpi, the host plasma membranes surrounding the BICs were still intact, and there was no co-localisation of Pwl2-GFP and Pwl2-mRFP in either of the BICs (Figure 5.12). In these cells, each invasive hypha was shown to

secrete either green or red fluorescently labelled Pwl2 separately. This result confirms the observation that the BIC is a host derived structure and that the movement of Pwl2 is in one direction – from the fungus to the BIC and into the host cell.

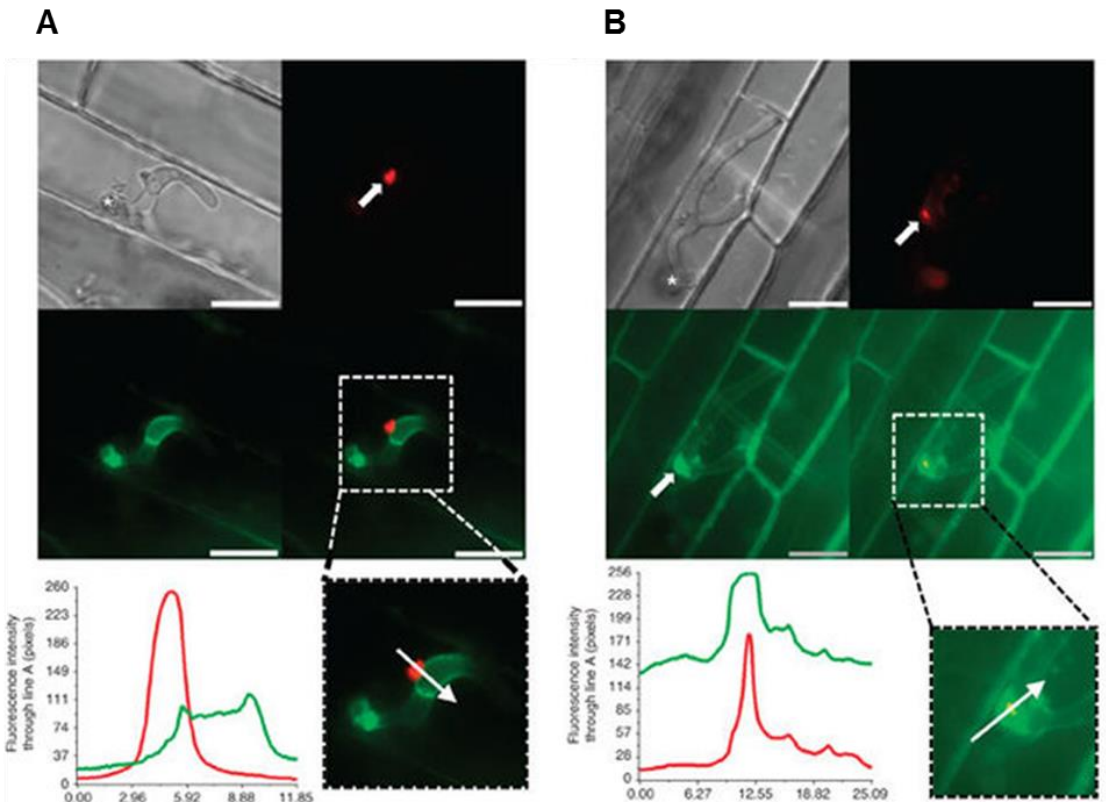

Figure 5.11 The BIC is a plant derived structure.

The BIC contains secreted proteins from *M. oryzae.* (A) A strain expressing *M. oryzae* ATPase Pma1-GFP and Pwl2-mRFP did not show co-localisation of these fluorescent proteins with fluorescence intensity scan shown on the bottom left (B) In a transgenic rice line expressing LTi6B-GFP, the BIC containing Pwl2-mRFP co-localises with fluorescence LTi6B-GFP, a rice plasma membrane marker. Fluorescence intensity scan is shown on the bottom left. Scale bar represent 10 µm. Figure from Giraldo *et al* [2].

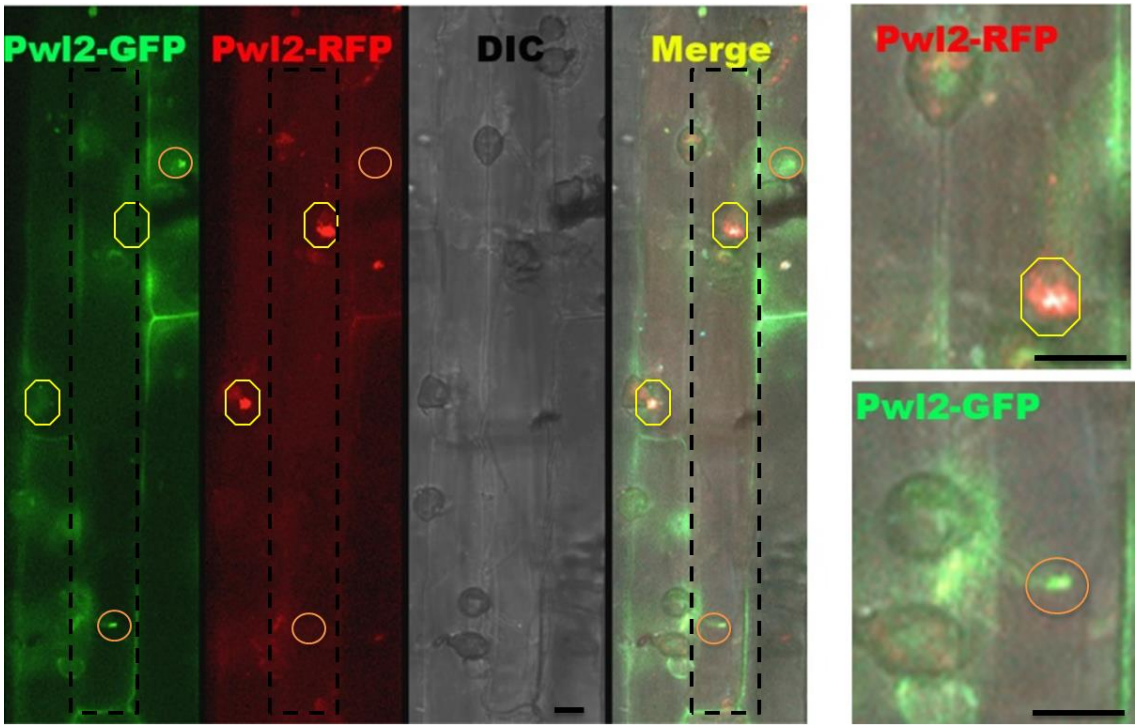

Figure 5.12 *S*ecretion and translocation of *M. oryzae* effectors into the BIC and host cytoplasm during rice blast infections.

Moukoto rice leaf sheath was co-infected with *M. oryzae* Guy11, expressing Pwl2-mRFP and Ina168 expressing Pwl2-GFP at 30 hpi. BICs are labelled by cycles for GFP and octagons for mRFP. Lack of co-localisation between Pwl2-mRFP and Pwl2-GFP, indicated by separate red and green fluorescence confirms the BIC does not contain Pwl2 sequestered from rice cell. Dotted lines indicate a single rice cell co-infected by two different *M. oryzae* strains. Scale bar represent 10 µm.

Like most cytoplasmic effectors, successful secretion of Pwl2 appears to require two exocyst components, Exo70 and Sec5. The Exo70 and Sec5 proteins are involved in tethering secretory vesicles to the fungal plasma membrane. In *Δexo70* and *Δsec5* mutants, cytoplasmic effectors are not delivered into the BIC, but are partially retained inside the BIC-associated IH cells [2]. However, secretion of apoplastic effectors is normal in these mutants [2]. Genes associated with polarised growth and secretion have also been demonstrated to be involved in effector delivery [2]. A t-SNARE protein Sso1 for example is required for efficient effector secretion into the BIC. Together with Snc1, these proteins mediate vesicle docking and fusion to the plasma membrane [2]. This means that BIC-associated cells are responsible for effector secretion via a Golgi-independent secretory process [2].

In this study, high-resolution microscopy was used to study secretion of effectors into the BIC during biotrophic growth. In order to gain more insight into how effectors are packaged, secreted into the BIC and translocated across the plant plasma membrane into the host cytoplasm, a *M. oryzae* strain expressing two BIC localising effectors Pwl2-mRFP and Bas1-GFP was generated and I compared localisation of these effectors at 36-h post-infection. To generate this strain, a BIC-localising effector Bas1 expressing GFP at the C-terminal (*BAS1p: BAS1-GFP*) - was transformed into a Guy11 strain expressing Pwl2-mRFP. This strain was used to infect a susceptible rice cultivar *Moukoto* leaf sheath tissue and accumulation of effectors was imaged by laser confocal microscopy. All microscopic observations of live leaf sheath were carried out from 25-36 h post-infection. To clearly differentiate fluorescence emitted by different vesicular structures, a sequential scanning at different position of the sample was carried

out and Z-stack images obtained. Sections that produced unfocused images were discarded.

Fluorescence of the two effectors was observed as small punctate signals in the BIC, suggesting that they translocate through the fungal plasma membrane as extracellular vesicles. However, there was no co-localisation of Pwl2 and Bas1 fusion proteins. This observation suggests that mature proteins of the two different effectors are packaged differently and cross the fungal plasma membrane, as separate extracellular vesicles despite both being secreted through the same exocyst-dependent mechanism [2]. Fluorescence intensity of Pwl2-mRFP was constantly higher that Bas1-GFP in all the analysed invasive hypha. The function of these two effectors remains unknown, but both move into uninvaded cells before invasive hyphae [3].

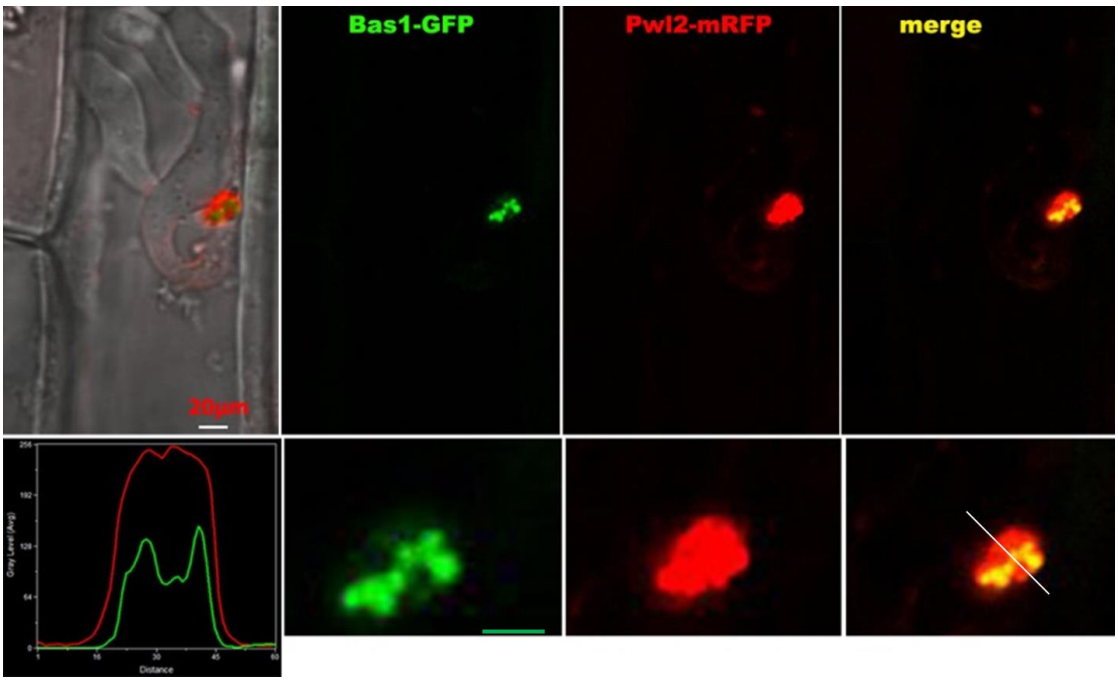

Figure 5.13 BIC contains expressed effectors in form of extracellular vesicles.

Rice leaf sheath of a susceptible *Moukoto* was infected with a *M. oryzae* strain expressing Pwl2-mRFP and Bas1-GFP. No co-localisation between Pwl2-mRFP was observed and Bas1-GFP was observed as indicated by fluorescence intensity distribution linescan at bottom left. Scale bar represent 20 µm.

5.3.5 CRISPR/Cas9 genome editing approach for targeted deletion of multiple *PWL2* copies

In recent years, gene editing mediated by clustered regularly interspaced short palindromic repeats (CRISPR)-associated with a DNA endonuclease and guided by RNA has been used in genome editing. CRISPR/Cas9 has become a powerful tool that can be used for genome editing in mammals, plants and filamentous microbes [260-262]. In this technology, a single chimeric RNA (sgRNA) is used to guide the Cas9 endonuclease to a DNA sequence of interest, allowing Cas9 to introduce a double strand break (DSB) at this locus [260-262]. The sgRNA contains a protospacer sequence that defines the target DNA [260-262]. The sgRNA can essentially be a 17-20 bp nucleotide sequence found adjacent to a 5' PAM sequence [263]. The PAM sequence is required for Cas9 to recognise a target sequence. Genome editing by CRISPR/Cas9 relies on a double strand break repair mechanism (Figure 5.14) [263].

The break caused by Cas9 is repaired by the non-homologous end-joining (NHEJ) mechanism that is error prone and leads to deletions or insertions within the target sequence [263], which may result in loss of function of a targeted gene [263]. If a donor DNA (DNA sequence with homology near the double strand break is also introduced, organisms can repair the break by homologous recombination (HR) allowing insertion of the donor DNA [263]. This has enabled the use of CRISPR/Cas9 for gene deletion, insertions or targeted gene replacements [263]. Moreover, gene editing using this technology has another advantage; it can be used marker-free and can therefore replace marker-based gene deletions [264]. Recently Pohl *et al* [264] has shown that

Cas9 protein can be delivered in fungal protoplasts together with an *in vitro* synthesised sgRNA (targeting the *pks17* gene) and a selectable marker in form of *amds* cassette flanked with sequences homologous to *pks17*. They showed that the efficiency of homology directed repair (HDR) was increased when a selectable marker was used [264]. The study concluded that genome editing can be carried out using preassembled CRISPR-Cas9 ribonucleoproteins (RNPs) as already reported for animals and plants [264].

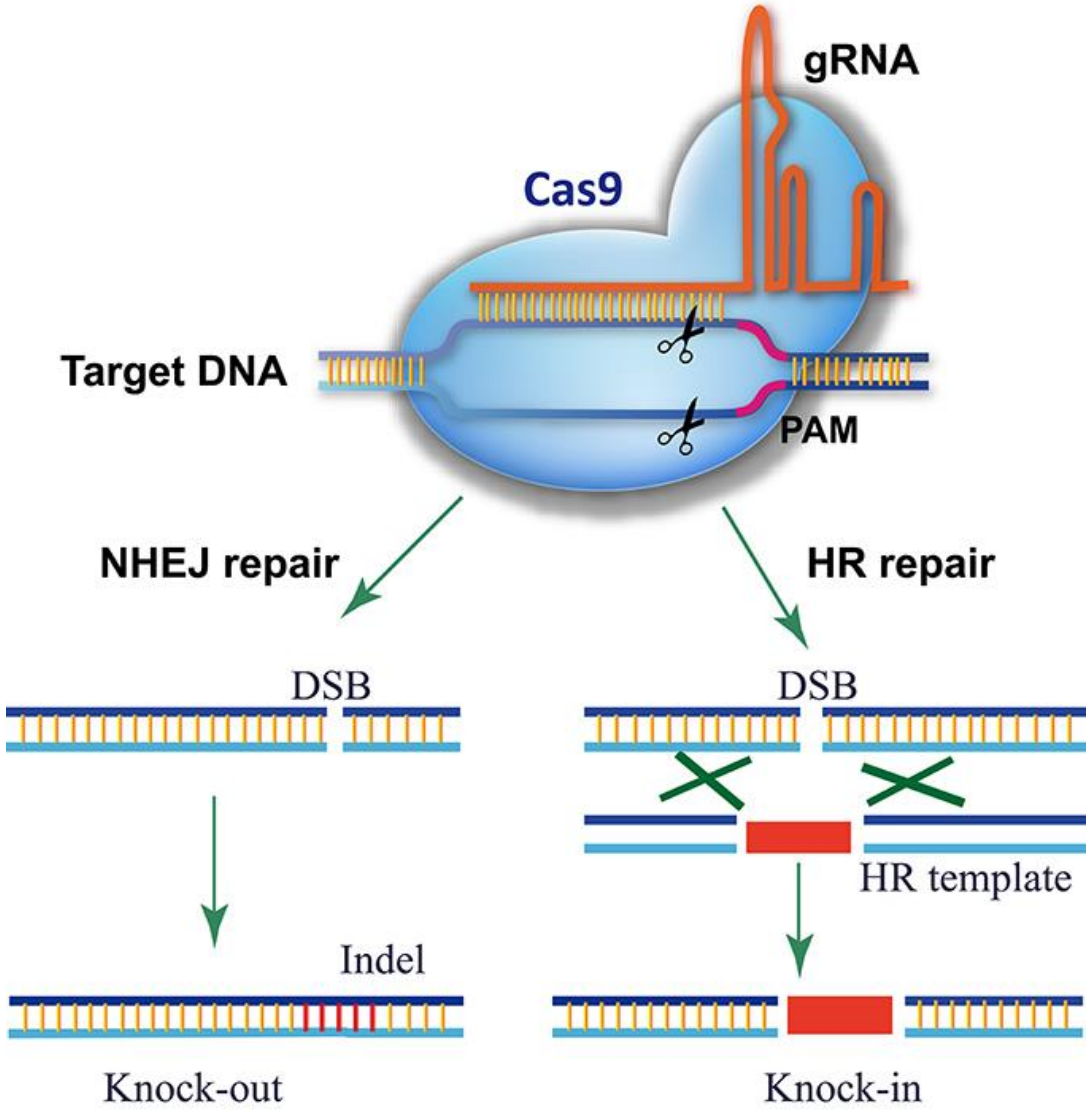

Figure 5.14 Schematic illustration of genome editing using Cas9/gRNA.

Base pairing between the gRNA and target DNA directs Cas9 to this target. Downstream of the gRNA-binding region is a PAM motif that is required for Cas9 recognition and cleavage. A cut by Cas9 triggers an endogenous double stranded break (DSB) repair that results in a knockout via error-prone NHEJ pathway that introduces an indel and causes loss of gene function. In presence of donor template, the break is repaired by homology directed repair resulting in the integration of donor DNA. Figure from Ding *et al* [265]

Earlier studies of *PWL2* have shown that strains that are non-pathogenic to weeping lovegrass can become pathogenic by spontaneous mutation. Spontaneous mutants could be obtained from a parental strain 6043 at a lower rate than from progeny resulting from crosses generated by 6043 [118]. According to Chang *et al,* attempts to generate spontaneous mutants from Guy11 (or the 6043 parental strain) were not successful [118]. Similarly, in this study, several initial attempts to delete *PWL2* from Guy11 were not successful. Using Southern blot analysis, I observed that Guy11 appeared to possess two copies of *PWL2* (Figure 5.7). The well characterised *M. oryzae* genome 70-15 possesses two copies of this gene that are annotated as *MGG_04301* and *MGG_13863* [24].

The reference genome 70-15 is the result of a cross between Guy11 and a weeping love grass pathogen, followed by backcrossing with Guy11 [24]. This may explain the two copies identified in Guy11, although Illumina sequencing of wild Guy11 suggested a single copy of *PWL2*. We re-sequenced Guy11 using single-molecule real-time (SMRT) Pacbio sequencing which has provided a more contiguous genome sequence (see Table 3.1). Analysis of this improved genome sequence has revealed that there are in fact three copies of *PWL2* in Guy11, of which one was not annotated in the 70-15 or previous Guy11 genome assemblies, and not elucidated by Southern blot or PCR analysis. This may explain the difficulty in generating null mutants for this gene and might also explain the inability to generate spontaneous mutants in Guy11, as reported in Chang *et al* [117].

To overcome this challenge, CRISPR/Cas9 was used to introduce a double strand break in the *PWL2* gene in combination with PCR-based marker-assisted gene replacement. The donor DNA consisted of the hygromycin resistance gene cassette fused with sequences flanking the *PWL2* gene to enhance homologous recombination DNA repair and facilitate gene replacement (Figure 5.15). Briefly, sequences upstream and downstream of *PWL2* coding sequence were PCR-amplified using primers KOLFF, KOLFR, KORFF and KORFR (Table 5.3) designed to include overhangs complementary to a fragment of Hygromycin phosphotransferase resistance gene cassette. These upstream and down-stream flanking regions of *PWL2* gene were fused with either overlapping fragments of hygromycin resistance gene cassette at regions with overhang sequences generated in the first-round of PCR using primers KOLFF, HY, YG and KORFR (Figure 5.15). To increase the efficiency of the repair with the donor DNA, a third-round PCR was carried out to make one fragment with 766bp sequence upstream and 892bp downstream of the *PWL2* flanking the hygromycin resistance gene cassette using primers KOLFF and KORFR. Regions surrounding the *PWL2* gene allow for homologous combination during homology-directed repair.

