

**The Genetic Basis of Resistance and  
Susceptibility in the *Albugo laibachii*- *Arabidopsis*  
*thaliana* pathosystem**

**Oliver John Furzer**

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

**Philosophy**

**BIO/TSL**

**September 2014**

© This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with the author and that no quotation from the thesis, nor any information derived therefrom, may be published without the author's prior written consent.

## TABLE OF CONTENTS

<b>ABSTRACT</b>	<b>5</b>
<b>ACKNOWLEDGMENTS</b>	<b>6</b>
<b>PUBLICATIONS ARISING FROM THIS THESIS</b>	<b>7</b>
Published papers	7
Papers in preparation	7
<b>MAJOR ABBREVIATIONS</b>	<b>8</b>
<b>CHAPTER 1: GENERAL INTRODUCTION</b>	<b>9</b>
1.1 Pattern-Triggered Immunity, the first line of defense	9
1.2 Pathogen evasion and suppression of PTI	11
1.3 Effector Triggered Immunity by Resistance proteins	13
1.4 Population genetics of gene-for-gene interactions	16
1.6 Oomycetes, effectors and genomics	18
1.7 <i>Albugo</i> species	21
1.8 Aims of this study	25
<b>CHAPTER 2A: STANDARD MATERIALS AND METHODS</b>	<b>26</b>
<b>2a.1 Molecular Methods</b>	<b>26</b>
Isolation of DNA for PCR	26
Rapid isolation of Arabidopsis DNA for screening	26
Isolation of <i>Albugo laibachii</i> DNA for next generation sequencing	26
Polymerase Chain Reaction (PCR)	27
Gel electrophoresis and gel extraction	27
DNA quantification	27
Isolation of <i>Albugo</i> RNA	28
Rapid Amplification of cDNA Ends-PCR	28
Gateway cloning	28
GoldenGate cloning	28
USER cloning	29
Quantitative PCR (qPCR)	29
Plasmid extraction	29
Sequencing of PCR products and plasmids	29
Next-generation sequencing	29

List of oligonucleotide primers	30
<b>2a.2 Microbial methods</b>	<b>30</b>
Isolation, Purification, Propagation and Screening of <i>Albugo laibachii</i>	30
Transformation and growth of <i>Escherichia coli</i>	31
Transformation of <i>Agrobacterium tumefaciens</i>	31
Triparental mating	31
Estimation of colony forming units / ml	32
<b>2a.3 Plant methods</b>	<b>32</b>
<i>Arabidopsis thaliana</i> growth	32
<i>Nicotiana</i> spp. growth	32
<i>Agrobacterium</i> transformation of Arabidopsis	32
Seed sterilization	32
<i>Agrobacterium tumefaciens</i> transient assays in <i>Nicotiana</i> spp.	33
<i>Pseudomonas syringae</i> infiltration and growth assays	33
Transient expression with biolistics	33
GUS staining and quantification	34
Trypan blue staining of <i>Albugo</i> structures	34
Confocal microscopy	34
List of plant accessions and mutants	34
<b>2a.4 Bioinformatics methods</b>	<b>35</b>
Quality assesment of Illumina reads	35
Read alignment	35
Variant calling	35
Genome alignment visualization	35
Prediction of the effects of polymorphisms	35
The Varitale pipeline	35
Identification of recombination events	36
Bayes-empirical-bayes analysis	36
Gene-density plotting in R	36
Genome assembly	36
BLAST	36
Genome coverage statistics	36
Alignments and phylogenetic trees	37
<b>CHAPTER 2B: DEVELOPMENT OF NOVEL GENOMICS TECHNIQUES</b>	<b>38</b>
<b>2b.1 Association genomics in plant pathogens</b>	<b>38</b>
Introduction	38
Methodology	39
<i>Hyaloperonospora arabidopsidis</i> as a test case	41
Conclusions	43
<b>2b.2 Brassicaceae R-gene enrichment sequencing</b>	<b>45</b>
Introduction	45
Methodology	46
Results	48
Conclusions	51