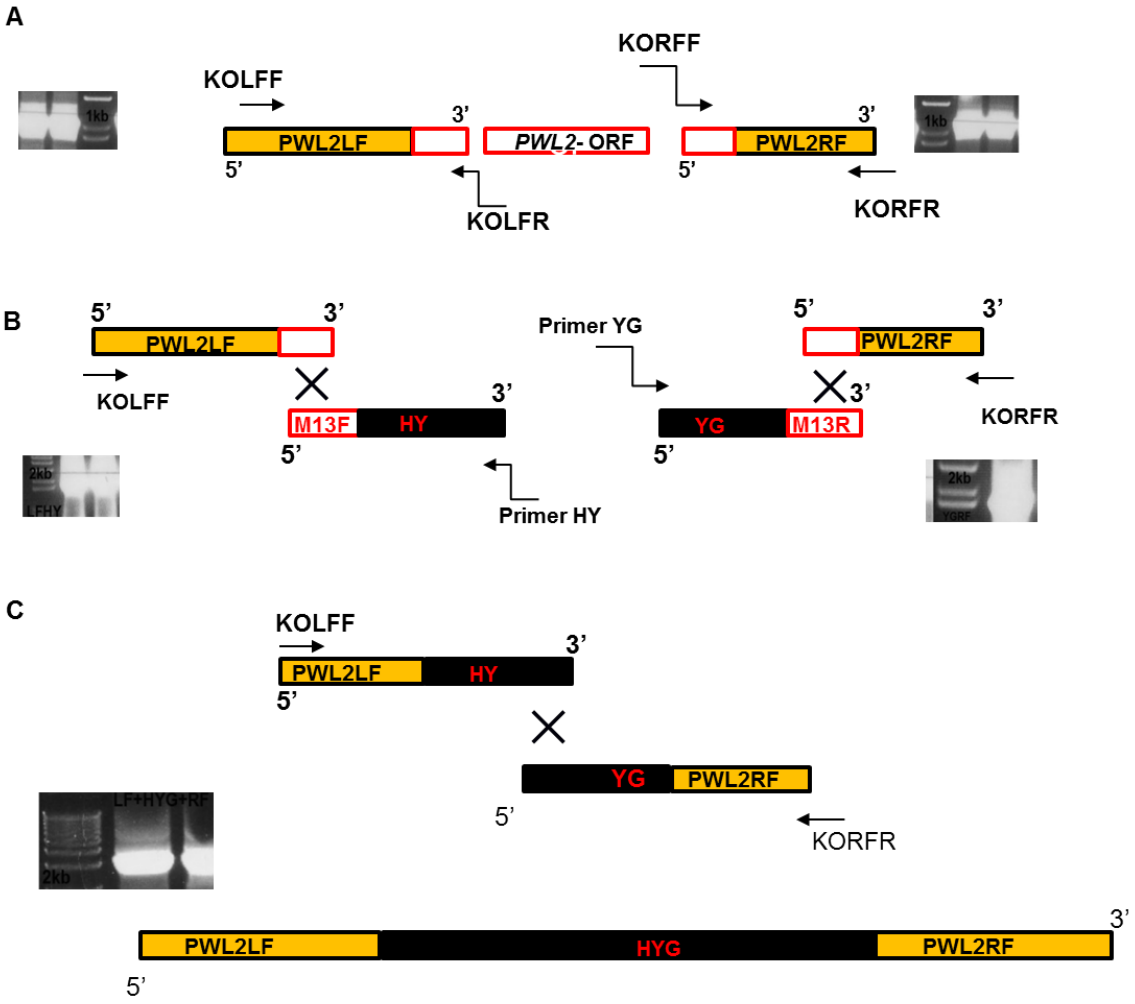

Figure 5.15 Schematic representation of PCR-based split-marker deletion via CRISPR/Cas9 induced homology directed repair.

(A) In the first-round PCR, 766bp sequence upstream and 892bp downstream of *PWL2* coding region were amplified using the primers as shown. (B) In the second-round PCR, the *PWL2* left flank was fused to one half of hygromycin gene cassette (HY) using a reverse primer that generated overhangs complementary to sequences in the YG fragment. The *PWL2* right flank was fused to the other half of Hygromycin gene cassette (YG) using a forward primer that generated overhangs complementary to sequences in the HY region. (C) In the third-round PCR, the second-round PCR products were fused together and used for fungal transformation as a donor template.

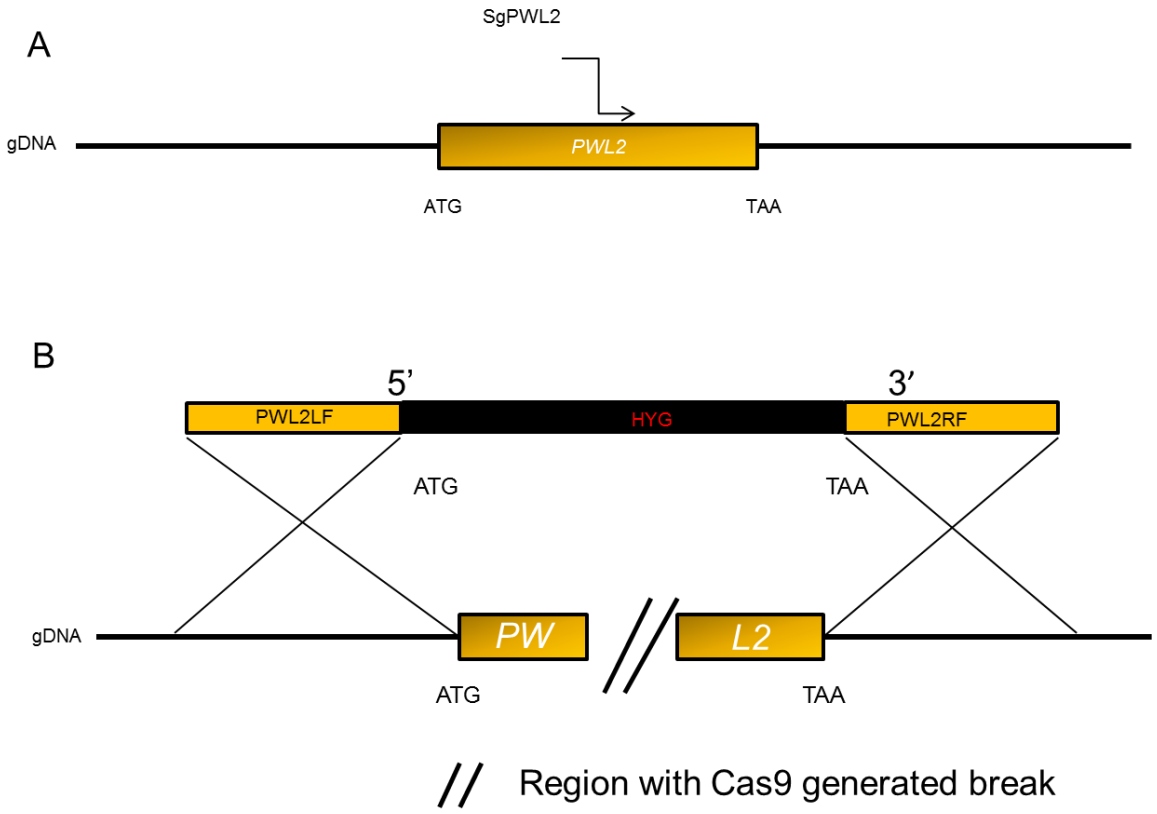

Figure 5.16 Schematic illustration of inserting Hygromycin gene cassette at the *PWL2* locus using CRISPR-Cas9 gene editing.

(A) A guided sequence (sg*PWL2*) was used to direct Cas9 and introduce a double strand break at the *PWL2* locus. (B) The DNA repair template constitutes the Hygromycin resistance gene cassette and flanking regions of the *PWL2* gene**.**

The sgRNA was designed using an online tool E-CRISP http://www.e-crisp.org/E-CRISP/. A 20-nucleotide sequence was selected from the *PWL2* locus (not including the PAM NGG-sequence). At the 5' end of this sequence, a T7 promoter sequence was appended and 14 nucleotides overlap sequence added at the 3' end. To synthesis the sgRNA, EnGen 2 x sgRNA Reaction Mix kit was used, as described in methods Section 5.2.4. The enzyme mix contains a DNA polymerase that synthesises a dsDNA that is used as a template from which RNA polymerase synthesises a guide RNA. The RNP complex was generated as described in Section 5.2.4. After incubation at room temperature for 10 min the RNP was added to 150 µl Guy11 protoplasts in STC (concentration of 1 X 10^6), was mixed with 4 µL of RNP complex and 4 µg of donor DNA, after which a transformation and regeneration was carried out, as previously described.

Putative positive transformants were selected on hygromycin and DNA extracted. Transformants were analysed by PCR using primers designed to amplify the *PWL2* coding sequence from 140bp upstream region (primer PWL2f) to 67bp downstream (primer PWL2r). Transformants with ectopic integrations or where not all copies of *PWL2* were deleted gave an amplicon of the predicted size of 645 bp (Figure 5.17) similar to that obtained from Guy11 genomic DNA. In successful mutants, the amplicon showed an increase in size consistent with incorporation of the hygromycin resistance cassette at the *PWL2* locus and deletion of all three copies of *PWL2*. Primers PWL2f and PWL2r amplified the hygromycin resistance gene (1.4 kb) from the position *PWL2* coding sequence was replaced. These primers also amplified 140 bp and 67bp upstream and downstream of *PWL2* coding sequence together with the

hygromycin resistance gene resulting in a 1.5 Kb PCR product (see Figure 5.17). The amplicons from putative mutants were excised, purified and submitted for DNA sequencing using primers PWL2f and PWL2r (Table 5.3).

Sequencing results of the region flanking *PWL2* confirmed insertion of the hygromycin resistance gene cassette and deletion of *PWL2*. These transformants were further analysed using Southern blot (See section 2.5.8). The genomic DNA was digested with *Pst*1 and fractionated by gel electrophoresis. The membrane was probed using a radio-labelled (α-^{32} P) *PWL2* coding region (See section 2.5.8.2). The probe hybridised to a 6.5kb fragment in the wild type Guy11 but not in transformant 12 (T12) as shown in Figure 5.17. The probe hybridised to a faint restriction fragment in transformants T5 and T6. Transformants T5 and T12 were sent for whole genome sequencing to investigate efficiency of CRISPR and analyse for double-stranded breaks and complete gene deletion at all three *PWL2* loci as well as potential occurrence of off target effects.

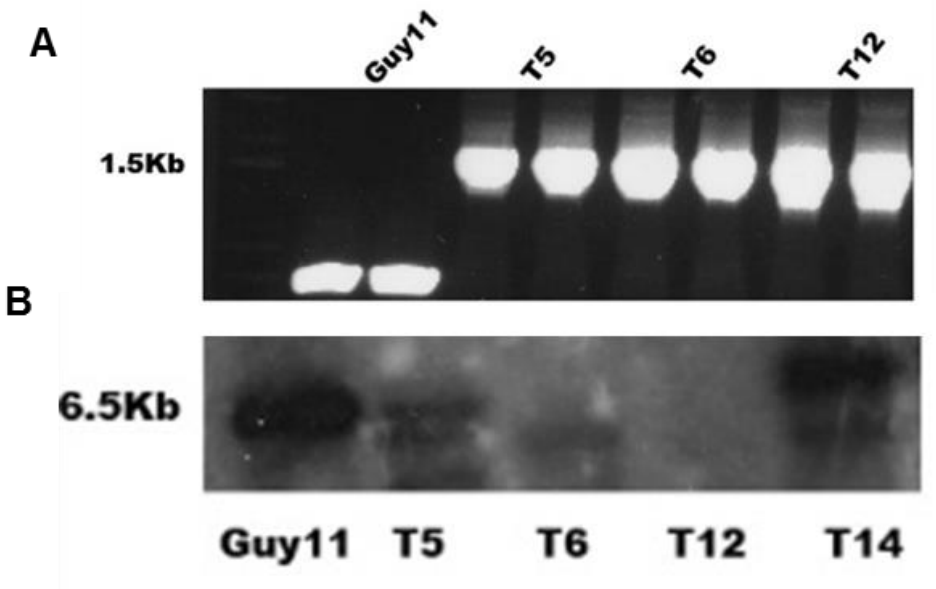

Figure 5.17 PCR screens and southern blotting of selected *Δpwl2* putative mutants.

(A) Genomic DNA of Guy11 and transformants were used to amplify the *PWL2* coding sequence using primers listed in Table 5.3. *PWL2* coding sequence was amplified from wild type Guy11 while the hygromycin cassette was amplified in transformant T5, T6 and T12 (B) Southern blot analysis showing *Pst*1 and *Bam*H1 restriction digest probed with *PWL2* coding sequence probe. The probe did not hybridise to fragmented T12 genomic DNA demonstrating deletion of all three loci of *PWL2*.

5.3.5.1 Whole genome sequencing of putative *Δpwl2* mutants confirm occurrence of Cas9-induced break and complete gene target replacement of all three copies.

The genomic DNA from putative deletion mutants (transformants T5 and T6) were sequenced to confirm and characterise the Cas9 induced insertion/deletion events. Genomic DNA and template quality was assessed using a Qubit® dsDNA BR assay before sequencing on HiSeq 2500 (Illumina) at Exeter Sequencing services. After checking the quality of reads and trimming off adaptor sequences, the reads were aligned to the reference genome (70-15) [24] using Burrow Wheeler Aligner (BWA) and SPAdes (http://bioinf.spbau.ru/en/spades_for_remove) was used to generate *de novo* genomic assemblies. To determine targeted deletion either caused by NHEJ repair or homology directed repair, BLASTn search was performed to identify presence or absence of *PWL2* and *PWL3* in the two transformants in comparison with the Guy11 genome. BLASTn was also employed to detect integration of the hygromycin resistance gene cassette at the loci.

The *PWL2* coding sequence was present in the Guy11 genome but absent in T12 and the hygromycin resistance nucleotide sequence was integrated at this specific locus (Figure 5.19). To validate this result, flanking regions upstream of the coding sequence was searched for in both the genomes and were aligned. The two sequences showed 100 % similarity upstream of the coding sequence but no similarities at the *PWL2* locus in which the hygromycin resistance gene cassette was inserted in transformant T12 (Figure 5.18). This confirmed that the *PWL2* loci in T12 have been replaced by

hygromycin resistance gene cassette. In transformants T5, the *PWL2* nucleotide sequence was split into two halves, occurring in two different short contigs of 0.399kb and 0.34 kb (Figure 5.19). In the Guy11 genome sequence, three *PWL2* loci were present in continuity and in two contigs. This suggested that a double strand break was introduced in the *PWL2* coding sequence. However, the PCR amplification result had suggested insertion of Hygromycin at the *PWL2* locus. It is possible that one or two loci of *PWL2* may be replaced by Hygromycin gene cassette while the other one or two loci contain a double stranded break (primer un-specific) observed from whole genome sequencing analysis. The two mutants were sequenced using short-read sequencing and we could not accurately determine changes at the three loci in T5. *PWL3* was present in the two mutants' genome sequences and shared 100% identity with Guy11 sequence and was not targeted by the guide RNA.

```
WT_GUY11  100 AAGATTCGGGTAGCCAGAATGCGGGGGGTGTTAATTTTAAATCCTTAAATTACATCCCTTA
Mut12     120 AAGATTCGGGTAGCCAGAATGCGGGGGGTGTTAATTTTAAATCCTTAAATTACATCCCTTA

WT_GUY11  160 CTCCGCCACTTTTCTCATTCCCTTAACGATCAACTCCCGCGTGGTTAATGATATAGTTTA
Mut12     180 CTCCGCCACTTTTCTCATTCCCTTAACGATCAACTCCCGCGTGGTTAATGATATAGTTTA

WT_GUY11  220 AAATAATTTGCTTCATCGCATTATAATAATAAAAAACTTTGAACCAGTTCGGGCACTCCG
Mut12     240 AAATAATTTGCTTCATCGCATTATAATAATAAAAAACTTTGAACCAGTTCGGGCACTCCG

WT_GUY11  280 TTACTAATTTAAAATCGAGGTAAGTGAATGAATTACGTACTAATATATATAATTATATAT
Mut12     300 TTACTAATTTAAAATCGAGGTAAGTGAATGAATTACGTACTAATATATATAATTATATAT

WT_GUY11  340 TTTTTTATTTATGCAAGCTTACTCGCGGACGGGACGAGTAAAAAACATACCTTTTTATTT
Mut12     360 TTTTTTATTTATGCAAGCTTACTCGCGGACGGGACGAGTAAAAAACATACCTTTTTATTT

WT_GUY11  400 ATGCAAGCTTACTCGTGGACAGGACGAATAAAAAAACATAATATATTTATATATGCAAGC
Mut12     420 ATGCAAGCTTACTCGTGGACAGGACGAATAAAAAAACATAATATATTTATATATGCAAGC

WT_GUY11  460 TTACTCGCGGATGGGACGAATAAAAAACATATACAATAAGGGGTTGGCTAATTTATAAGC
Mut12     480 TTACTCGCGGATGGGACGAATAAAAAACATATACAATAAGGGGTTGGCTAATTTATAAGC

WT_GUY11  520 ATACATAGGAAAGGTTCTTATTATGGTCCCGGGTGATAAAATCTTCACAGCTCCCAATTA
Mut12     540 ATACATAGGAAAGGTTCTTATTATGGTCCCGGGTGATAAAATCTTCACAGCTCCCAATTA

WT_GUY11  580 CTTTAAGGGTTTTTGTTTCGTTCTTTCATTTTTTATGTTCAGAATTACAATTAAGCTCGG
Mut12     600 CTTTAAGGGTTTTTGTTTCGTTCTTTCATTTTTTATGTTCAGAATTACAATTAAGCTCGG

WT_GUY11  640 AAAATCTCTTTTTAAAATTAAAAACTTTCAAAATGAAATGCAACAACATCATCCTCCCTT
Mut12     660 AAAATCTCTTTTTAAAATTAAAAACTTTCAAAATGAAATGCAACAACATCATCCTCCCTT

WT_GUY11  700 TTGCTTTGGTCT--------TTTTTTCGACCAC-----TGTAACCGCCGGTGGCGGGTGG
Mut12     720 TTGCTTTGGTCTCGCCAGGGTTTTCCCAGTCACGACGTTGTAA--AACGACGGCCAGTGA

WT_GUY11  747 ACTAACAAACAATTT--TACAACGACAAAGGCGAAAGAGAGGGCTCAATTTCAATTAGGA
Mut12     778 ATTGTAATACGACTCACTATAGGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCA

WT_GUY11  805 AGGGCTCGGAAGGCGATTTTAACTATGGCCCCAGTTATCCTGGAGGGCCCGATAGGA--T
Mut12     838 TGGCCGCGGGA--TTGCTCTAGATATTGAAGGAGCATTTTTGG---GCTTGGCTGGAGCT
```

Figure 5.18 Nucleotide sequence alignment of the *PWL2* locus from Guy11 and transformants T12.

Sequences in red show 100% similarity and represent the region upstream of *PWL2* start codon. The regions with black and blue coloured sequences depict region with no similarity that represent the beginning of *PWL2* coding sequence in Guy11 and hygromycin resistance gene cassette sequence in transformant T12.

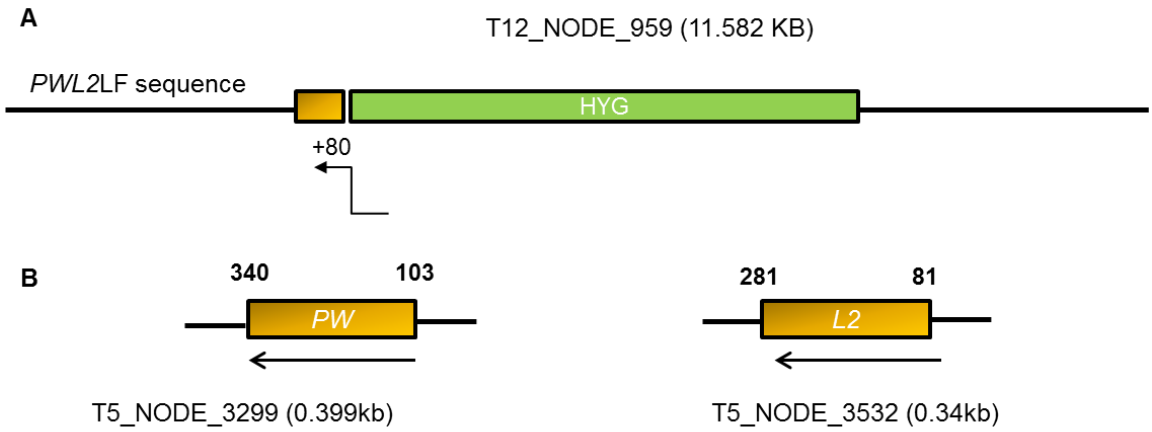

Figure 5.19 Schematic illustration of Cas9 generated *Δpwl2* mutants sequencing results.

(A) In transformant T12, BLASTn search using Hygromycin nucleotide sequence produced a hit at the *PWL2* coding sequence locus 80bp after the start codon. Arrow shows the position of KOLFR primer used to amplify left flank incorporated in donor template. (B) In transformant T5, BLASTn search using *PWL2* coding sequence nucleotide sequence produced hits on two small contigs of 0.399 and 0.34kb. The arrow shows direction of 5' to 3'.

5.3.5.2 Phenotypic and pathogenicity assay of *Δpwl2* mutants

Vegetative growth and colony morphology of the *Δpwl2* mutants were analysed. Guy11, T5 and T12 were grown on CM plates and observed 10 days post sub-culturing. T5 displayed vegetative growth like Guy11 and had normal dark concentric rings and light growing edges. T12 displayed similar phenotype but had a fluffy, less melanised growth and produced less conidia. These results suggest that *PWL2* is not involved in the vegetative growth of rice blast fungus on plates. To investigate the ability of mutants T5 and T12 to cause infection on rice, three-week old *Moukoto* rice plants were inoculated with conidia collected from 8-12 days old cultures of wild type Guy11 and the two mutants. Disease symptoms were analysed 5 -7 days' post-infection. These mutants showed no significant differences in infection compared to Guy11 and could produce sporulating lesion 5-6 days' post-infection (Figure 5.20). This shows that *Δpwl2* mutants have no obvious pathogenicity defects on rice.

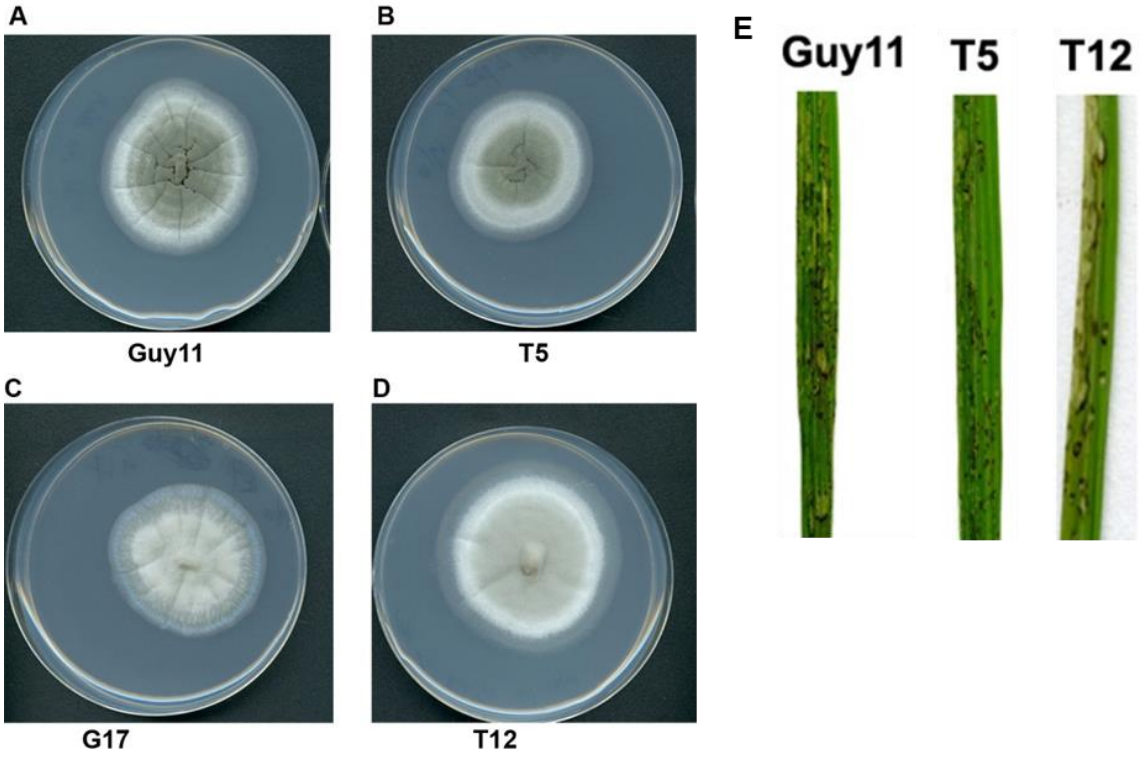

Figure 5.20 Colony morphology and compatibility assay of *M. oryzae* Δ*pwl2* Cas9 induced mutants.

(A-D) Guy11, T5, G17and T12 were inoculated on CM plates and incubated at 25 °C. Images were obtained after 10 days using an Epson Expression 1680 Pro scanner. (E) Wild type Guy11, T5 and T12 spores were used to spray 3-week old susceptible *Moukoto* seedlings and images obtained 6 days post-infection. Both T5 and T12 were pathogenic on rice cultivar *Moukoto* in three replicates. Observations were consistent in 3/3 infection replicates.

5.3.5.3 Deletion of multiple copies of *PWL2* causes gain of virulence on Weeping lovegrass

Putative *Δpwl2* gene deletion transformants T5, T6 and T12, were selected for further analysis. *PWL2* is a host-range determinant gene that controls pathogenicity towards weeping lovegrass and therefore, I reasoned that null mutants will gain virulence towards this host. To test this hypothesis, I inoculated weeping lovegrass seedlings with *Δpwl2* deletion mutants T5, T6, T12 and Guy11, the background strain, selected as a non-pathogenic negative control and weeping love grass pathogen G17 as a positive control. In the G17 pathogenic interaction, all inoculated leaf tissues started showing disease symptoms, including shrivelling of leaves after 4-5 days and by 7 days had developed full symptoms in the form of brown, shrivelled leaves (Figure 5.21). The pathogenic phenotype of G17 and Guy11 could be easily distinguished. The Guy11 inoculations showed no disease symptoms as shown in (Figure 5.21). However, transformants T5, T6 and T12 showed a dramatic gain of virulence colonising inoculated leaves and showing symptoms of infection similar to those exhibited by G17. This result confirms that the function of *PWL2* as a host range determinant for virulence towards weeping lovegrass. Moreover, the result shows that CRISPR Cas9 gene editing can be used to functionally characterise genes with multiple copies in *M. oryzae* genome through either gene disruption or gene replacement.

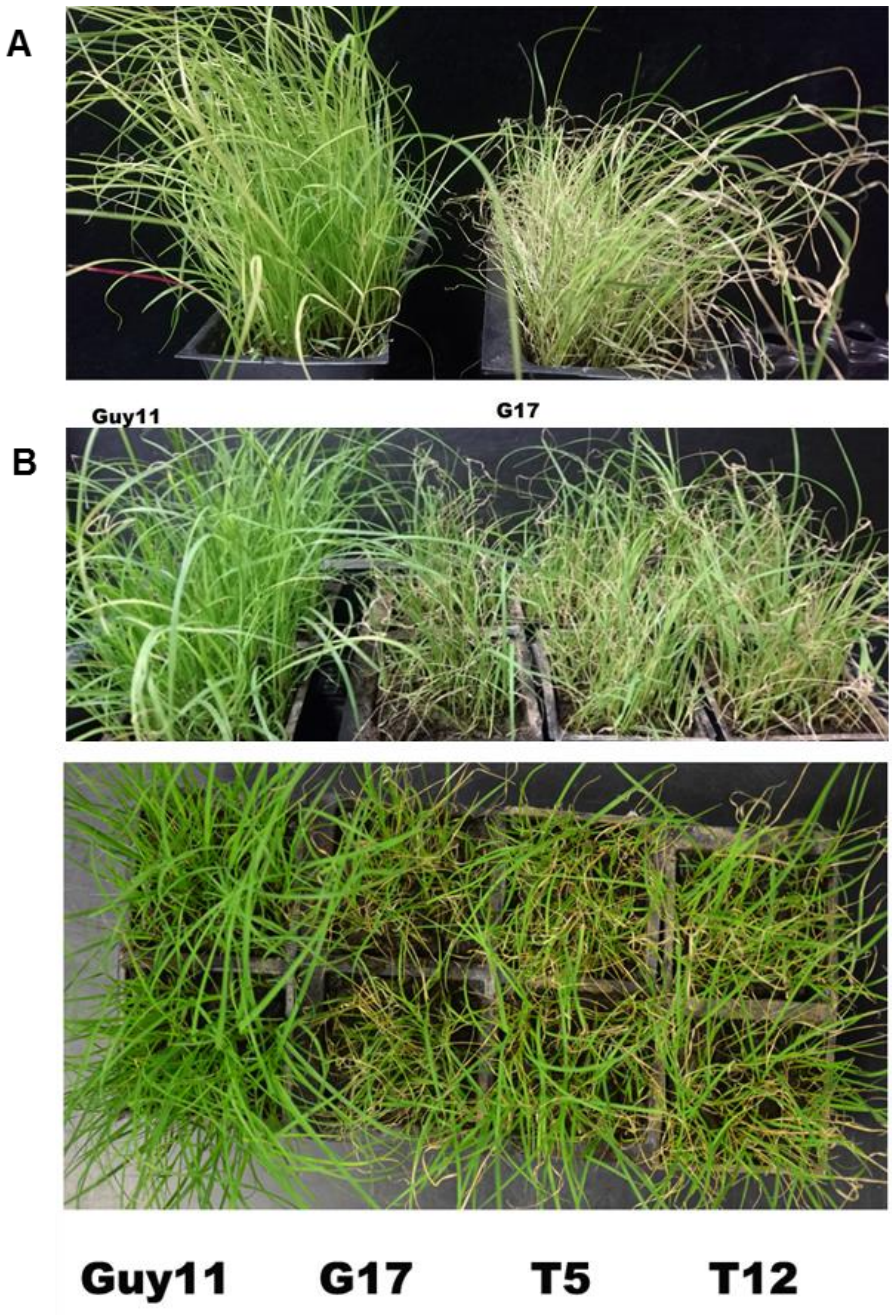

Figure 5.21 Pathogenicity assay of *Δpwl2* Cas9 induced mutants.

(A) Typical blast disease symptoms caused by G17, on weeping lovegrass, *Eragrostis curvula* compared to a non-pathogenic Guy11 the rice pathogen that possess three copies of *PWL2*. (B) G17 and *Δpwl2* mutants T5 and T12 were pathogenic to weeping lovegrass. Guy11 was non-pathogenic. Images were obtained 7 days post-infection**.** Two pots of weeping lovegrass seedlings were sprayed with different *M. oryzae* strains in four replicates. Observations were consistent in 4/4 infection replicates.

5.3.6 Investigating for Pwl2 rice-interacting proteins during microbe-host interaction

5.3.6.1 Purification of *in-planta* expressed fluorescently labelled effectors reveals putative Pwl2 interactors

To gain insight into the molecular function of the *PWL2* gene product during host cell colonisation, co-immunoprecipitation was used to identify plant protein that interacts with Pwl2. To increase the chances of detecting effector proteins in a mixture of fungal-plant proteins, leaf drop infections were carried out to optimise the amount of fungal biomass in infected plant tissue. Rice leaves were inoculated with two different *M. oryzae* strains, one expressing Pwl2:mRFP, and other Bas4:mRFP. Proteins extracts were made from inoculated leaves collected at 36 and 48 hpi, to coincide with a time when effector proteins would typically be expressed. A sub-set of inoculated rice plants were left for 3 - 4 more days to develop full infection and disease.

Leaf tissue infected with *M. oryzae* strains expressing two effectors, Mep1-GFP and Mep3-GFP, and Mep3 promoter signal peptide fused to GFP (Mep3sp-GFP) (Yan *et al* unpublished) were used as negative control to eliminate promiscuous interactors that include ‘sticky’ proteins or highly abundant proteins such as plant proteins associated with the translation machinery, chlorophyll-binding proteins or other abundant enzymes. Infected leaf tissue was collected and frozen in liquid nitrogen before being used for total protein extraction. An aliquot of the sample was fractionated on a sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel and a Western blot used to detect accumulation of Pwl2-mRFP and Bas4-mRFP.

Separated proteins were transferred onto a Nitrocellulose membrane and probed with an anti-RFP antibody. A secondary antibody conjugate (alkaline phosphatase) was then added.

Both proteins were detected as single protein bands of correct predicted sizes (39kDa and 43kDa for Bas4-mRFP and Pwl2-mRFP respectively) (Figure 5.22). Bas4 has a molecular weight (MW) of 12 kDa, Pwl2 has 16.16kDa MW while RFP has a MW of 27kDa [117, 266]. Pwl2-mRFP was abundantly expressed and had a brighter and more intensive band compared to Bas4:mRFP. These results are consistent with the localisation patterns of fluorescently-tagged Pwl2 and Bas4 at 48 h. Secreted Pwl2 is delivered into the host cytoplasm and already translocated into neighbouring cells whereas Bas4 remains in the extra-invasive hyphal membrane (EIHM) compartment between the fungal cell wall and the host cell plasma membrane. The Pwl2-mRFP and Bas4-mRFP protein with potential interacting proteins were isolated from the total extracted protein using RFP-trap beads and washed several times to release unbound immobilised proteins, before submitting for liquid chromatography-tandem mass-spectrometry (LC-MS/MS) and an aliquot used for Western blot analysis.

The proteins were digested using trypsin, and peptides then identified through mass spectrometry. The peptides were mapped to a database of fungal and rice proteomes and a score given for each match of a spectra to a predicted peptide. Proteins with the highest score were considered as the best match. This list of matched proteins was compared to control experiments (Mep3sp-GFP, Mep1-GFP and Mep3-GFP) and any protein occurring in all the

controls was considered an artifact and disregarded. Proteins with a score less than 15 were also disregarded. A total of 13 rice proteins were selected as candidate effector targets for Pwl2 (Table 5.2). The List containing Bas4 matching proteins was identical to the list of rice proteins that matched to either the controls or Pwl2, and were considered as not likely to be genuine interactors. Pwl2 was associated with several high scoring chloroplast-localised proteins which were selected for confirmatory experiments. LC-MS/MS data processing and protein identification was carried out as described in Petre *et al* [199].

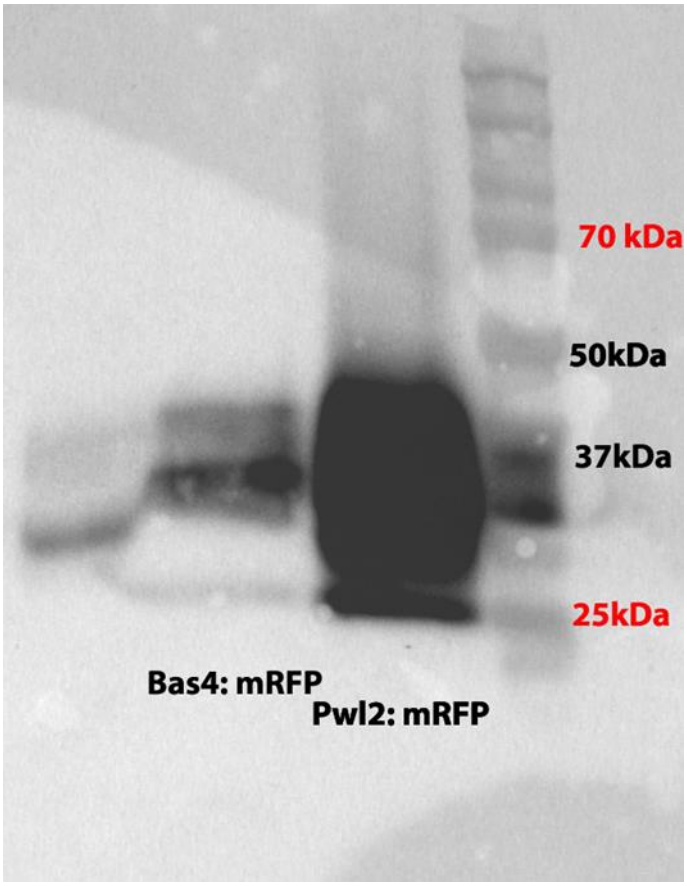

Figure 5.22 Western blot detection of fluorescently-labelled Bas4 and Pwl2 during infection on rice cultivar CO39.

Leaves of a susceptible rice line *Moukoto* were inoculated with two different Guy11 strains, one expressing Pwl2-mRFP and another Bas4-mRFP. Separated proteins were transferred onto a Nitrocellulose membrane and probed with an anti-RFP antibody. Pwl2-mRFP accumulation was detected as bright band of 43 - 50 kDa. Bas4: RFP accumulation was detected as a band of 39 – 45 kDa. Bas-4 has a molecular weight (MW) of 12 kDa, Pwl2 is 16.16kDa while RFP has a MW of 27kDa

Table 5.2 List of putative Pwl2 interacting rice target proteins

Description	Score	No. of Proteins	MW (kDa)
Chlorophyll a-b binding protein, chloroplastic	127.5	1	27.5
Ribulose bisphosphate carboxylase small chain	123.84	2	19.6
Chlorophyll a-b binding protein, chloroplastic	104.3	1	33.4
Putative 40S ribosomal protein; contains	50.20	1	26
Plasma membrane ATPase	47.91	7	104.8
Plasma membrane ATPase	46.10	1	39.3
ATP/ADP translocator protein	43.94	1	41.5
Cytochrome b5 protein	27.46	1	15.3
Ovp1	22.93	4	80.6
Putative photosystem I chain V (Fragment)	21.68	1	8.1
ATP synthase epsilon chain, chloroplastic	19.85	1	15.2
Pathogenesis-related thaumatin-like protein	19.39	1	19.2
Vacuolar ATPase B subunit	18.82	1	54.0

[8]

[8] List of selected matched proteins after comparing to the controls (Mep3sp-GFP, Mep1: GFP and Mep3: GFP). Most of the proteins identified were chloroplast proteins.

To complement the mass spectrometry results, we used transient expression to observe for sub-cellular localisation of the GFP-tagged Pwl2 protein in *N. benthamiana*. To determine where Pwl2 accumulates in plant cells, the *PWL2* gene without its signal sequence was amplified from cDNA obtained from rice tissue infected with *M. oryzae*. The sequence of the mature protein-encoding gene sequence (without signal peptide) was cloned into a binary vector to express GFP downstream of an *Arabidopsis* ubiquitin10 promoter. This plasmid was transformed in *Agrobacterium tumefaciens* strain GV3101 and expressed in *Nicotiana benthamiana* using agroinfiltration method [113, 252, 253]. All analysis was done on live leaf tissue 48 h post-infection.

Square leaf discs from the *N. benthamiana* infiltrated leaf area were mounted in a perfluorocarbon immersion and observed by laser confocal microscopy [199, 267]. GFP-tagged proteins were excited using a 488 nm laser and auto-fluorescence from chlorophyll detected at 650-740 nm, to determine whether Pwl2 interacts with chloroplast proteins, resulting in localisation of Pwl2-GFP in this organelle. Pwl2-GFP was however only observed in the cytosol as shown in Figure 5.23. This suggests that Pwl2 does not translocate to the chloroplast. However, as a non-host for *M. oryzae*, *N. benthamiana* might have different cellular mechanisms and not recognise this effector.

We next carried out yeast-two-hybrid (Y2H) interaction assays to further investigate the molecular mechanism underlying the contribution of Pwl2 to biotrophic proliferation of *M. oryzae* and its role in immune suppression. The *PWL2* gene without the predicted signal sequence was amplified from cDNA obtained from rice tissue infected with *M. oryzae*. The PCR product was then

cloned into the bait vector pGBKT7 DNA-BD, while potential host targets of Pwl2 including the chlorophyll a-b binding protein and ATP synthase predicted by co-immunoprecipitation experiment (Table 5.2) isolated from cDNA library of rice tissues infected with *M. oryzae* were cloned into a prey vector pGADT7.