<b>CHAPTER 3: COLLECTION, IDENTIFICATION AND PHENOTYPING OF NEW <i>ALBUGO LAIBACHII</i> ISOLATES</b>	<b>52</b>
<b>3.1 Introduction</b>	<b>52</b>
<b>3.2 Results</b>	<b>53</b>
3.2.1 Collection of new isolates	53
3.2.2 Molecular identification of new isolates	54
3.2.3 Phenotyping of new isolates	55
<b>3.3 Discussion</b>	<b>56</b>
<b>CHAPTER 4: COMPARATIVE GENOMIC ANALYSIS OF SIX <i>ALBUGO LAIBACHII</i> ISOLATES AND PREDICTION OF RECOGNISED EFFECTORS</b>	<b>58</b>
<b>4.1 Introduction</b>	<b>58</b>
<b>4.2 Results</b>	<b>60</b>
4.2.1 Genome Sequencing of Six <i>Albugo laibachii</i> Isolates	60
4.2.2 Polymorphism in Six <i>Albugo laibachii</i> Isolates	61
4.2.3 Association Genomics in Six <i>Albugo laibachii</i> Isolates	63
4.2.4 Population-Genetic Analysis of Six <i>Albugo laibachii</i> Isolates	64
4.2.5 Analysis of Recombination in Six <i>Albugo laibachii</i> Isolates	70
<b>4.3 Discussion</b>	<b>71</b>
<b>CHAPTER 5: TESTING AND CHARACTERISATION OF POTENTIAL RECOGNISED EFFECTORS</b>	<b>75</b>
<b>5.1 Introduction</b>	<b>75</b>
<b>5.2 Results</b>	<b>76</b>
5.2.1 SSP16 <sup>Nc14</sup> is recognised by HR-5	76
5.2.2 SSP18 <sup>Nc14</sup> is recognised by Ksk-1	79
5.2.3 Allelic diversity of SSP16 and SSP18	81
5.2.4 Paralogs and Homologs of SSP16 and SSP18	86
5.2.5 The C-terminus of SSP16 <sup>Nc14</sup> is sufficient for recognition	89
5.2.6 The sub-cellular localisations of SSP16 and SSP18 in <i>N. benthamiana</i>	90
<b>5.3 Discussion</b>	<b>92</b>
<b>CHAPTER 6: IDENTIFICATION AND CHARACTERISATION OF THE <i>RAL4</i> LOCUS</b>	<b>95</b>
<b>6.1 Introduction</b>	<b>95</b>
<b>6.2 Results</b>	<b>95</b>
6.2.1 Co-segregation of SSP16 <sup>Nc14</sup> recognition and Nc14 resistance in HR-5	95
6.2.2 Mapping of the Resistance to <i>Albugo laibachii</i> 4 ( <i>RAL4</i> ) locus	97
6.2.3 <i>RAL4</i> candidate genes	99

6.2.4 Cloning and testing of <i>RAL4</i> candidate genes	101
6.2.5 R gene enrichment sequencing and Illumina based mapping	102
6.2.6 <i>RAL4</i> immunity cannot be suppressed by virulent <i>Albugo laibachii</i>	104
<b>6.3 Discussion</b>	<b>105</b>
<b>CHAPTER 7: GENERAL DISCUSSION AND OUTLOOK</b>	<b>109</b>
<b>7.1 Population variation in <i>Albugo laibachii</i></b>	<b>110</b>
<b>7.2 Association genomics to identify <i>Avr</i> gene candidates</b>	<b>111</b>
<b>7.3 Discovery of recognized effectors from <i>Albugo laibachii</i></b>	<b>112</b>
<b>7.3 <i>RAL4</i>; a genetic link between recognition and resistance</b>	<b>114</b>
<b>7.4 Conclusions and outlook</b>	<b>115</b>
<b>9. LIST OF REFERENCES</b>	<b>119</b>
<b>10. APPENDICES</b>	<b>134</b>

## ABSTRACT

*Albugo* is a genus of biotrophic plant pathogens that can infect an extensive range of hosts including many Brassicaceae crop species. Little is known about the molecular mechanisms by which *Albugo* species can suppress host immunity and the mechanisms by which plants can resist *Albugo* infection.

*Albugo laibachii* (*Al*) is a specialized pathogen of *Arabidopsis thaliana* (*At*). It can colonize ~90% of *At* accessions and suppress effector-triggered-immunity to other pathogens. It is postulated that *Al* secretes effector proteins. Analysis of the *A. laibachii* genome by Kemen et al, (2011, *PLoS Biology*) revealed a potential class of effectors with a 'CHXC' motif in their N-terminus that can mediate translocation into host cells. However, there are only ~35 CHXC effectors in *A. laibachii*, suggesting that they might not represent its entire effector complement.

I took a traditional method to identify *Al* effectors: clone "avirulence (*Avr*) genes". These typically encode effectors that are recognized and trigger a strong response by the immune system of some host accessions. I identified and sequenced four *Al* isolates from field samples. Using differential phenotype information to guide a genome-wide analysis, and my expectations of the allelic diversity of *Avr* genes, I identified two novel recognized effectors. These effectors, short secreted proteins named "SSP16" and "SSP18", are recognized by the *Arabidopsis* accessions HR-5 and Ksk-1 respectively.

I used classical and Illumina-based genetic mapping to identify the locus conferring SSP16 recognition in HR-5, *Resistance to A. laibachii 4 (RAL4)*. This locus contains three putative CC-NB-LRR class Resistance protein-encoding genes with similarity to *Resistance to Peronospora parasitica 7 (RPP7)*.

I demonstrated the utility of combined genomics approaches to identify recognized effectors without known motifs. The identification of the first *Avr-Resistance* gene pair will pave the way for further dissection of the molecular interactions in this pathosystem.

## ACKNOWLEDGMENTS

My first acknowledgment is to Professor Jonathan Jones. JJ has been a constant source of ideas, encouragement and guidance. I feel very fortunate to have been able to complete my PhD with such a positive, patient and intellectually stimulating Professor.

Professor Sophien Kamoun, who has been part of my supervisory committee, provided invaluable advice and insight. He has been an excellent example of a thoughtful and responsible scientific leader.