The two vectors were transformed into Y2H gold yeast cells and interaction stringency analysed on three selection medium (-Leu/-Trp), (-Leu/-Trp/-His) and (-Leu/-Trp/-His/-Ade). Low stringency interaction on -Leu/-Trp medium indicated that growing colonies had both plasmids integrated and could synthesise both amino acids. These colonies were sub-cultured onto medium and high stringency medium. Growth on medium stringency medium (–Leu/-Trp/-His) indicated that colonies possessed interacting proteins encoded by both plasmids and could synthesis Leucine, Trypsin and Histidine. However, none of these proteins showed high stringency interactions with Pwl2 (data not shown).

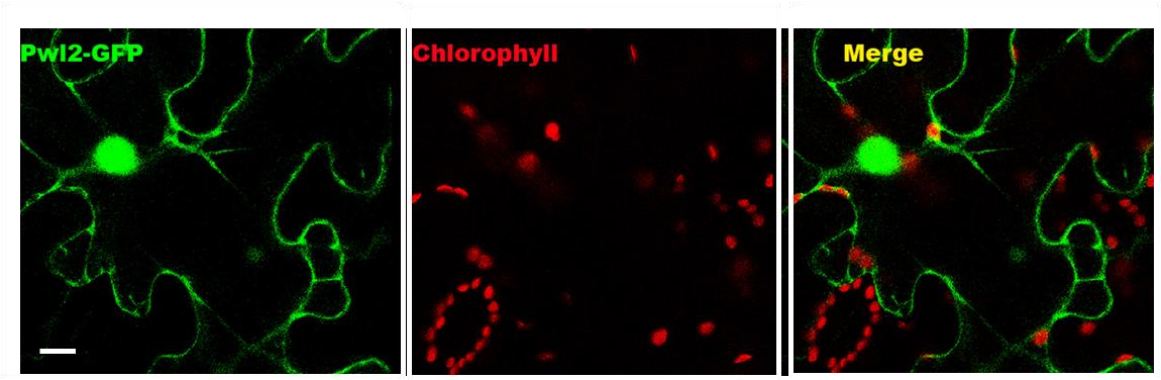

Figure 5.23 Subcellular localisation of Pwl2-GFP in *Nicotiana benthamiana* cells.

Pwl2-GFP was observed in cell nucleus and cytoplasm of *N. benthamiana* cells. Autofluorescence from chloroplast is indicated in the red channel. Merger of green and red channel did not show specific overlap in fluorescence. Scale bar represent 10 µm.

5.4 Discussion

In this chapter I have provided evidence that suggest that, *PWL2* is under constant selection pressure and is an example of rapidly evolving avirulence gene in *M. oryzae*. Moreover, the occurrence of multiple copies of *PWL2* gene in *M. oryzae* demonstrates that the gene has experienced gene duplication and gene family expansion that is consistent with the occurrence of observed repeated sequences and transposable elements adjacent to the *PWL2* locus. Elucidating the function of the *PWL2* gene will help to explain the reason for these events. It is intriguing that as *M. oryzae* has evolved, the *PWL2* gene has been maintained in the population and in fact been amplified in most isolates analysed from a worldwide collection of *M. oryzae.* If the importance of *PWL2* outweighs the consequences of losing the gene, then this may explain the reason for maintenance and amplification of *PWL2* in the population. Moreover, if Pwl2 serves an important function in fungal fitness or during biotrophic growth of the fungus, it will explain why most isolates carry alleles of *PWL2* that are no longer recognised by the cognate R gene in *Eragrostis curvula* [118]. The specific modified region in Pwl2 alleles can perhaps help to identify amino acids patterns involved in AVR/R-gene recognition and binding of which would likely to be essential for conferring recognition and a successful immune response by *Eragrostis curvula*.

In most *M. oryzae* isolates collected from Japan, *PWL2* was missing or there was occurrence of *pwl2* alleles. Rice pathogens Ina168 and P131 for example did not possess *PWL2* gene and were consequently pathogenic on weeping lovegrass. The two most aggressive pathogens, of weeping lovegrass

G17and 4091.5.8 were isolated from Japan consistent with this observation. If we consider selection pressure to be the driving force for rapid evolution and expansion of *PWL2*, then we can conclude that deployment of rice cultivars carrying R-genes cognate to Pwl2 in Asia, and particularly in Japan may have occurred in the past. In addition, it is possible that some *M. oryzae* isolates might infect more than one host species in the field which might include weeping lovegrass as it is a common forage grass. In this study, I have presented data that suggests the Pwl2 virulence allele is still expressed normally and might be still serving an unknown function during biotrophic growth.

To investigate further, a null mutant was generated using CRISPR/Cas9 genome editing approach. Two *Δpwl2* mutants generated by this gene editing technology in a Guy11 background resulted in deletion of all three copies and were able to infect weeping lovegrass, confirming the function of *PWL2* as a host range determinant. However *Δpwl2* mutants failed to show any phenotypic or pathogenicity defects when inoculated on a susceptible rice cultivar *Moukoto*. It is not clear whether *M. oryzae* specificity to difference grass hosts is controlled by R-genes, or by other unknown mechanisms. Understanding these molecular mechanisms involved in host specificity can be used as a reservoir for R-gene specificity that could be used in cereal crops. This will also improve strategies to breed for or engineer crops, by transferring resistance across the *Poaceae* family.

All *M. oryzae* isolates from Sub-Saharan Africa analysed in this study possessed *PWL2,* with the Kenyan KE002 confirmed to possess five copies of

PWL2, while BF48 had at least four copies present. Using the example of Guy11 and KE002 in this study it is not possible to determine accurately the extent of *PWL2* gene family expansion in most field isolates unless a long-read sequencing approach is used. Short-read sequencing consistently failed to identify additional *PWL2* in Guy11, for example. As demonstrated by Khang *et al* [117], the *PWL* family genes are often associated with repetitive sequences. In this study, we found that regions surrounding the three *PWL2* genes in Guy11, 10kb upstream and 14kb downstream were similar in all the three copies. Regions nearly 1 kb upstream and downstream of *PWL2* were highly repeated and produced more than 130 hits when blasted in the Guy11 genome. This made gene replacement conventional homologous recombination very challenging indeed.

I have also used high resolution microscopy to gain an insight on how effectors are secreted into the BIC and confirmed that, the BIC is a plant-derived structure used by the rice blast fungus to secrete and deliver effectors across host plasma membrane. The data generated in this chapter confirm that the cytoplasmic effector Pwl2 is moving in one direction through the BIC from the invasive hypha into the targeted plant cell. The fact that vesicles containing BIC-localised effectors are associated with individual sites of infection shows that there is no sequestration of effectors from the host cytoplasm to the BIC. I have also shown that, after the fungus contacts host tissue, expression of *PWL2* starts prior to rice cell penetration. However, translation into mature protein appears to occur after penetration in the host cell after infection has occurred. This is consistent with the description of fungal effectors as encoded by pathogens and functioning in host cells.

Using *M. oryzae* Pwl2 and Bas1, I have shown that fungal effectors translocate across the plasma membrane packaged in vesicles which are visible within the BIC, and that these effectors are sorted into distinct vesicles before crossing the plasma membrane. There is increasing evidence that *M. oryzae* effectors will translocate into the host cell as extracellular vesicles (Oliveira-Garcia, Valent *et al* unpublished) but this was not observed in this study. It will be of interest to study how *M. oryzae* effectors with conserved structural domains such as MAX-domain containing effectors [210], with the same host targets, translocate into the BIC and if this involves separate secretory vesicles.

Table 5.3 List of oligonucleotide primer used in chapter 5

Primer	Oligo sequence (5' to 3')
MutPWL2F	CCTGATCATCAGGAAAATAAAAGCGATCGT
MutPWL2F	GACGATCGCTTTTATTTTCCTGATGATCAG
PWL2ORF	ATGAAATGCAACAACATCATCCTCCC
PWL2ORR	ACATAATATTGCAGCCCTCTTCTCGC
KOLFF	ACAACGCGGTGTAAAGATTCGGGT
KOLFR	GTCGTGACTGGGAAAACCCTGGCGAGACCAAAGCAAAAG GGAGGATGA
KORFF	TCCTGTGTGAAATTGTTATCCGCTGCGAGAAGAGGG CTGCAATATTA
KORFR	CGGCGTGGCTGGTAGGTCGAGTGG
PWL2f	GGTTCTTATTATGGTCCCGGGTGA
PWL2r	GGGCGTGATCCCTCACACCTAAGT
Cons-F PWL2	CGCGGTGGCGGCCGCTCTAGAATGTGTCTCGATGATTGTTG
Cons-FPWL2 GFP	GGGCTGCAATATTATGATGGTGAGCAAGGGCGAGGAGCTG
Cons-R PWL2	CATAATATTGCAGCCCTCTT
PWL2848F	CGCGGTGGCGGCCGCTCTAGAGACCGGGCACGAACCCGGC
AGGC	
YGF	GTGATTTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAA
HYR	CAGCAATCGCGCATATGAAATCAC
BaitPWL2F	CATGGAGGCCGAATTCCCACCATGGGTGGCGGGTGGACTAAC
BaitPWL2R	GCAGGTCGACGGATCCTTACATAATATTGCAGCCCT
GTWPWL2F TGGACTAAC	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCACCATGGGTGGCGGG
GTWPWL2R	GGGGACCACTTTGTACAAGAAAGCTGGGTGCATAATATTGCAGCC
PSY1F	GGAGGCCAGTGAATTCccaccATTCCCGGGGCGGGAGAA CGTGGC
PSY1R	CGATGCCCACCCGGGTCAGAAGAAGTTTGGGTTGTATCCG
ADPF	GGAGGCCAGTGAATTCccaccATGGCTGAGCAGGCTAATCAACCG

ADPR	CGATGCCCACCCGGGTTAGGCACCGCCCGAGCCGTACTT
Chloroa-bF	GGAGGCCAGTGAATTCCCACCATGGCCGCCGCCACCATGGCCCTC
Chloroa-bR-	CGATGCCCACCCGGGTCACTTGCCGGGGACGAAGTTGGT
RBSCF	GGAGGCCAGTGAATTCCCACCATGGGCCCCACCGTGATGGCCTCC
RBSCR	CGATGCCCACCCGGGTTAGTTGCCACCAAACTCCGCCCG

Chapter 6 General conclusion and discussion

Rice consumption in Africa has surpassed domestic rice production. Recently, new rice cultivars - New Rice for Africa (NERICA), a cross between *Oryza glaberrima* Steud (African rice) and *Oryza sativa* L. (Asian rice), have been introduced to farmers by Africa Rice Centre, Benin [10]. Although these varieties have advantages, such as high yields, early maturity and increased resistance to pathogens, they are still susceptible to rice blast [10]. Several studies have shown that deployment of rice cultivars with rice blast R-genes can only be effective in a specific region for short period, because of resistance gene breakdown [228, 229]. The best strategy to deal with breakdown of host resistance is therefore to pyramid several R- genes in local adapted rice cultivars, for example NERICA and Basmati varieties in Sub-Saharan Africa [231].

This thesis set out to investigate the effector repertoire of a rice blast population, as a means of guiding a breeding programme for rice blast resistance, in Sub-Saharan Africa. The study was aimed at achieving two major objectives. First, we assessed the virulence spectrum of 122 rice blast isolates from Sub-Saharan Africa. Out of this, a set of 23 *M. oryzae* isolates were selected on the basis of their pathotype and genotype variability and their genomes were sequenced. We then used association genetics to predict *M. oryzae* effector protein-encoding genes that determine virulence towards rice cultivars carrying known resistance genes. This was carried out to identify appropriate R-genes that can be deployed in-order-to achieve durable resistance. We also set out to use this information to facilitate the mining of

novel putative effector encoding gene. Secondly, I assessed a host-specificity determinant gene, *PWL2,* as a means of investigating a totally novel form of disease resistance based on identifying fungal host specificity determinants and their cognate R-gene in either a wild grass host, or among diverse collections of rice. I then used a range of different techniques to characterise the role of *PWL2*, which we discovered was a gene that has undergone gene duplication and expansion in several *M. oryzae* isolates, including Guy11.

To achieve these objectives, I had to overcome two major challenges. First of all, there was a need to develop and apply new techniques to yield well-assembled contiguous genome sequences from *M. oryzae*, to accurately predict and annotate putative secreted effector protein encoding genes. Secondly, there was also a need for a new method to generate null mutants for genes occurring at multiple loci in the same genome.

Long-read sequencing in combination with RNA-sequencing essential for accurate gene prediction

Comparative genomics offers a rapid way of identifying new avirulence genes. The number of cloned *AVR* genes (11 to date) is still low compared to the number of reported cloned R-genes (25 to date in 2018), especially when considering the size of the respective genomes [42]. Despite its small genome size, *M. oryzae* has a high percentage of repetitive sequences which comprise 10% of the whole genome [44]. This is common in phytopathogenic fungi, in *Verticillium dahliae* for example, a new completed genome assembly proved that the extent of repetitive sequences is more than previous estimates – 12% of the genome compared to earlier prediction of 4% [37].

Multiple genome studies have indicated the importance of the repetitive regions, non-coding regions and genome structural re-arrangements in the lifestyle of living organisms [45]. Transposable elements play a crucial role in effector evolution, causing gene deletion and expansion, or affecting gene expression [45]. Short-read sequencing falls short of determining genome arrangements and continuity [37]. Any repetitive sequences falling between these ends will, for instance, still be lost during genome assembly [37]. However, it is now possible to use recent technologies that generate long reads to characterise repeat-rich regions of a genome.

Our large-scale pathotype analysis of isolates of *M. oryzae* on the international rice differential lines, carrying known R genes, identified two R-genes that could be deployed in combination with other genes to acquire durable resistance. *Pi9* and *Pita2* were found to confer resistance to more than 90% of *M. oryzae* isolates analysed from across sub-Saharan Africa [231]. Consistent with this, avirulence gene analysis of these isolates revealed that the *AVR-Pi9* gene occurs in all Sub-Saharan Africa isolates. It is therefore essential to understand the function served by highly conserved avirulence genes, such as *AVR-Pi9* during fungal infection. It is possible that the importance of *AVR-Pi9*, for example, outweighs the consequences of losing the gene, which may explain the reason why most analysed *M. oryzae* isolates carry *AVR-Pi9*. The putative importance of *AVR-Pi9* in the *M. oryzae* genome may be inferred by how frequently mutations or deletions occur in this locus. Above all, monitoring the breakdown of a highly efficient R-gene like *Pi9*, could offer more insight into the pathogen-host interaction. For example, information regarding how long it takes for a deployed broad-range R-gene to be broken down can be obtained

through blast population surveillance of *Pi9* in blast hotspots, where severe disease pressure exists. There was no clear correlation between occurrence of the *AVR-Pib* and *AVR-Pik* alleles to virulence shown on *Pib* and *Pik* alleles, respectively. This observation may be as a result of selection pressure imposed on *AVR-Pib* and the *AVR-Pik* alleles, by the corresponding *Pib* and *Pik* alleles, respectively. For example, in West Africa deployment of rice cultivars that contain *Pik* alleles that recognises *AVR-Pik*-D might have resulted in selection for *AVR-Pik*-E to evade recognition.

Genome-wide association studies to determine the relationship between the occurrence of novel predicted effector genes and virulence of rice monogenic lines was not yet successful in definitively identifying new *AVR* genes. I hypothesise that this is because of inaccurately predicted genes, or omissions during gene calling in all genome sequences analysed. For example, gene prediction carried out on short-read sequencing assembled genomes produced fewer genes than expected. To our surprise, we could not accurately predict known avirulence genes in the Kenyan strain KE002, as a consequence. We related this observation to the highly fragmented *M. oryzae* genome sequences that were obtained from inaccurate assemblies.

Long-read sequencing technology from Pacbio, in combination with RNA analysis, was therefore used to improve gene predictions in KE002 and Guy11 *M. oryzae* genome sequences. Pacbio sequencing yielded larger genome assemblies than those obtained using Illumina technology. It is possible that the observed increase in genome size and contiguous sequenced fragments is a result of successfully assembled repeat-rich regions. Using this combination, we

could predict many more genes in both Guy11 and KE002 long-read assembled genome sequence compared to short-reads assembled genome sequences. Moreover, it was also possible to predict more small secreted protein-encoding genes using this approach, including known avirulence genes occurring in KE002 genome. In both genomes for example, several secreted proteins including two effectors *MEP13* and *PWL2* were found to have undergone genome duplication. This observation is consistent with several studies that have reported genome translocation, duplication and deletion of effector-encoding genes [217, 218]. I conclude that gene expansion and gene duplication events are common in *M. oryzae*, but cannot be well studied using data generated from short-read sequencing. Cases of gene expansion and gene duplication, moreover, make it difficult to generate null mutants and functionally characterise putative effectors.

I used an effector prediction program *EffectorP* to further improve putative effector annotation. With this approach, it was possible to differentiate putative effector genes from non-effector genes. I propose the program can be incorporated in effector identification pipeline for newly sequenced genomes and can be an essential tool in effector biology. Three predicted effector proteins encoding genes from this study, *MEP13*, *MEP14* and *MEP15* exhibited BIC localisation when expressed in rice leaf sheath. This is typical of all cloned avirulence proteins apart from Ace1 which is a well-characterised polyketide synthase [96]. I reasoned that if *MEP13* is an avirulence gene, *Δmep13* mutants would gain virulence on one specific rice monogenic lines resistant to WT KE002. However, *Δmep13* mutant did not show a gain of virulence when analysed on selected rice monogenic lines resistant to KE002. Additionally,

MEP14 genetic complement did not show loss virulence defects when analysed on selected rice monogenic lines. In this study, 621 genes were annotated as genes encoding for putative effector proteins in KE002, out of which 139 were unique to KE002. An extensive study of isolate-specific putative effector encoding genes is in progress. Analysis of genetic complements of Guy11 expressing KE002 isolate-specific putative effector encoding genes on monogenic rice lines will now be carried to mine for more avirulence genes.

***PWL2*, a host-range determinant gene**

Results presented in Chapter 5 confirm that *PWL2* is a host range determinant with a predicted cognate R gene in weeping lovegrass, *Eragrostis curvula. M. oryzae* is known to infect a wide range of grasses in a manner thought to be controlled by host specificity determinants. Most recently, Inoue et al [20] for example reported two avirulence genes *PWT3* and *PWT4* that control infection of *Magnaporthe* isolates towards *Triticum aestivum* (Wheat). Deployment of cultivars lacking Rtw3 resulted in wheat susceptibility towards *Lolium*-infecting isolates of *M. oryzae*, an event that coincided with the outbreak of wheat blast in the province of Parana', Brazil [20]. It is not clear whether specificity of *M. oryzae* to different grass species is controlled exclusively by single avirulence to R-gene interaction. However, from these two examples, we hypothesised that single major R-genes can be used to exclude different *M. oryzae* pathotypes. For example, introgressing resistance genes from wild species or specific grass species can be used as a novel way of achieving durable resistance in cultivated crops in the *Poaceae* family. All *M. oryzae* isolates from Sub-Saharan Africa analysed in this study, for example, possessed *PWL2.* From this

observation, we hypothesised that durable resistance could be achieved by introgressing a resistance gene cognate to *PWL2*, or by pyramiding this with other promising R-genes.

First, I studied the mechanism by which this effector is regulated, expressed and translocated into rice cells. I have also used a range of different techniques to understand the putative function of this gene during rice blast infection. Results presented in Chapter 5 suggest that, the expression of *PWL2* starts immediately upon contact with the leaf surface, but translation into mature Pwl2 protein appears to occur in host cells. Additionally, I have presented preliminary data suggesting that effectors translocate from the fungal plasma membrane as extracellular vesicles. These results are consistent with the idea that the invasive hyphae are responsible for active secretion of effector proteins. Another possibility is, however, that after transcription, *PWL2* mRNA is trafficked at the site near the BIC before translation [268]. In eukaryotes, transport of mRNA is an important process involved in expression and precise sub-cellular localisation of proteins [269, 270]. A recent report has shown that a pathogenic fungus *Botrytis cinerea* can deliver small RNAs into host plant cell and hijack host RNA interference defense response in *Arabidopsis thaliana* [271]. To my knowledge, this phenomenon has not been reported in *M. oryzae* and more experiments are needed to test this hypothesis.

In this study, I also identified that five copies and three copies of *PWL2* exist in *M. oryzae* genome sequences of a Kenyan isolate KE002 and Guy11, respectively. I observed that it is not possible to accurately determine the extent of *PWL2* gene family expansion in most field isolates unless a long-read

sequencing approach is used. The regions surrounding the three *PWL2* gene copies in Guy11, 10kb upstream and 14kb downstream were similar in all the three copies. It seems likely that the repeated sequences surrounding the *PWL2* loci are responsible for translocation and generation of multiple copies of the gene. For this effect, there was a need for a new technique to delete and functionally characterise *PWL2*.

Two *Δpwl2* mutants were therefore generated by CRISPR-Cas9 gene editing technology in a Guy11 background and this resulted in deletion or induced mutations in all three copies of *PWL2*. Successful deletion of all three copies of *PWL2* was first confirmed using Southern blot and then by whole genome sequencing. Successful mutants were further confirmed due to gain of virulence on weeping lovegrass. So far, we did not observe any off targets effects after sequencing genomes of the two *Δpwl2* mutants. For example, the nucleotide sequence *PWL3,* a member of *PWL* family with 57% homology to *PWL2* occurring in Guy11 genome sequence was not deleted or targeted for editing. This suggested a low chance of off-target effects by the CRISPR-Cas9 gene editing system in *M. oryzae.* Considered together, I have shown that it is possible to functionally characterise genes with multiple copies in *M. oryzae* genome either through gene disruption or gene replacement, using CRISPR/Cas9 gene editing approach.

Knowledge of host specificity determinants can be used to carry out interspecies R-gene transfer as a strategy to acquire durable resistance. Additionally, this knowledge will assist rice breeders and plant pathologists to plan for potential cases of host jumps and emergence of new pathotypes of *M.*

oryzae. To my knowledge, the genome of weeping lovegrass has not been published, but with the application of recently developed technologies it may be possible to identify the NLR complement to, and potentially define the putative cognate R-gene to *PWL2*. Recent studies have shown that, resistance to phytopathogenic fungi can be transferred from wild species into susceptible commercial varieties [272, 273]. With this approach, a reference genome can be mined for NLRs, from which biotinylated RNA sequence capture libraries are synthesised [272, 273]. These libraries could then be used to clone NLRs from distant related species or unsequenced taxa [273]. Most recently, resistance gene sequence capture (RenSeq) in combination with long-read sequencing was used, for example, to clone Rpi-amr3i from wild, diploid non-tuber-bearing *Solanum americanum* which harbours multiple Rpi genes [272]. This approach could be adopted to identify the resistance gene specific to Pwl2. In a different approach, the host Pwl2 target could be identified using co-immunoprecipitation (coIP), whereby Pwl2 and cognate resistance protein complexes are identified using antibodies specific to a protein fused to the effector and then analysed using liquid chromatography-tandem and mass spectrometry [254, 255]. A yeast two-hybrid screen could then be further used to confirm this interaction before using this for testing in transgenic plants. Alternatively, an association genetics study in weeping lovegrass or rice population and map-based cloning could be carried out to identify NLRs that recognises Pwl2.

Conclusion and long term goals of durable blast resistance

Regional differences in rice blast pathotypes pose a major challenge in deployment of R-genes in form of quick break down of resistance [228]. Break down of resistance is dependent on the interaction between *M. oryzae* and its host R–genes and renders deployment of one R-gene not a durable solution [228, 229]. The best strategy to deal with breakdown of host resistance is to pyramid several R- genes in local adapted rice cultivars [231]. As part of this study, *Pi9* and *Pita2* were found to confer resistance to a majority of isolates analysed and this correlated with presence of *AVR-Pi9* in the analysed rice blast population. Ultimately, a combination of different stacks of resistance genes can be used to achieve a durable blast control in a specific region [231]. From this study it is clear that a combination of *Pi9* and *Pita2* introduced into local adapted cultivar like NERICA and Basmati rice cultivars will achieve a durable resistance in West Africa. In East Africa for example, combining *Pi9* and *Pi12* into NERICA or a local commercial variety like Basmati 370 will offer durable resistance. As part of this study, we are currently introgressing the identified gene combinations into local adapted varities using marker-assisted breeding to verify our observation. Before deployment, we will also carry out a performance analysis of the gene pyramids in blast disease hotspots both in East and West Africa. Ultimately, the best strategy will involve stacking major resistance genes *Pi9*, *Pi2*, *Piz5*, *and Pita2* followed by hotspot testing to ascertain their durability. These genes can then be combined with recessive resistance allele like *pi21* and other QTLs such as *Pi-A-35* into major R-gene stacks, offering the possibility of long-term durable control of rice blast disease.

7 Bibliography

1. Mosquera G., Giraldo M.C., Khang C.H., Coughlan S., B. V. Interaction Transcriptome Analysis Identifies Magnaporthe oryzae BAS1-4 as Biotrophy-Associated Secreted Proteins in Rice Blast Disease. American Society of Plant Biologists. 2009;21:1273–90.

2. Giraldo MC, Dagdas YF, Gupta YK, Mentlak TA, Yi M, Martinez-Rocha AL, et al. Two distinct secretion systems facilitate tissue invasion by the rice blast fungus Magnaporthe oryzae. Nat Commun. 2013;4:1996. doi: 10.1038/ncomms2996. PubMed PMID: 23774898; PubMed Central PMCID: PMCPMC3709508.

3. Khang C. H., Berruyer R., Giraldo M.C., Kankanal P., Park S-Y., Kirk C., et al. Translocation of Magnaporthe oryzae Effectors into Rice Cells and Their Subsequent Cell-to-Cell Movement. American Society of Plant Biologists. 2010;22:1388–403.

4. Dai L, Wu J, Li X, Wang X, Liu X, Jantasuriyarat C, et al. Genomic structure and evolution of the Pi2/9 locus in wild rice species. Theor Appl Genet. 2010;121(2):295-309. Epub 2010/03/17. doi: 10.1007/s00122-010-1310-0. PubMed PMID: 20229250.

5. Godfray HC, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, et al. Food security: the challenge of feeding 9 billion people. Science. 2010;327(5967):812-8. doi: 10.1126/science.1185383. PubMed PMID: 20110467.

6. Organization FaA. THE STATE OF FOOD AND AGRICULTURE 1990. FAO Agriculture Series. 1990;23.

7. Goff SA. Rice as a model for cereal genomics. Curr Opin Plant Biol. 1999;2(2):86-9. doi: Doi 10.1016/S1369-5266(99)80018-1. PubMed PMID: WOS:000079695500003.

8. Balasubramanian V, Sie M, Hijmans RJ, Otsuka K. Increasing Rice Production in Sub-Saharan Africa: Challenges and Opportunities. Advances in Agronomy. 2007;94:55-130.

9. M. S. Rice is life in 2004 and beyond. International Rice Commission Newsletter. 2005;54:1-10.

10. <Phillips Abstract and Biog.pdf>.

11. Couch BC, Kohn LM. A multilocus gene genealogy concordant with host preference indicates segregation of a new species, Magnaporthe oryzae, from M. grisea. Mycologia. 2002;94(4):683-93. PubMed PMID: 21156541.

12. Wilson RA, Talbot NJ. Under pressure: investigating the biology of plant infection by Magnaporthe oryzae. Nat Rev Microbiol. 2009;7(3):185-95. doi: 10.1038/nrmicro2032. PubMed PMID: 19219052.

13. Talbot NJ. On the trail of a cereal killer: Exploring the biology of Magnaporthe grisea. Annu Rev Microbiol. 2003;57:177-202. doi: 10.1146/annurev.micro.57.030502.090957. PubMed PMID: 14527276.

14. Zeigler RS, Leong SA, Peng PS. Rice blast disease. CAB International, Wallingford, UK. 1994.

15. Skamnioti P, Gurr SJ. Against the grain: safeguarding rice from rice blast disease. Trends Biotechnol. 2009;27(3):141-50. doi: 10.1016/j.tibtech.2008.12.002. PubMed PMID: 19187990.

16. Barr ME. Magnaporthe, Telimenella, and Hyponectria (Physosporellaceae). Mycologia. 1977;69(5):952-66. doi: Doi 10.2307/3758778. PubMed PMID: WOS:A1977EE42400006.

17. Zhang H, Zheng X, Zhang Z. The Magnaporthe grisea species complex and plant pathogenesis. Mol Plant Pathol. 2016;17(6):796-804. Epub 2015/11/18. doi: 10.1111/mpp.12342. PubMed PMID: 26575082.

18. Ou SH. Pathogen Variability and Host-Resistance in Rice Blast Disease. Annual Review of Phytopathology. 1980;18:167-87. doi: DOI 10.1146/annurev.py.18.090180.001123. PubMed PMID: WOS:A1980KF31300011.

19. Cruz CD, Valent B. Wheat blast disease: danger on the move. Trop Plant Pathol. 2017;42(3):210-22. doi: 10.1007/s40858-017-0159-z. PubMed PMID: WOS:000403934200010.

20. Inoue Y, Vy TTP, Yoshida K, Asano H, Mitsuoka C, Asuke S, et al. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. Science. 2017;357(6346):80-3. doi: 10.1126/science.aam9654. PubMed PMID: 28684523.

21. Silva CP, Nomura E, Freitas EG, Brugnaro C, Urashima AS. Efficiency of alternative treatments in the control of Pyricularia grisea in wheat seeds. Trop Plant Pathol. 2009;34(2):127-31. PubMed PMID: WOS:000270261000009.

22. Islam MT, Croll D, Gladieux P, Soanes DM, Persoons A, Bhattacharjee P, et al. Emergence of wheat blast in Bangladesh was caused by a South American lineage of Magnaporthe oryzae. Bmc Biol. 2016;14. doi: ARTN 84

10.1186/s12915-016-0309-7. PubMed PMID: WOS:000384374900001.

23. Sadat MA, Choi J. Wheat Blast: A New Fungal Inhabitant to Bangladesh Threatening World Wheat Production. Plant Pathol J. 2017;33(2):103-8. Epub 2017/04/07. doi: 10.5423/PPJ.RW.09.2016.0179. PubMed PMID: 28381956; PubMed Central PMCID: PMCPMC5378430.

24. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, et al. The genome sequence of the rice blast fungus Magnaporthe grisea. Nature. 2005;434(7036):980-6. doi: 10.1038/nature03449. PubMed PMID: 15846337.

25. International Rice Genome Sequencing P. The map-based sequence of the rice genome. Nature. 2005;436(7052):793-800. doi: 10.1038/nature03895. PubMed PMID: 16100779.

26. Hamer JE, Howard RJ, Chumley FG, Valent B. A mechanism for surface attachment in spores of a plant pathogenic fungus. Science. 1988;239(4837):288-90. doi: 10.1126/science.239.4837.288. PubMed PMID: 17769992.

27. Talbot NJ. Having a blast: exploring the pathogenicity of Magnaporthe grisea. Trends Biotechnol. 1995;3(1):9-16.

28. Dean RA. Signal pathways and appressorium morphogenesis. Annu Rev Phytopathol. 1997;35:211-34. doi: 10.1146/annurev.phyto.35.1.211. PubMed PMID: 15012522.

29. Howard RJ, Ferrari MA, Roach DH, Money NP. Penetration of hard substrates by a fungus employing enormous turgor pressures. Proceedings of the National Academy of Sciences of the United States of America. 1991;88(24):11281-4. PubMed PMID: 1837147; PubMed Central PMCID: PMCPMC53118.

30. Ryder LS, Talbot NJ. Regulation of appressorium development in pathogenic fungi. Curr Opin Plant Biol. 2015;26:8-13. doi: 10.1016/j.pbi.2015.05.013. PubMed PMID: 26043436; PubMed Central PMCID: PMCPMC4781897.

31. Talbot NJ, Kershaw MJ. The emerging role of autophagy in plant pathogen attack and host defence. Curr Opin Plant Biol. 2009;12(4):444-50. Epub 2009/07/25. doi: 10.1016/j.pbi.2009.05.008. PubMed PMID: 19625208.

32. Veneault-Fourrey C, Barooah M, Egan M, Wakley G, Talbot NJ. Autophagic fungal cell death is necessary for infection by the rice blast fungus. Science. 2006;312(5773):580-3. doi: 10.1126/science.1124550. PubMed PMID: 16645096.

33. Oses-Ruiz M, Sakulkoo W, Littlejohn GR, Martin-Urdiroz M, Talbot NJ. Two independent S-phase checkpoints regulate appressorium-mediated plant infection by the rice blast fungus Magnaporthe oryzae. Proc Natl Acad Sci U S A. 2017;114(2):E237-E44. Epub 2016/12/29. doi: 10.1073/pnas.1611307114. PubMed PMID: 28028232; PubMed Central PMCID: PMCPMC5240714.

34. Bourett TM, Howard RJ. Invitro Development of Penetration Structures in the Rice Blast Fungus Magnaporthe-Grisea. Can J Bot. 1990;68(2):329-42. PubMed PMID: WOS:A1990CW63000014.

35. Chumley FG, Valent B. Genetic-Analysis of Melanin-Deficient, Nonpathogenic Mutants of Magnaporthe-Grisea. Molecular Plant-Microbe Interactions. 1990;3(3):135-43. doi: Doi 10.1094/Mpmi-3-135. PubMed PMID: WOS:A1990DD45200001.

36. deJong JC, MCCormack BJ, Smirnoff N, Talbot NJ. Glycerol generates turgor in rice blast. Nature. 1997;389(6648):244-5. doi: Doi 10.1038/38418. PubMed PMID: WOS:A1997XW77200029.

37. Faino L, Seidl MF, Datema E, van den Berg GC, Janssen A, Wittenberg AH, et al. Single-Molecule Real-Time Sequencing Combined with Optical Mapping Yields Completely Finished Fungal Genome. MBio. 2015;6(4). doi: 10.1128/mBio.00936-15. PubMed PMID: 26286689; PubMed Central PMCID: PMCPMC4542186.

38. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why bigger can be better. Nature reviews Microbiology.

2012;10(6):417-30. Epub 2012/05/09. doi: 10.1038/nrmicro2790. PubMed PMID: 22565130.

39. Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C, et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc Natl Acad Sci U S A. 2011;108(22):9166-71. Epub 2011/05/04. doi: 10.1073/pnas.1019315108. PubMed PMID: 21536894; PubMed Central PMCID: PMCPMC3107277.

40. Cantu D, Govindarajulu M, Kozik A, Wang M, Chen X, Kojima KK, et al. Next generation sequencing provides rapid access to the genome of Puccinia striiformis f. sp. tritici, the causal agent of wheat stripe rust. Plos One. 2011;6(8):e24230. Epub 2011/09/13. doi: 10.1371/journal.pone.0024230. PubMed PMID: 21909385; PubMed Central PMCID: PMCPMC3164196.

41. Bao JD, Chen ML, Zhong ZH, Tang W, Lin LY, Zhang XT, et al. PacBio Sequencing Reveals Transposable Elements as a Key Contributor to Genomic Plasticity and Virulence Variation in Magnaporthe oryzae. Molecular plant. 2017;10(11):1465-8. doi: 10.1016/j.molp.2017.08.008. PubMed PMID: WOS:000414599700010.

42. Wu J, Kou Y, Bao J, Li Y, Tang M, Zhu X, et al. Comparative genomics identifies the Magnaporthe oryzae avirulence effector AvrPi9 that triggers Pi9-mediated blast resistance in rice. New Phytol. 2015;206(4):1463-75. doi: 10.1111/nph.13310. PubMed PMID: 25659573.

43. Xue M, Yang J, Li Z, Hu S, Yao N, Dean RA, et al. Comparative analysis of the genomes of two field isolates of the rice blast fungus Magnaporthe

oryzae. PLoS Genet. 2012;8(8):e1002869. doi: 10.1371/journal.pgen.1002869. PubMed PMID: 22876203; PubMed Central PMCID: PMCPMC3410873.

44. Chiapello H, Mallet L, Guerin C, Aguileta G, Amselem J, Kroj T, et al. Deciphering Genome Content and Evolutionary Relationships of Isolates from the Fungus Magnaporthe oryzae Attacking Different Host Plants. Genome Biol Evol. 2015;7(10):2896-912. Epub 2015/10/11. doi: 10.1093/gbe/evv187. PubMed PMID: 26454013; PubMed Central PMCID: PMCPMC4684704.

45. Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GC, Wittenberg AH, et al. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. Genome Res. 2016;26(8):1091-100. doi: 10.1101/gr.204974.116. PubMed PMID: 27325116; PubMed Central PMCID: PMCPMC4971763.

46. Yoshida K., Saitoh H., Fujisawa S., Kanzaki H., Matsumura H., Yoshida K., et al. Association Genetics Reveals Three Novel Avirulence Genes from the Rice Blast Fungal Pathogen Magnaporthe oryzae. 2009;21:1573–91.

47. Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature. 2010;464(7287):367-73. Epub 2010/03/20. doi: 10.1038/nature08850. PubMed PMID: 20237561; PubMed Central PMCID: PMCPMC3048781.

48. Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, Grimwood J, et al. The genome of Nectria haematococca: contribution of supernumerary chromosomes to gene expansion. PLoS Genet.

2009;5(8):e1000618. Epub 2009/08/29. doi: 10.1371/journal.pgen.1000618. PubMed PMID: 19714214; PubMed Central PMCID: PMCPMC2725324.

49. Stukenbrock EH, Jorgensen FG, Zala M, Hansen TT, McDonald BA, Schierup MH. Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen Mycosphaerella graminicola. PLoS Genet. 2010;6(12):e1001189. Epub 2011/01/05. doi: 10.1371/journal.pgen.1001189. PubMed PMID: 21203495; PubMed Central PMCID: PMCPMC3009667.

50. Johnson LJ, Johnson RD, Akamatsu H, Salamiah A, Otani H, Kohmoto K, et al. Spontaneous loss of a conditionally dispensable chromosome from the Alternaria alternata apple pathotype leads to loss of toxin production and pathogenicity. Curr Genet. 2001;40(1):65-72. Epub 2001/09/26. PubMed PMID: 11570518.

51. Tzeng TH, Lyngholm LK, Ford CF, Bronson CR. A restriction fragment length polymorphism map and electrophoretic karyotype of the fungal maize pathogen Cochliobolus heterostrophus. Genetics. 1992;130(1):81-96. Epub 1992/01/01. PubMed PMID: 1346261; PubMed Central PMCID: PMCPMC1204808.

52. Leclair S, Ansan-Melayah D, Rouxel T, Balesdent M. Meiotic behaviour of the minichromosome in the phytopathogenic ascomycete Leptosphaeria maculans. Curr Genet. 1996;30(6):541-8. Epub 1996/12/01. PubMed PMID: 8939816.

53. Talbot NJ, Ebbole DJ, Hamer JE. Identification and characterization of MPG1, a gene involved in pathogenicity from the rice blast fungus Magnaporthe grisea. The Plant cell. 1993;5(11):1575-90. Epub 1993/11/01. doi: 10.1105/tpc.5.11.1575. PubMed PMID: 8312740; PubMed Central PMCID: PMCPMC160387.

54. Talbot NJ, Salch YP, Ma M, Hamer JE. Karyotypic Variation within Clonal Lineages of the Rice Blast Fungus, Magnaporthe-Grisea. Appl Environ Microb. 1993;59(2):585-93. PubMed PMID: WOS:A1993KK91600037.

55. Chen C, Lian B, Hu J, Zhai H, Wang X, Venu RC, et al. Genome comparison of two Magnaporthe oryzae field isolates reveals genome variations and potential virulence effectors. Bmc Genomics. 2013;14:887. Epub 2013/12/18. doi: 10.1186/1471-2164-14-887. PubMed PMID: 24341723; PubMed Central PMCID: PMCPMC3878650.

56. Yi M, Valent B. Communication Between Filamentous Pathogens and Plants at the Biotrophic Interface. Annual Review of Phytopathology, Vol 51. 2013;51:587-611. doi: 10.1146/annurev-phyto-081211-172916. PubMed PMID: WOS:000323889000027.

57. Giraldo MC, Valent B. Filamentous plant pathogen effectors in action. Nat Rev Microbiol. 2013;11(11):800-14. doi: 10.1038/nrmicro3119. PubMed PMID: 24129511.

58. Dodds PN, Rathjen JP. Plant immunity: towards an integrated view of plant-pathogen interactions. Nat Rev Genet. 2010;11(8):539-48. doi: 10.1038/nrg2812. PubMed PMID: 20585331.

59. Jones JDG, Dangl JL. The plant immune system. Nature. 2006;444:323–9.

60. Sanchez-Vallet A, Mesters JR, Thomma BP. The battle for chitin recognition in plant-microbe interactions. FEMS microbiology reviews. 2015;39(2):171-83. doi: 10.1093/femsre/fuu003. PubMed PMID: 25725011.

61. Jones JD, Dangl JL. The plant immune system. Nature. 2006;444(7117):323-9. doi: 10.1038/nature05286. PubMed PMID: 17108957.

62. Zipfel C, Oldroyd GE. Plant signalling in symbiosis and immunity. Nature. 2017;543(7645):328-36. Epub 2017/03/17. doi: 10.1038/nature22009. PubMed PMID: 28300100.

63. Boller T, Felix G. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. Annual review of plant biology. 2009;60:379-406. Epub 2009/04/30. doi: 10.1146/annurev.arplant.57.032905.105346. PubMed PMID: 19400727.

64. Fliegmann J, Felix G. IMMUNITY Flagellin seen from all sides. Nat Plants. 2016;2(9). doi: Artn 16136

10.1038/Nplants.2016.136. PubMed PMID: WOS:000385011300011.

65. Furukawa T, Inagaki H, Takai R, Hirai H, Che FS. Two distinct EF-Tu epitopes induce immune responses in rice and Arabidopsis. Mol Plant Microbe Interact. 2014;27(2):113-24. Epub 2013/11/10. doi: 10.1094/MPMI-10-13-0304-R. PubMed PMID: 24200076.

66. Liu T, Liu Z, Song C, Hu Y, Han Z, She J, et al. Chitin-induced dimerization activates a plant immune receptor. Science. 2012;336(6085):1160-4. doi: 10.1126/science.1218867. PubMed PMID: 22654057.

67. Miya A, Albert P, Shinya T, Desaki Y, Ichimura K, Shirasu K, et al. CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(49):19613-8. doi: 10.1073/pnas.0705147104. PubMed PMID: 18042724; PubMed Central PMCID: PMC2148337.

68. Shimizu T, Nakano T, Takamizawa D, Desaki Y, Ishii-Minami N, Nishizawa Y, et al. Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. The Plant journal : for cell and molecular biology. 2010;64(2):204-14. doi: 10.1111/j.1365-313X.2010.04324.x. PubMed PMID: 21070404; PubMed Central PMCID: PMC2996852.

69. Liu W, Liu J, Ning Y, Ding B, Wang X, Wang Z, et al. Recent progress in understanding PAMP- and effector-triggered immunity against the rice blast fungus Magnaporthe oryzae. Mol Plant. 2013;6(3):605-20. doi: 10.1093/mp/sst015. PubMed PMID: 23340743.

70. Takken FLW, Goverse A. How to build a pathogen detector: structural basis of NB-LRR function. Curr Opin Plant Biol. 2012;15(4):375-84. doi: 10.1016/j.pbi.2012.05.001. PubMed PMID: WOS:000308055600005.

71. Bernoux M, Ve T, Williams S, Warren C, Hatters D, Valkov E, et al. Structural and functional analysis of a plant resistance protein TIR domain

reveals interfaces for self-association, signaling, and autoregulation. Cell Host Microbe. 2011;9(3):200-11. doi: 10.1016/j.chom.2011.02.009. PubMed PMID: 21402359; PubMed Central PMCID: PMCPMC3142617.

72. Qi D, Innes RW. Recent Advances in Plant NLR Structure, Function, Localization, and Signaling. Front Immunol. 2013;4:348. doi: 10.3389/fimmu.2013.00348. PubMed PMID: 24155748; PubMed Central PMCID: PMCPMC3801107.

73. Maekawa T, Cheng W, Spiridon LN, Toller A, Lukasik E, Saijo Y, et al. Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. Cell Host Microbe. 2011;9(3):187-99. doi: 10.1016/j.chom.2011.02.008. PubMed PMID: 21402358.

74. Maqbool A, Saitoh H, Franceschetti M, Stevenson CE, Uemura A, Kanzaki H, et al. Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. Elife. 2015;4. doi: 10.7554/eLife.08709. PubMed PMID: 26304198; PubMed Central PMCID: PMCPMC4547098.

75. Eitas TK, Dangl JL. NB-LRR proteins: pairs, pieces, perception, partners, and pathways. Curr Opin Plant Biol. 2010;13(4):472-7. doi: 10.1016/j.pbi.2010.04.007. PubMed PMID: 20483655; PubMed Central PMCID: PMCPMC2910844.

76. Wu CH, Krasileva KV, Banfield MJ, Terauchi R, Kamoun S. The "sensor domains" of plant NLR proteins: more than decoys? Front Plant Sci. 2015;6. doi: ARTN 134

10.3389/fpls.2015.00134. PubMed PMID: WOS:000350968400001.

77. Bialas A, Zess EK, De la Concepcion JC, Franceschetti M, Pennington HG, Yoshida K, et al. Lessons in Effector and NLR Biology of Plant-Microbe Systems. Mol Plant Microbe Interact. 2017:MPMI08170196FI. Epub 2017/11/17. doi: 10.1094/MPMI-08-17-0196-FI. PubMed PMID: 29144205.

78. Stergiopoulos I, van den Burg HA, Okmen B, Beenen HG, van Liere S, Kema GH, et al. Tomato Cf resistance proteins mediate recognition of cognate homologous effectors from fungi pathogenic on dicots and monocots. Proc Natl Acad Sci U S A. 2010;107(16):7610-5. Epub 2010/04/07. doi: 10.1073/pnas.1002910107. PubMed PMID: 20368413; PubMed Central PMCID: PMCPMC2867746.

79. Dixon MS, Jones DA, Keddie JS, Thomas CM, Harrison K, Jones JD. The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell. 1996;84(3):451-9. Epub 1996/02/09. PubMed PMID: 8608599.

80. Cesari S, Bernoux M, Moncuquet P, Kroj T, Dodds PN. A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. Front Plant Sci. 2014;5. doi: UNSP 606

10.3389/fpls.2014.00606. PubMed PMID: WOS:000347323700001.

81. Williams SJ, Sohn KH, Wan L, Bernoux M, Sarris PF, Segonzac C, et al. Structural basis for assembly and function of a heterodimeric plant immune receptor. Science. 2014;344(6181):299-303. doi: 10.1126/science.1247357. PubMed PMID: 24744375.

82. Fukuoka S, Yamamoto SI, Mizobuchi R, Yamanouchi U, Ono K, Kitazawa N, et al. Multiple functional polymorphisms in a single disease resistance gene in rice enhance durable resistance to blast. Sci Rep-Uk. 2014;4. doi: ARTN 4550

10.1038/srep04550. PubMed PMID: WOS:000333555300015.

83. Cesari S, Kanzaki H, Fujiwara T, Bernoux M, Chalvon V, Kawano Y, et al. The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. EMBO J. 2014;33(17):1941-59. doi: 10.15252/embj.201487923. PubMed PMID: 25024433; PubMed Central PMCID: PMCPMC4195788.

84. Cesari S, Thilliez G, Ribot C, Chalvon V, Michel C, Jauneau A, et al. The rice resistance protein pair RGA4/RGA5 recognizes the Magnaporthe oryzae effectors AVR-Pia and AVR1-CO39 by direct binding. Plant Cell. 2013;25(4):1463-81. doi: 10.1105/tpc.112.107201. PubMed PMID: 23548743; PubMed Central PMCID: PMCPMC3663280.

85. Kanzaki H, Yoshida K, Saitoh H, Fujisaki K, Hirabuchi A, Alaux L, et al. Arms race co-evolution of Magnaporthe oryzae AVR-Pik and rice Pik genes driven by their physical interactions. Plant Journal. 2012;72(6):894-907. doi: 10.1111/j.1365-313X.2012.05110.x. PubMed PMID: WOS:000314182000003.

86. Fukuoka S, Saka N, Koga H, Ono K, Shimizu T, Ebana K, et al. Loss of function of a proline-containing protein confers durable disease resistance in rice. Science. 2009;325(5943):998-1001. Epub 2009/08/22. doi: 10.1126/science.1175550. PubMed PMID: 19696351.

87. Gabriels SHEJ, Vossen JH, Ekengren SK, van Ooijen G, Abd-El-Haliem AM, van den Berg GCM, et al. An NB-LRR protein required for HR signalling mediated by both extra- and intracellular resistance proteins. Plant Journal. 2007;50(1):14-28. doi: 10.1111/j.1365-313X.2007.03027.x. PubMed PMID: WOS:000245258300002.

88. Bonardi V, Tang S, Stallmann A, Roberts M, Cherkis K, Dangl JL. Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(39):16463-8. doi: 10.1073/pnas.1113726108. PubMed PMID: 21911370; PubMed Central PMCID: PMCPMC3182704.

89. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal effectors and plant susceptibility. Annu Rev Plant Biol. 2015;66:513-45. doi: 10.1146/annurev-arplant-043014-114623. PubMed PMID: 25923844.

90. Stergiopoulos I, Collemare J, Mehrabi R, De Wit PJ. Phytotoxic secondary metabolites and peptides produced by plant pathogenic Dothideomycete fungi. FEMS microbiology reviews. 2013;37(1):67-93. Epub 2012/08/31. doi: 10.1111/j.1574-6976.2012.00349.x. PubMed PMID: 22931103.

91. de Jonge R, Bolton MD, Thomma BPHJ. How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. Curr Opin Plant Biol. 2011;14(4):400-6. doi: 10.1016/j.pbi.2011.03.005. PubMed PMID: WOS:000294525400009.

92. O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, et al. Lifestyle transitions in plant pathogenic Colletotrichum fungi deciphered by genome and transcriptome analyses. Nat Genet. 2012;44(9):1060-5. Epub 2012/08/14. doi: 10.1038/ng.2372. PubMed PMID: 22885923.

93. Jones K, Kim DW, Park JS, Khang CH. Live-cell fluorescence imaging to investigate the dynamics of plant cell death during infection by the rice blast fungus Magnaporthe oryzae. BMC Plant Biol. 2016;16:69. doi: 10.1186/s12870-016-0756-x. PubMed PMID: 27000073; PubMed Central PMCID: PMCPMC4802709.

94. Mentlak TA, Kombrink A, Shinya T, Ryder LS, Otomo I, Saitoh H, et al. Effector-mediated suppression of chitin-triggered immunity by magnaporthe oryzae is necessary for rice blast disease. Plant Cell. 2012;24(1):322-35. doi: 10.1105/tpc.111.092957. PubMed PMID: 22267486; PubMed Central PMCID: PMCPMC3289562.

95. Saitoh H, Fujisawa S, Mitsuoka C, Ito A, Hirabuchi A, Ikeda K, et al. Large-scale gene disruption in Magnaporthe oryzae identifies MC69, a secreted protein required for infection by monocot and dicot fungal pathogens. PLoS

Pathog. 2012;8(5):e1002711. doi: 10.1371/journal.ppat.1002711. PubMed PMID: 22589729; PubMed Central PMCID: PMCPMC3349759.

96. Bohnert HU, Fudal I, Dioh W, Tharreau D, Notteghem JL, Lebrun MH. A putative polyketide synthase/peptide synthetase from Magnaporthe grisea signals pathogen attack to resistant rice. Plant Cell. 2004;16(9):2499-513. doi: 10.1105/tpc.104.022715. PubMed PMID: 15319478; PubMed Central PMCID: PMCPMC520948.

97. Rafiqi M, Ellis JG, Ludowici VA, Hardham AR, Dodds PN. Challenges and progress towards understanding the role of effectors in plant-fungal interactions. Curr Opin Plant Biol. 2012;15(4):477-82. Epub 2012/06/05. doi: 10.1016/j.pbi.2012.05.003. PubMed PMID: 22658704.

98. Hogenhout SA, Van der Hoorn RA, Terauchi R, Kamoun S. Emerging concepts in effector biology of plant-associated organisms. Mol Plant Microbe Interact. 2009;22(2):115-22. Epub 2009/01/10. doi: 10.1094/MPMI-22-2-0115. PubMed PMID: 19132864.

99. Giraldo MC, Valent B. Filamentous plant pathogen effectors in action. Nature Reviews Microbiology. 2013;11(11):800-14. doi: 10.1038/nrmicro3119. PubMed PMID: WOS:000325797200013.

100. Jelenska J, van Hal JA, Greenberg JT. Pseudomonas syringae hijacks plant stress chaperone machinery for virulence. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(29):13177-82. doi: 10.1073/pnas.0910943107. PubMed PMID: 20615948; PubMed Central PMCID: PMC2919979.

101. Doehlemann G, van der Linde K, Assmann D, Schwammbach D, Hof A, Mohanty A, et al. Pep1, a secreted effector protein of Ustilago maydis, is required for successful invasion of plant cells. PLoS pathogens. 2009;5(2):e1000290. doi: 10.1371/journal.ppat.1000290. PubMed PMID: 19197359; PubMed Central PMCID: PMC2631132.

102. Djamei A, Schipper K, Rabe F, Ghosh A, Vincon V, Kahnt J, et al. Metabolic priming by a secreted fungal effector. Nature. 2011;478(7369):395-8. Epub 2011/10/07. doi: 10.1038/nature10454. PubMed PMID: 21976020.

103. Tanaka S, Brefort T, Neidig N, Djamei A, Kahnt J, Vermerris W, et al. A secreted Ustilago maydis effector promotes virulence by targeting anthocyanin biosynthesis in maize. Elife. 2014;3:e01355. doi: 10.7554/eLife.01355. PubMed PMID: 24473076; PubMed Central PMCID: PMCPMC3904489.

104. Gonzalez-Lamothe R, Tsitsigiannis DI, Ludwig AA, Panicot M, Shirasu K, Jones JD. The U-box protein CMPG1 is required for efficient activation of defense mechanisms triggered by multiple resistance genes in tobacco and tomato. The Plant cell. 2006;18(4):1067-83. doi: 10.1105/tpc.106.040998. PubMed PMID: 16531490; PubMed Central PMCID: PMC1425846.

105. Park CH, Chen S, Shirsekar G, Zhou B, Khang CH, Songkumarn P, et al. The Magnaporthe oryzae effector AvrPiz-t targets the RING E3 ubiquitin ligase APIP6 to suppress pathogen-associated molecular pattern-triggered immunity in rice. Plant Cell. 2012;24(11):4748-62. doi: 10.1105/tpc.112.105429. PubMed PMID: 23204406; PubMed Central PMCID: PMCPMC3531864.

106. Wang R, Ning Y, Shi X, He F, Zhang C, Fan J, et al. Immunity to Rice Blast Disease by Suppression of Effector-Triggered Necrosis. Curr Biol. 2016;26(18):2399-411. doi: 10.1016/j.cub.2016.06.072. PubMed PMID: 27641772.

107. Fujisaki K, Abe Y, Ito A, Saitoh H, Yoshida K, Kanzaki H, et al. Rice Exo70 interacts with a fungal effector, AVR-Pii, and is required for AVR-Pii-triggered immunity. Plant Journal. 2015;83(5):875-87. doi: 10.1111/tpj.12934. PubMed PMID: WOS:000360100200011.

108. Singh R, Dangol S, Chen Y, Choi J, Cho YS, Lee JE, et al. Magnaporthe oryzae Effector AVR-Pii Helps to Establish Compatibility by Inhibition of the Rice NADP-Malic Enzyme Resulting in Disruption of Oxidative Burst and Host Innate Immunity. Mol Cells. 2016;39(5):426-38. doi: 10.14348/molcells.2016.0094. PubMed PMID: 27126515; PubMed Central PMCID: PMCPMC4870191.

109. Kleemann J, Rincon-Rivera LJ, Takahara H, Neumann U, Ver Loren van Themaat E, van der Does HC, et al. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen Colletotrichum higginsianum. PLoS Pathog. 2012;8(4):e1002643. doi: 10.1371/journal.ppat.1002643. PubMed PMID: 22496661; PubMed Central PMCID: PMCPMC3320591.

110. Mogga V, Delventhal R, Weidenbach D, Langer S, Bertram PM, Andresen K, et al. Magnaporthe oryzae effectors MoHEG13 and MoHEG16

interfere with host infection and MoHEG13 counteracts cell death caused by Magnaporthe-NLPs in tobacco. Plant Cell Reports. 2016;35(5):1169–85.

111. Mueller AN, Ziemann S, Treitschke S, Assmann D, Doehlemann G. Compatibility in the Ustilago maydis-maize interaction requires inhibition of host cysteine proteases by the fungal effector Pit2. PLoS Pathog. 2013;9(2):e1003177. doi: 10.1371/journal.ppat.1003177. PubMed PMID: 23459172; PubMed Central PMCID: PMCPMC3573112.

112. Rooney HC, Van't Klooster JW, van der Hoorn RA, Joosten MH, Jones JD, de Wit PJ. Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance. Science. 2005;308(5729):1783-6. doi: 10.1126/science.1111404. PubMed PMID: 15845874.

113. Song J, Win J, Tian M, Schornack S, Kaschani F, Ilyas M, et al. Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(5):1654-9. doi: 10.1073/pnas.0809201106. PubMed PMID: 19171904; PubMed Central PMCID: PMCPMC2635833.

114. Tian M, Win J, Song J, van der Hoorn R, van der Knaap E, Kamoun S. A Phytophthora infestans cystatin-like protein targets a novel tomato papain-like apoplastic protease. Plant Physiol. 2007;143(1):364-77. doi: 10.1104/pp.106.090050. PubMed PMID: 17085509; PubMed Central PMCID: PMCPMC1761951.

115. Sharpee W, Oh Y, Yi M, Franck W, Eyre A, Okagaki LH, et al. Identification and characterization of suppressors of plant cell death (SPD) effectors from Magnaporthe oryzae. Mol Plant Pathol. 2017;18(6):850-63. Epub 2016/06/16. doi: 10.1111/mpp.12449. PubMed PMID: 27301772.

116. Dong Y, Li Y, Zhao M, Jing M, Liu X, Liu M, et al. Global genome and transcriptome analyses of Magnaporthe oryzae epidemic isolate 98-06 uncover novel effectors and pathogenicity-related genes, revealing gene gain and lose dynamics in genome evolution. PLoS Pathog. 2015;11(4):e1004801. doi: 10.1371/journal.ppat.1004801. PubMed PMID: 25837042; PubMed Central PMCID: PMCPMC4383609.

117. Kang S.C., Sweigard J.A., B V. The PWL host specificity gene family in the blast fungus Magnaporthe grisea. Molecular Plant-Microbe Interactions 1995;8:939-48.

118. Sweigard JA, Carroll AM, Kang S, Farrall L, Chumley FG, Valent B. Identification, cloning, and characterization of PWL2, a gene for host species specificity in the rice blast fungus. The Plant cell. 1995;7:1221-33.

119. Dawkins R, Krebs JR. Arms races between and within species. Proceedings of the Royal Society of London Series B, Biological sciences. 1979;205(1161):489-511. PubMed PMID: 42057.

120. Thrall PH, Laine AL, Ravensdale M, Nemri A, Dodds PN, Barrett LG, et al. Rapid genetic change underpins antagonistic coevolution in a natural host-pathogen metapopulation. Ecol Lett. 2012;15(5):425-35. Epub 2012/03/01. doi:

10.1111/j.1461-0248.2012.01749.x. PubMed PMID: 22372578; PubMed Central PMCID: PMCPMC3319837.

121. Dodds PN, Lawrence GJ, Catanzariti AM, Teh T, Wang CI, Ayliffe MA, et al. Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(23):8888-93. Epub 2006/05/30. doi: 10.1073/pnas.0602577103. PubMed PMID: 16731621; PubMed Central PMCID: PMCPMC1482673.

122. van der Hoorn RA, Kamoun S. From Guard to Decoy: a new model for perception of plant pathogen effectors. Plant Cell. 2008;20(8):2009-17. doi: 10.1105/tpc.108.060194. PubMed PMID: 18723576; PubMed Central PMCID: PMCPMC2553620.

123. Huang J, Si W, Deng Q, Li P, Yang S. Rapid evolution of avirulence genes in rice blast fungus Magnaporthe oryzae. BMC Genet. 2014;15:45. doi: 10.1186/1471-2156-15-45. PubMed PMID: 24725999; PubMed Central PMCID: PMCPMC4021558.

124. Zhang S, Wang L, Wu W, He L, Yang X, Pan Q. Function and evolution of Magnaporthe oryzae avirulence gene AvrPib responding to the rice blast resistance gene Pib. Sci Rep. 2015;5:11642. doi: 10.1038/srep11642. PubMed PMID: 26109439.

125. Orbach MJ, Farrall L, Sweigard JA, Chumley FG, Valent B. A telomeric avirulence gene determines efficacy for the rice blast resistance gene Pi-ta. The

Plant cell. 2000;12(11):2019-32. PubMed PMID: 11090206; PubMed Central PMCID: PMCPMC152363.

126. Hacquard S, Kracher B, Maekawa T, Vernaldi S, Schulze-Lefert P, Ver Loren van Themaat E. Mosaic genome structure of the barley powdery mildew pathogen and conservation of transcriptional programs in divergent hosts. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(24):E2219-28. Epub 2013/05/23. doi: 10.1073/pnas.1306807110. PubMed PMID: 23696672; PubMed Central PMCID: PMCPMC3683789.

127. Joly DL, Feau N, Tanguay P, Hamelin RC. Comparative analysis of secreted protein evolution using expressed sequence tags from four poplar leaf rusts (Melampsora spp.). Bmc Genomics. 2010;11. doi: Artn 422

10.1186/1471-2164-11-422. PubMed PMID: WOS:000282785900002.

128. Sharma R, Mishra B, Runge F, Thines M. Gene Loss Rather Than Gene Gain Is Associated with a Host Jump from Monocots to Dicots in the Smut Fungus Melanopsichium pennsylvanicum. Genome Biol Evol. 2014;6(8):2034-49. doi: 10.1093/gbe/evu148. PubMed PMID: WOS:000342976600011.

129. Rep M, Kistler HC. The genomic organization of plant pathogenicity in Fusarium species. Curr Opin Plant Biol. 2010;13(4):420-6. Epub 2010/05/18. doi: 10.1016/j.pbi.2010.04.004. PubMed PMID: 20471307.

130. Franceschetti M, Maqbool A, Jimenez-Dalmaroni MJ, Pennington HG, Kamoun S, Banfield MJ. Effectors of Filamentous Plant Pathogens:

Commonalities amid Diversity. Microbiol Mol Biol Rev. 2017;81(2). Epub 2017/03/31. doi: 10.1128/MMBR.00066-16. PubMed PMID: 28356329; PubMed Central PMCID: PMCPMC5485802.

131. Dong S, Stam R, Cano LM, Song J, Sklenar J, Yoshida K, et al. Effector specialization in a lineage of the Irish potato famine pathogen. Science. 2014;343(6170):552-5. doi: 10.1126/science.1246300. PubMed PMID: 24482481.

132. Chuma I, Isobe C, Hotta Y, Ibaragi K, Futamata N, Kusaba M, et al. Multiple Translocation of the AVR-Pita Effector Gene among Chromosomes of the Rice Blast Fungus Magnaporthe oryzae and Related Species. PLoS pathogens. 2011;7(7). doi: ARTN e1002147

10.1371/journal.ppat.1002147. PubMed PMID: WOS:000293339300035.

133. Valent B, Chumley FG. Molecular genetic analysis of the rice blast fungus, magnaporthe grisea. Annu Rev Phytopathol. 1991;29:443-67. doi: 10.1146/annurev.py.29.090191.002303. PubMed PMID: 18479196.

134. Tosa Y, Hirata K, Tamba H, Nakagawa S, Chuma I, Isobe C, et al. Genetic Constitution and Pathogenicity of Lolium Isolates of Magnaporthe oryzae in Comparison with Host Species-Specific Pathotypes of the Blast Fungus. Phytopathology. 2004;94(5):454-62. doi: 10.1094/PHYTO.2004.94.5.454. PubMed PMID: 18943763.

135. Tosa Y, Tamba H, Tanaka K, Mayama S. Genetic Analysis of Host Species Specificity of Magnaporthe oryzae Isolates from Rice and Wheat.

Phytopathology. 2006;96(5):480-4. doi: 10.1094/PHYTO-96-0480. PubMed PMID: 18944307.

136. Tosa Y, Osue J, Eto Y, Oh HS, Nakayashiki H, Mayama S, et al. Evolution of an avirulence gene, AVR1-CO39, concomitant with the evolution and differentiation of Magnaporthe oryzae. Mol Plant Microbe Interact. 2005;18(11):1148-60. doi: 10.1094/MPMI-18-1148. PubMed PMID: 16353550.

137. Doehlemann G, Hemetsberger C. Apoplastic immunity and its suppression by filamentous plant pathogens. The New phytologist. 2013;198(4):1001-16. Epub 2013/04/19. doi: 10.1111/nph.12277. PubMed PMID: 23594392.

138. Tyler BM, Kale SD, Wang QQ, Tao K, Clark HR, Drews K, et al. Microbe-Independent Entry of Oomycete RxLR Effectors and Fungal RxLR-Like Effectors Into Plant and Animal Cells Is Specific and Reproducible. Molecular Plant-Microbe Interactions. 2013;26(6):611-6. doi: 10.1094/Mpmi-02-13-0051-Ia. PubMed PMID: WOS:000318960300001.

139. Stergiopoulos I, de Wit PJ. Fungal effector proteins. Annu Rev Phytopathol. 2009;47:233-63. doi: 10.1146/annurev.phyto.112408.132637. PubMed PMID: 19400631.

140. Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell RJ, Narusaka Y, et al. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of Colletotrichum fungi. The New phytologist. 2013;197(4):1236-49. Epub 2012/12/21. doi: 10.1111/nph.12085. PubMed PMID: 23252678.

141. Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stuber K, et al. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science. 2010;330(6010):1543-6. Epub 2010/12/15. doi: 10.1126/science.1194573. PubMed PMID: 21148392.

142. Kobayashi N, Telebanco-Yanoria MJ, Tsunematsu H, Kato H, Imbe T, Fukuta Y. Development of new sets of international standard differential varieties for blast resistance in rice (Oryza sativa L.). Jarq-Jpn Agr Res Q. 2007;41(1):31-7. PubMed PMID: WOS:000245634900007.

143. Tsunematsu H, Yanoria MJT, Ebron LA, Hayashi N, Ando I, Kato H, et al. Development of monogenic lines of rice for blast resistance. Breeding Sci. 2000;50(3):229-34. PubMed PMID: WOS:000089862800012.

144. Valent B, Farrall L, Chumley FG. Magnaporthe-Grisea Genes for Pathogenicity and Virulence Identified through a Series of Backcrosses. Genetics. 1991;127(1):87-101. PubMed PMID: WOS:A1991EP81100009.

145. Kankanala P, Czymmek K, Valent B. Roles for rice membrane dynamics and plasmodesmata during biotrophic invasion by the blast fungus. Plant Cell. 2007;19(2):706-24. doi: 10.1105/tpc.106.046300. PubMed PMID: 17322409; PubMed Central PMCID: PMCPMC1867340.

146. Katzen F. Gateway (R) recombinational cloning: a biological operating system. Expert Opin Drug Dis. 2007;2(4):571-89. doi: 10.1517/17460441.2.4.571. PubMed PMID: WOS:000207616700012.

147. Southern E. Southern blotting. Nat Protoc. 2006;1(2):518-25. Epub 2007/04/05. doi: 10.1038/nprot.2006.73. PubMed PMID: 17406277.

148. Sambrook J F, E.F., and Maniatis, T Molecular Cloning - a laboratory manual, second edition. . Cold Spring Harbour Laboratory press. 1989.

149. Feinberg AP, Vogelstein B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem. 1983;132(1):6-13. Epub 1983/07/01. PubMed PMID: 6312838.

150. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(4):R36. Epub 2013/04/27. doi: 10.1186/gb-2013-14-4-r36. PubMed PMID: 23618408; PubMed Central PMCID: PMCPMC4053844.

151. Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, et al. A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science. 2002;296(5565):79-92. Epub 2002/04/06. doi: 10.1126/science.1068037. PubMed PMID: 11935017.

152. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106. Epub 2010/10/29. doi: 10.1186/gb-2010-11-10-r106. PubMed PMID: 20979621; PubMed Central PMCID: PMCPMC3218662.

153. Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-9. doi: 10.1093/bioinformatics/btu638. PubMed PMID: WOS:000347832300003.

154. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. Epub 2009/05/20. doi: 10.1093/bioinformatics/btp324. PubMed PMID: 19451168; PubMed Central PMCID: PMCPMC2705234.

155. Holsters M, Silva B, Van Vliet F, Genetello C, De Block M, Dhaese P, et al. The functional organization of the nopaline A. tumefaciens plasmid pTiC58. Plasmid. 1980;3(2):212-30. Epub 1980/03/01. PubMed PMID: 6100894.

156. Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature. 1989;340(6230):245-6. Epub 1989/07/20. doi: 10.1038/340245a0. PubMed PMID: 2547163.

157. WN B. Burnette WN. “Western Blotting”: electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein. Analytical Biochemistry. 1981;112(2):195 - 203.

158. de Jonge R, Bolton MD, Kombrink A, van den Berg GC, Yadeta KA, Thomma BP. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res. 2013;23(8):1271-82. Epub 2013/05/21. doi: 10.1101/gr.152660.112. PubMed PMID: 23685541; PubMed Central PMCID: PMCPMC3730101.

159. Croll D, Zala M, McDonald BA. Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. PLoS Genet. 2013;9(6):e1003567. Epub 2013/06/21. doi:

10.1371/journal.pgen.1003567. PubMed PMID: 23785303; PubMed Central PMCID: PMCPMC3681731.

160. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, et al. Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae. Nature. 2005;438(7071):1105-15. Epub 2005/12/24. doi: 10.1038/nature04341. PubMed PMID: 16372000.

161. Wapinski I, Pfeffer A, Friedman N, Regev A. Natural history and evolutionary principles of gene duplication in fungi. Nature. 2007;449(7158):54-61. Epub 2007/09/07. doi: 10.1038/nature06107. PubMed PMID: 17805289.

162. Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM, et al. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature. 2009;461(7262):393-8. Epub 2009/09/11. doi: 10.1038/nature08358. PubMed PMID: 19741609.

163. Gardiner DM, McDonald MC, Covarelli L, Solomon PS, Rusu AG, Marshall M, et al. Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. PLoS pathogens. 2012;8(9):e1002952. Epub 2012/10/03. doi: 10.1371/journal.ppat.1002952. PubMed PMID: 23028337; PubMed Central PMCID: PMCPMC3460631.

164. Soanes DM, Richards TA, Talbot NJ. Insights from sequencing fungal and oomycete genomes: What can we learn about plant disease and the evolution of pathogenicity? The Plant cell. 2007;19(11):3318-26. doi: 10.1105/tpc.107.056663. PubMed PMID: WOS:000252268700002.

165. Xu JR, Peng YL, Dickman MB, Sharon A. The dawn of fungal pathogen genomics. Annu Rev Phytopathol. 2006;44:337-66. Epub 2006/05/18. doi: 10.1146/annurev.phyto.44.070505.143412. PubMed PMID: 16704358.

166. Soanes DM, Alam I, Cornell M, Wong HM, Hedeler C, Paton NW, et al. Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. Plos One. 2008;3(6):e2300. Epub 2008/06/05. doi: 10.1371/journal.pone.0002300. PubMed PMID: 18523684; PubMed Central PMCID: PMCPMC2409186.

167. Stergiopoulos I, Kourmpetis YA, Slot JC, Bakker FT, De Wit PJ, Rokas A. In silico characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. Mol Biol Evol. 2012;29(11):3371-84. Epub 2012/05/26. doi: 10.1093/molbev/mss143. PubMed PMID: 22628532.

168. Beare PA, Unsworth N, Andoh M, Voth DE, Omsland A, Gilk SD, et al. Comparative Genomics Reveal Extensive Transposon-Mediated Genomic Plasticity and Diversity among Potential Effector Proteins within the Genus Coxiella. Infect Immun. 2009;77(2):642-56. doi: 10.1128/Iai.01141-08. PubMed PMID: WOS:000262777300009.

169. Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, et al. Effector diversification within compartments of the Leptosphaeria maculans genome affected by Repeat-Induced Point mutations. Nature Communications. 2011;2. doi: 10.1038/ncomms1189. PubMed PMID: WOS:000288225900031.

170. Grandaubert J, Lowe RGT, Soyer JL, Schoch CL, de Wouw APV, Fudal I, et al. Transposable element-assisted evolution and adaptation to host plant within the Leptosphaeria maculans-Leptosphaeria biglobosa species complex of fungal pathogens. Bmc Genomics. 2014;15. doi: Artn 891

10.1186/1471-2164-15-891. PubMed PMID: WOS:000344566000001.

171. Raffaele S, Farrer RA, Cano LM, Studholme DJ, MacLean D, Thines M, et al. Genome evolution following host jumps in the Irish potato famine pathogen lineage. Science. 2010;330(6010):1540-3. Epub 2010/12/15. doi: 10.1126/science.1193070. PubMed PMID: 21148391.

172. Okagaki LH, Nunes CC, Sailsbery J, Clay B, Brown D, John T, et al. Genome Sequences of Three Phytopathogenic Species of the Magnaporthaceae Family of Fungi. G3 (Bethesda). 2015;5(12):2539-45. Epub 2015/09/30. doi: 10.1534/g3.115.020057. PubMed PMID: 26416668; PubMed Central PMCID: PMCPMC4683626.

173. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. Plos One. 2012;7(11):e47768. Epub 2012/11/28. doi: 10.1371/journal.pone.0047768. PubMed PMID: 23185243; PubMed Central PMCID: PMCPMC3504050.

174. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013;10(6):563-9. doi: 10.1038/nmeth.2474. PubMed PMID: 23644548.

175. Worley KC. A golden goat genome. Nat Genet. 2017;49(4):485-6. Epub 2017/03/31. doi: 10.1038/ng.3824. PubMed PMID: 28358125.

176. Levy-Sakin M, Ebenstein Y. Beyond sequencing: optical mapping of DNA in the age of nanotechnology and nanoscopy. Curr Opin Biotechnol. 2013;24(4):690-8. Epub 2013/02/23. doi: 10.1016/j.copbio.2013.01.009. PubMed PMID: 23428595.

177. Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nat Biotechnol. 2013;31(2):135-41. Epub 2012/12/25. doi: 10.1038/nbt.2478. PubMed PMID: 23263233.

178. De Wit PJ, Mehrabi R, Van den Burg HA, Stergiopoulos I. Fungal effector proteins: past, present and future. Mol Plant Pathol. 2009;10(6):735-47. Epub 2009/10/24. doi: 10.1111/j.1364-3703.2009.00591.x. PubMed PMID: 19849781.

179. Sperschneider J, Dodds PN, Gardiner DM, Manners JM, Singh KB, Taylor JM. Advances and challenges in computational prediction of effectors from plant pathogenic fungi. PLoS pathogens. 2015;11(5):e1004806. Epub 2015/05/29. doi: 10.1371/journal.ppat.1004806. PubMed PMID: 26020524; PubMed Central PMCID: PMCPMC4447458.

180. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8(10):785-6. Epub 2011/10/01. doi: 10.1038/nmeth.1701. PubMed PMID: 21959131.

181. Martin F, Aerts A, Ahren D, Brun A, Danchin EGJ, Duchaussoy F, et al. The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis. Nature. 2008;452(7183):88-U7. doi: 10.1038/nature06556. PubMed PMID: WOS:000253671900052.

182. Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, et al. EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytologist. 2016;210(2):743-61. doi: 10.1111/nph.13794. PubMed PMID: WOS:000373380700034.

183. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77. Epub 2012/04/18. doi: 10.1089/cmb.2012.0021. PubMed PMID: 22506599; PubMed Central PMCID: PMCPMC3342519.

184. Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol. 2012;30(8):771-6. Epub 2012/07/17. doi: 10.1038/nbt.2303. PubMed PMID: 22797562; PubMed Central PMCID: PMCPMC3817024.

185. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 2005;33:W465-W7. doi: 10.1093/nar/gki458. PubMed PMID: WOS:000230271400094.

186. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinformatics. 2011;12:124. Epub 2011/04/30. doi: 10.1186/1471-2105-12-124. PubMed PMID: 21526987; PubMed Central PMCID: PMCPMC3114741.

187. Nielsen H, Engelbrecht J, Brunak S, vonHeijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering. 1997;10(1):1-6. doi: DOI 10.1093/protein/10.1.1. PubMed PMID: WOS:A1997WJ04100001.

188. Testa AC, Hane JK, Ellwood SR, Oliver RP. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. Bmc Genomics. 2015;16. doi: ARTN 170

10.1186/s12864-015-1344-4. PubMed PMID: WOS:000350947400002.

189. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008;18(1):188-96. doi: 10.1101/gr.6743907. PubMed PMID: 18025269; PubMed Central PMCID: PMCPMC2134774.

190. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5:59. Epub 2004/05/18. doi: 10.1186/1471-2105-5-59. PubMed PMID: 15144565; PubMed Central PMCID: PMCPMC421630.

191. Feng J, Meyer CA, Wang Q, Liu JS, Shirley Liu X, Zhang Y. GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-

seq data. Bioinformatics. 2012;28(21):2782-8. Epub 2012/08/28. doi: 10.1093/bioinformatics/bts515. PubMed PMID: 22923299.

192. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674-6. Epub 2005/08/06. doi: 10.1093/bioinformatics/bti610. PubMed PMID: 16081474.

193. Jarvis P. Targeting of nucleus-encoded proteins to chloroplasts in plants. The New phytologist. 2008;179(2):257-85. Epub 2008/12/17. PubMed PMID: 19086173.

194. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols. 2007;2(4):953-71. doi: 10.1038/nprot.2007.131. PubMed PMID: WOS:000253138600024.

195. Hicks SW, Galan JE. Exploitation of eukaryotic subcellular targeting mechanisms by bacterial effectors. Nature reviews Microbiology. 2013;11(5):316-26. Epub 2013/04/17. doi: 10.1038/nrmicro3009. PubMed PMID: 23588250; PubMed Central PMCID: PMCPMC3859125.

196. Caillaud MC, Piquerez SJM, Fabro G, Steinbrenner J, Ishaque N, Beynon J, et al. Subcellular localization of the Hpa RxLR effector repertoire identifies a tonoplast-associated protein HaRxL17 that confers enhanced plant susceptibility. Plant Journal. 2012;69(2):252-65. doi: 10.1111/j.1365-313X.2011.04787.x. PubMed PMID: WOS:000298874300006.

197. Schornack S, van Damme M, Bozkurt TO, Cano LM, Smoker M, Thines M, et al. Ancient class of translocated oomycete effectors targets the host nucleus. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(40):17421-6. Epub 2010/09/18. doi: 10.1073/pnas.1008491107. PubMed PMID: 20847293; PubMed Central PMCID: PMCPMC2951462.

198. Rivas S, Genin S. A plethora of virulence strategies hidden behind nuclear targeting of microbial effectors. Front Plant Sci. 2011;2:104. Epub 2011/01/01. doi: 10.3389/fpls.2011.00104. PubMed PMID: 22639625; PubMed Central PMCID: PMCPMC3355726.

199. Petre B, Lorrain C, Saunders DGO, Win J, Sklenar J, Duplessis S, et al. Rust fungal effectors mimic host transit peptides to translocate into chloroplasts. Cell Microbiol. 2016;18(4):453-65. doi: 10.1111/cmi.12530. PubMed PMID: WOS:000373066100001.

200. Sperschneider J, Catanzariti AM, DeBoer K, Petre B, Gardiner DM, Singh KB, et al. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Sci Rep-Uk. 2017;7. doi: ARTN 44598

10.1038/srep44598. PubMed PMID: WOS:000396543100002.

201. Rep M. Small proteins of plant-pathogenic fungi secreted during host colonization. FEMS Microbiol Lett. 2005;253(1):19-27. Epub 2005/10/12. doi: 10.1016/j.femsle.2005.09.014. PubMed PMID: 16216445.

202. Ciuffetti LM, Manning VA, Pandelova I, Betts MF, Martinez JP. Host-selective toxins, Ptr ToxA and Ptr ToxB, as necrotrophic effectors in the Pyrenophora tritici-repentis-wheat interaction. New Phytologist. 2010;187(4):911-9. doi: 10.1111/j.1469-8137.2010.03362.x. PubMed PMID: WOS:000280998600008.

203. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14(7):1394-403. Epub 2004/07/03. doi: 10.1101/gr.2289704. PubMed PMID: 15231754; PubMed Central PMCID: PMCPMC442156.

204. Ames RM. Using Network Extracted Ontologies to Identify Novel Genes with Roles in Appressorium Development in the Rice Blast Fungus Magnaporthe oryzae. Microorganisms. 2017;5(3).

205. Jacob S, Grotsch T, Foster AJ, Schuffler A, Rieger PH, Sandjo LP, et al. Unravelling the biosynthesis of pyriculol in the rice blast fungus Magnaporthe oryzae. Microbiology. 2017;163(4):541-53. Epub 2016/12/03. doi: 10.1099/mic.0.000396. PubMed PMID: 27902426.

206. Parker D, Beckmann M, Zubair H, Enot DP, Caracuel-Rios Z, Overy DP, et al. Metabolomic analysis reveals a common pattern of metabolic re-programming during invasion of three host plant species by Magnaporthe grisea. The Plant journal : for cell and molecular biology. 2009;59(5):723-37. Epub 2009/05/21. doi: 10.1111/j.1365-313X.2009.03912.x. PubMed PMID: 19453445.

207. Petre B, Saunders DG, Sklenar J, Lorrain C, Win J, Duplessis S, et al. Candidate Effector Proteins of the Rust Pathogen Melampsora larici-populina Target Diverse Plant Cell Compartments. Mol Plant Microbe Interact. 2015;28(6):689-700. doi: 10.1094/MPMI-01-15-0003-R. PubMed PMID: 25650830.

208. Kroj T, Chanclud E, Michel-Romiti C, Grand X, Morel JB. Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. New Phytologist. 2016;210(2):618-26. doi: 10.1111/nph.13869. PubMed PMID: WOS:000373380700024.

209. Sarris PF, Cevik V, Dagdas G, Jones JD, Krasileva KV. Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. BMC Biol. 2016;14:8. Epub 2016/02/20. doi: 10.1186/s12915-016-0228-7. PubMed PMID: 26891798; PubMed Central PMCID: PMCPMC4759884.

210. de Guillen K, Ortiz-Vallejo D, Gracy J, Fournier E, Kroj T, Padilla A. Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi. PLoS Pathog. 2015;11(10):e1005228. doi: 10.1371/journal.ppat.1005228. PubMed PMID: 26506000; PubMed Central PMCID: PMCPMC4624222.

211. Axtell MJ, Chisholm ST, Dahlbeck D, Staskawicz BJ. Genetic and molecular evidence that the Pseudomonas syringae type III effector protein AvrRpt2 is a cysteine protease. Molecular Microbiology. 2003;49(6):1537-46.

doi: 10.1046/j.1365-2958.2003.03666.x. PubMed PMID: WOS:000185190700007.

212. Mackey D, Holt BF, Wiig A, Dangl JL. RIN4 interacts with Pseudomonas syringae type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis. Cell. 2002;108(6):743-54. doi: Doi 10.1016/S0092-8674(02)00661-X. PubMed PMID: WOS:000174563000005.

213. Andersson MX, Kourtchenko O, Dangl JL, Mackey D, Ellerstrom M. Phospholipase-dependent signalling during the AvrRpm1- and AvrRpt2-induced disease resistance responses in Arabidopsis thaliana. The Plant journal : for cell and molecular biology. 2006;47(6):947-59. Epub 2006/08/24. doi: 10.1111/j.1365-313X.2006.02844.x. PubMed PMID: 16925603.

214. Sarris PF, Duxbury Z, Huh SU, Ma Y, Segonzac C, Sklenar J, et al. A Plant Immune Receptor Detects Pathogen Effectors that Target WRKY Transcription Factors. Cell. 2015;161(5):1089-100. doi: 10.1016/j.cell.2015.04.024. PubMed PMID: WOS:000355152600016.

215. Le Roux C, Huet G, Jauneau A, Camborde L, Tremousaygue D, Kraut A, et al. A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. Cell. 2015;161(5):1074-88. doi: 10.1016/j.cell.2015.04.025. PubMed PMID: 26000483.

216. Weidenbach D, Esch L, Moller C, Hensel G, Kumlehn J, Hofle C, et al. Polarized Defense Against Fungal Pathogens Is Mediated by the Jacalin-Related Lectin Domain of Modular Poaceae-Specific Proteins. Molecular Plant.

2016;9(4):514-27. doi: 10.1016/j.molp.2015.12.009. PubMed PMID: WOS:000373742700005.

217. Sone T, Takeuchi S, Miki S, Satoh Y, Ohtsuka K, Abe A, et al. Homologous recombination causes the spontaneous deletion of AVR-Pia in Magnaporthe oryzae. FEMS Microbiol Lett. 2013;339(2):102-9. doi: 10.1111/1574-6968.12058. PubMed PMID: 23198972.

218. Yoshida K, Saunders DGO, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, et al. Host specialization of the blast fungus Magnaporthe oryzae is associated with dynamic gain and loss of genes linked to transposable elements. Bmc Genomics. 2016;17. doi: ARTN 370

10.1186/s12864-016-2690-6. PubMed PMID: WOS:000376348900001.

219. Macho AP, Zipfel C. Targeting of plant pattern recognition receptor-triggered immunity by bacterial type-III secretion system effectors. Curr Opin Microbiol. 2015;23:14-22. Epub 2014/12/03. doi: 10.1016/j.mib.2014.10.009. PubMed PMID: 25461568.

220. Win J, Krasileva KV, Kamoun S, Shirasu K, Staskawicz BJ, Banfield MJ. Sequence divergent RXLR effectors share a structural fold conserved across plant pathogenic oomycete species. PLoS pathogens. 2012;8(1):e1002400. Epub 2012/01/19. doi: 10.1371/journal.ppat.1002400. PubMed PMID: 22253591; PubMed Central PMCID: PMCPMC3257287.

221. Rovenich H, Boshoven JC, Thomma BP. Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. Curr Opin Plant Biol. 2014;20:96-103. doi: 10.1016/j.pbi.2014.05.001. PubMed PMID: 24879450.

222. Qutob D, Kemmerling B, Brunner F, Kufner I, Engelhardt S, Gust AA, et al. Phytotoxicity and innate immune responses induced by Nep1-like proteins. The Plant cell. 2006;18(12):3721-44. Epub 2006/12/30. doi: 10.1105/tpc.106.044180. PubMed PMID: 17194768; PubMed Central PMCID: PMCPMC1785393.

223. Stotz HU, Mitrousia GK, de Wit PJ, Fitt BD. Effector-triggered defence against apoplastic fungal pathogens. Trends Plant Sci. 2014;19(8):491-500. Epub 2014/05/27. doi: 10.1016/j.tplants.2014.04.009. PubMed PMID: 24856287; PubMed Central PMCID: PMCPMC4123193.

224. O'Connell RJ, Panstruga R. Tete a tete inside a plant cell: establishing compatibility between plants and biotrophic fungi and oomycetes. New Phytologist. 2006;171(4):699-718. doi: DOI 10.1111/j.1469-8137.2006.01829.x. PubMed PMID: WOS:000239517600004.

225. Ilyas M, Horger AC, Bozkurt TO, van den Burg HA, Kaschani F, Kaiser M, et al. Functional Divergence of Two Secreted Immune Proteases of Tomato. Curr Biol. 2015;25(17):2300-6. Epub 2015/08/25. doi: 10.1016/j.cub.2015.07.030. PubMed PMID: 26299516.

226. Keen NT, Tamaki S, Kobayashi D, Gerhold D, Stayton M, Shen H, et al. Bacteria Expressing Avirulence Gene D Produce a Specific Elicitor of the Soybean Hypersensitive Reaction. Molecular Plant-Microbe Interactions.

1990;3(2):112-21. doi: Doi 10.1094/Mpmi-3-112. PubMed PMID: WOS:A1990CT87700008.

227. Miah G, Rafii MY, Ismail MR, Puteh AB, Rahim HA, Asfaliza R, et al. Blast resistance in rice: a review of conventional breeding to molecular approaches. Mol Biol Rep. 2013;40(3):2369-88. Epub 2012/11/28. doi: 10.1007/s11033-012-2318-0. PubMed PMID: 23184051.

228. McDonald BA, Linde C. Pathogen population genetics, evolutionary potential, and durable resistance. Annu Rev Phytopathol. 2002;40:349-79. Epub 2002/07/31. doi: 10.1146/annurev.phyto.40.120501.101443. PubMed PMID: 12147764.

229. de Wit PJ. Molecular characterization of gene-for-gene systems in plant-fungus interactions and the application of avirulence genes in control of plant pathogens. Annu Rev Phytopathol. 1992;30:391-418. Epub 1992/01/01. doi: 10.1146/annurev.py.30.090192.002135. PubMed PMID: 18647100.

230. Fukuta Y, Araki E, Yanoria MJT, Imbe T, Tsunematsu H, Kato H, et al. Development of differential varieties for blast resistance in IRRI-Japan Collaborative Research Project. Rice Blast: Interaction with Rice and Control. 2004:229-33. PubMed PMID: WOS:000224548200027.

231. Mutiga SK, Rotich F, Ganeshan VD, Mwongera DT, Mgonja EM, Were VM, et al. Assessment of the Virulence Spectrum and Its Association with Genetic Diversity in Magnaporthe oryzae Populations from Sub-Saharan Africa. Phytopathology. 2017;107(7):852-63. Epub 2017/04/04. doi: 10.1094/PHYTO-08-16-0319-R. PubMed PMID: 28368237.

232. Tharreau D, Fudal I, Andriantsimialona D, Santoso, Utami D, Fournier E, et al. World Population Structure and Migration of the Rice Blast Fungus, Magnaporthe oryzae. Advances in Genetics, Genomics and Control of Rice Blast Disease. 2009:209-+. doi: 10.1007/978-1-4020-9500-9_21. PubMed PMID: WOS:000266407200021.

233. Liu S, Dean RA. G protein alpha subunit genes control growth, development, and pathogenicity of Magnaporthe grisea. Mol Plant Microbe Interact. 1997;10(9):1075-86. Epub 1997/12/09. doi: 10.1094/MPMI.1997.10.9.1075. PubMed PMID: 9390422.

234. Adachi K, Hamer JE. Divergent cAMP signaling pathways regulate growth and pathogenesis in the rice blast fungus Magnaporthe grisea. The Plant cell. 1998;10(8):1361-74. Epub 1998/08/26. PubMed PMID: 9707535; PubMed Central PMCID: PMCPMC144070.

235. Li GT, Zhang X, Tian H, Choi YE, Tao WA, Xu JR. MST50 is involved in multiple MAP kinase signaling pathways in Magnaporthe oryzae. Environ Microbiol. 2017;19(5):1959-74. doi: 10.1111/1462-2920.13710. PubMed PMID: WOS:000400029800018.

236. Qi L, Kim Y, Jiang C, Li Y, Peng Y, Xu JR. Activation of Mst11 and Feedback Inhibition of Germ Tube Growth in Magnaporthe oryzae. Mol Plant Microbe Interact. 2015;28(8):881-91. doi: 10.1094/MPMI-12-14-0391-R. PubMed PMID: 26057388.

237. Chen JS, Zheng W, Zheng SQ, Zhang DM, Sang WJ, Chen X, et al. Rac1 Is Required for Pathogenicity and Chm1-Dependent Conidiogenesis in

Rice Fungal Pathogen Magnaporthe grisea. PLoS pathogens. 2008;4(11). doi: ARTN e1000202

10.1371/journal.ppat.1000202. PubMed PMID: WOS:000261481200007.

238. Nishimura M, Park G, Xu JR. The G-beta subunit MGB1 is involved in regulating multiple steps of infection-related morphogenesis in Magnaporthe grisea. Molecular Microbiology. 2003;50(1):231-43. doi: 10.1046/j.1365-2958.2003.03676.x. PubMed PMID: WOS:000185446300019.

239. Kim S, Park SY, Kim KS, Rho HS, Chi MH, Choi J, et al. Homeobox transcription factors are required for conidiation and appressorium development in the rice blast fungus Magnaporthe oryzae. PLoS Genet. 2009;5(12):e1000757. Epub 2009/12/10. doi: 10.1371/journal.pgen.1000757. PubMed PMID: 19997500; PubMed Central PMCID: PMCPMC2779367.

240. Choi W, Dean RA. The adenylate cyclase gene MAC1 of Magnaporthe grisea controls appressorium formation and other aspects of growth and development. The Plant cell. 1997;9(11):1973-83. Epub 1997/12/24. doi: 10.1105/tpc.9.11.1973. PubMed PMID: 9401122; PubMed Central PMCID: PMCPMC157051.

241. Sabnam N, Barman SR. WISH, a novel CFEM GPCR is indispensable for surface sensing, asexual and pathogenic differentiation in rice blast fungus. Fungal Genet Biol. 2017;105:37-51. doi: 10.1016/j.fgb.2017.05.006. PubMed PMID: WOS:000406827300005.

242. Bruno KS, Tenjo F, Li L, Hamer JE, Xu JR. Cellular localization and role of kinase activity of PMK1 in Magnaporthe grisea. Eukaryot Cell. 2004;3(6):1525-32. Epub 2004/12/14. doi: 10.1128/EC.3.6.1525-1532.2004. PubMed PMID: 15590826; PubMed Central PMCID: PMCPMC539019.

243. Rayi S, Singh PK, Gupta DK, Mahato AK, Sarkar C, Rathour R, et al. Analysis of Magnaporthe oryzae Genome Reveals a Fungal Effector, Which Is Able to Induce Resistance Response in Transgenic Rice Line Containing Resistance Gene, Pi54. Front Plant Sci. 2016;7. doi: ARTN 1140

10.3389/fpls.2016.01140. PubMed PMID: WOS:000380923400001.

244. Mukhtar MS, Carvunis AR, Dreze M, Epple P, Steinbrenner J, Moore J, et al. Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science. 2011;333(6042):596-601. Epub 2011/07/30. doi: 10.1126/science.1203659. PubMed PMID: 21798943; PubMed Central PMCID: PMCPMC3170753.

245. Kamoun S. Groovy times: filamentous pathogen effectors revealed. Curr Opin Plant Biol. 2007;10(4):358-65. Epub 2007/07/06. doi: 10.1016/j.pbi.2007.04.017. PubMed PMID: 17611143.

246. Dagdas Y.F. PP, Sanguankiattichai., Tumtas Y. Belhaj., Duggan C., Segretin M.E., Kamoun S., Bozkur T.o.,. Host autophagosomes are diverted to a plant-pathogen interface. bioRxiv. 2017.

247. Maqbool A, Hughes RK, Dagdas YF, Tregidgo N, Zess E, Belhaj K, et al. Structural Basis of Host Autophagy-related Protein 8 (ATG8) Binding by the

Irish Potato Famine Pathogen Effector Protein PexRD54. J Biol Chem. 2016;291(38):20270-82. Epub 2016/07/28. doi: 10.1074/jbc.M116.744995. PubMed PMID: 27458016; PubMed Central PMCID: PMCPMC5025708.

248. Bozkurt TO, Schornack S, Banfield MJ, Kamoun S. Oomycetes, effectors, and all that jazz. Curr Opin Plant Biol. 2012;15(4):483-92. doi: 10.1016/j.pbi.2012.03.008. PubMed PMID: WOS:000308055600019.

249. Alfano JR. Roadmap for future research on plant pathogen effectors. Mol Plant Pathol. 2009;10(6):805-13. Epub 2009/10/24. doi: 10.1111/j.1364-3703.2009.00588.x. PubMed PMID: 19849786; PubMed Central PMCID: PMCPMC2792923.

250. Mak AN, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL. The crystal structure of TAL effector PthXo1 bound to its DNA target. Science. 2012;335(6069):716-9. Epub 2012/01/10. doi: 10.1126/science.1216211. PubMed PMID: 22223736; PubMed Central PMCID: PMCPMC3427646.

251. Win J, Chaparro-Garcia A, Belhaj K, Saunders DG, Yoshida K, Dong S, et al. Effector biology of plant-associated organisms: concepts and perspectives. Cold Spring Harb Symp Quant Biol. 2012;77:235-47. Epub 2012/12/12. doi: 10.1101/sqb.2012.77.015933. PubMed PMID: 23223409.

252. Torto TA, Li S, Styer A, Huitema E, Testa A, Gow NA, et al. EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen Phytophthora. Genome Res. 2003;13(7):1675-85. Epub 2003/07/04. doi: 10.1101/gr.910003. PubMed PMID: 12840044; PubMed Central PMCID: PMCPMC403741.

253. Du J, Rietman H, Vleeshouwers VG. Agroinfiltration and PVX agroinfection in potato and Nicotiana benthamiana. J Vis Exp. 2014;(83):e50971. Epub 2014/01/17. doi: 10.3791/50971. PubMed PMID: 24430891; PubMed Central PMCID: PMCPMC4063549.

254. Bozkurt TO, Schornack S, Win J, Shindo T, Ilyas M, Oliva R, et al. Phytophthora infestans effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(51):20832-7. doi: 10.1073/pnas.1112708109. PubMed PMID: WOS:000298289400109.

255. Win J, Kamoun S, Jones AM. Purification of effector-target protein complexes via transient expression in Nicotiana benthamiana. Methods Mol Biol. 2011;712:181-94. Epub 2011/03/02. doi: 10.1007/978-1-61737-998-7_15. PubMed PMID: 21359809.

256. Leigh H. THE RELATIVE PALATABILITY OF VARIOUS VARIETIES OF WEEPING LOVEGRASS (ERAGROSTIS CURVULA (SCHRAD) NEES). Grass and Forage Science. 1961;16(2

):135–40.

257. Schneider DR, Saraiva AM, Azzoni AR, Miranda HR, de Toledo MA, Pelloso AC, et al. Overexpression and purification of PWL2D, a mutant of the effector protein PWL2 from Magnaporthe grisea. Protein Expr Purif. 2010;74(1):24-31. Epub 2010/05/05. doi: 10.1016/j.pep.2010.04.020. PubMed PMID: 20438845.

258. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10. Epub 1990/10/05. doi: 10.1016/S0022-2836(05)80360-2. PubMed PMID: 2231712.

259. Cannarozzi G, Plaza-Wuthrich S, Esfeld K, Larti S, Wilson YS, Girma D, et al. Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (Eragrostis tef). Bmc Genomics. 2014;15:581. Epub 2014/07/11. doi: 10.1186/1471-2164-15-581. PubMed PMID: 25007843; PubMed Central PMCID: PMCPMC4119204.

260. Sung YH, Kim JM, Kim HT, Lee J, Jeon J, Jin Y, et al. Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. Genome Res. 2014;24(1):125-31. Epub 2013/11/21. doi: 10.1101/gr.163394.113. PubMed PMID: 24253447; PubMed Central PMCID: PMCPMC3875853.

261. Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife. 2014;3:e04766. Epub 2014/12/17. doi: 10.7554/eLife.04766. PubMed PMID: 25497837; PubMed Central PMCID: PMCPMC4383097.

262. Woo JW, Kim J, Kwon SI, Corvalan C, Cho SW, Kim H, et al. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat Biotechnol. 2015;33(11):1162-4. Epub 2015/10/20. doi: 10.1038/nbt.3389. PubMed PMID: 26479191.

263. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol. 2014;32(4):347-55. Epub 2014/03/04. doi:

10.1038/nbt.2842. PubMed PMID: 24584096; PubMed Central PMCID: PMCPMC4022601.

264. Pohl C, Kiel JA, Driessen AJ, Bovenberg RA, Nygard Y. CRISPR/Cas9 Based Genome Editing of Penicillium chrysogenum. ACS Synth Biol. 2016;5(7):754-64. Epub 2016/04/14. doi: 10.1021/acssynbio.6b00082. PubMed PMID: 27072635.

265. Ding YD, Li H, Chen LL, Xie KB. Recent Advances in Genome Editing Using CRISPR/Cas9. Front Plant Sci. 2016;7. doi: ARTN 703

10.3389/fpls.2016.00703. PubMed PMID: WOS:000376743300002.

266. Hink MA, Griep RA, Borst JW, van Hoek A, Eppink MH, Schots A, et al. Structural dynamics of green fluorescent protein alone and fused with a single chain Fv protein. J Biol Chem. 2000;275(23):17556-60. Epub 2000/04/05. doi: 10.1074/jbc.M001348200. PubMed PMID: 10748019.

267. Petre B, Saunders DGO, Sklenar J, Lorrain C, Krasileva KV, Win J, et al. Heterologous Expression Screens in Nicotiana benthamiana Identify a Candidate Effector of the Wheat Yellow Rust Pathogen that Associates with Processing Bodies. Plos One. 2016;11(2). doi: ARTN e0149035

10.1371/journal.pone.0149035. PubMed PMID: WOS:000370046600144.

268. Haag C, Steuten B, Feldbrugge M. Membrane-Coupled mRNA Trafficking in Fungi. Annu Rev Microbiol. 2015;69:265-81. Epub 2015/08/15. doi: 10.1146/annurev-micro-091014-104242. PubMed PMID: 26274025.

269. Jung H, Gkogkas CG, Sonenberg N, Holt CE. Remote control of gene function by local translation. Cell. 2014;157(1):26-40. Epub 2014/04/01. doi: 10.1016/j.cell.2014.03.005. PubMed PMID: 24679524; PubMed Central PMCID: PMCPMC3988848.

270. St Johnston D. Moving messages: the intracellular localization of mRNAs. Nat Rev Mol Cell Biol. 2005;6(5):363-75. Epub 2005/04/27. doi: 10.1038/nrm1643. PubMed PMID: 15852043.

271. Wang M, Weiberg A, Dellota E, Jr., Yamane D, Jin H. Botrytis small RNA Bc-siR37 suppresses plant defense genes by cross-kingdom RNAi. RNA Biol. 2017;14(4):421-8. Epub 2017/03/08. doi: 10.1080/15476286.2017.1291112. PubMed PMID: 28267415; PubMed Central PMCID: PMCPMC5411126.

272. Witek K, Jupe F, Witek AI, Baker D, Clark MD, Jones JDG. Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. Nature Biotechnology. 2016;34(6):656-60. doi: 10.1038/nbt.3540. PubMed PMID: WOS:000377846400034.

273. Jupe F, Witek K, Verweij W, Sliwka J, Pritchard L, Etherington GJ, et al. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. The Plant journal : for cell and molecular biology. 2013;76(3):530-44. Epub 2013/08/14. doi: 10.1111/tpj.12307. PubMed PMID: 23937694; PubMed Central PMCID: PMCPMC3935411.