I am particularly grateful to Dr Eric Kemen. Eric was responsible for the formulation of my project and was my mentor in the lab for the first two years of my PhD. I feel that during this time, under Eric's guidance, I made the most rapid progress in my understanding of the process of scientific research.

I am also very grateful to Dr Volkan Cevik, who came into the lab at the mid-point of the project. Volkan has been an invaluable source of both practical and philosophical advice and took over the role of my mentor. Volkan is one of the few people I have met who can match JJ's enthusiasm for science!

I would like to thank the other members of the *Albugo* team who have given time and energy to help me: principally Dr Kate Bailey, Dr Ariane Kemen, Dr Alexandre Robert-Seilaniantz, Dr Mark McMullan and Dr Torsten Schultz-Larsen, but all of the others too. I also owe great thanks to various other members of the JJ lab, including but not limited to: Dr Shuta Asai, Dr Florian Jupe, Dr Kamil Witek, Dr Laurence Tomlinson and Dr Sophie Piquerez.

I've made many wonderful friends during this PhD. In particular I'd like to thank Dr Simon Saucet, (future Dr) Yan Ma, Dr Torsten Schultz-Larsen, Dr Naveed Ishaque and (future Dr) Zane Duxbury for making the experience much more enjoyable, as well as providing a sounding board for the discussion of ideas.

Finally, I'd like to sincerely thank my parents, family, and friends for their love and support. The knowledge that such a strong network existed behind me provided the motivation needed during difficult times.

## PUBLICATIONS ARISING FROM THIS THESIS

### PUBLISHED PAPERS

**The Top 10 oomycete pathogens in molecular plant pathology.** (2014) Kamoun S, **Furzer O**, Jones JD, Judelson HS, Ali GS, Dalio RJ, Roy SG, Schena L, Zambounis A, Panabières F, Cahill D, Ruocco M, Figueiredo A, Chen XR, Hulvey J, Stam R, Lamour K, Gijzen M, Tyler BM, Grünwald NJ, Mukhtar MS, Tomé DF, Tör M, Van den Ackerveken G, McDowell J, Daayf F, Fry WE, Lindqvist-Kreuzer H, Meijer HJ, Petre B, Ristaino J, Yoshida K, Birch PR, Govers F. ***Molecular Plant Pathology***. Elements of chapter 1.

**Expression Profiling during Arabidopsis/Downy Mildew Interaction Reveals a Highly-Expressed Effector that Attenuates Salicylic Acid-Triggered Immunity** (2014) Asai S, G Rallapalli, S Piquerez, MC Caillaud, **OJ Furzer**, N Ishaque, L Wirthmueller, G Fabro and J. D.G. Jones. ***PLoS Pathogens***. Elements of chapter 2b.

### PAPERS IN PREPARATION

**Non-host resistance in Arabidopsis accessions to Brassica-infecting *Albugo candida* strains is conferred by three distinct resistance loci** (2014) Cevik V, K Bailey, **O Furzer**, E Kemen, A Robert-Seilaniantz, JDG Jones (*in preparation*). Elements of chapters 2b and 6.

***Albugo laibachii* effector SSP16 is recognized by the Arabidopsis HR-5 *RAL4* Resistance locus** (2014) **Furzer O**, N Ishaque N, Gardiner A, M McMullan, K Bailey, V Cevik, E Kemen, JDG Jones (*in preparation*). Elements of all chapters.

**Comparison of genome sequences of 6 *Albugo candida* strains reveals a rapidly diversifying CCG class of effector candidate gene** (2014) Bailey K, V Cevik, A Gardiner, M McMullan, **O Furzer**, E Kemen, JDG Jones (*in preparation*) Elements of chapter 2b.

## MAJOR ABBREVIATIONS

<i>Ac</i>	<i>Albugo candida</i>
<i>Al</i>	<i>Albugo laibachii</i>
Arabidopsis; <i>At</i>	<i>Arabidopsis thaliana</i>
Avr	Avirulence (gene or protein)
BAK1	Bri1-associated receptor kinase 1
BEB	Bayes-empirical-bayes
BIK1	<i>Botrytis</i> -induced kinase 1
BLAST	Basic local alignment search tool
C-terminus	Carboxyl-terminus
CC	Coiled-coil
cDNA	Complementary DNA
CDPK	Calcium dependent protein kinase
CDS	Coding sequence
CERK1	Chitin Elicitor Receptor Kinase 1
DNA	Deoxyribonucleic acid
EDS1	Enhanced disease susceptibility 1
ETI	Effector triggered immunity
EFR	Ef-Tu Receptor
FLS2	Flagellin sensing 2
gDNA	Genomic DNA
GUS	$\beta$ -glucuronidase
<i>Hpa</i>	<i>Hyaloperonospora arabidopsidis</i>
LB	Lysogeny broth
LRR	Leucine Rich Repeat
MAPK	Mitogen activated protein kinase
N-terminus	Amino-terminus
NB	Nucleotide-binding
NDR1	Nonrace-specific disease resistance 1
PAD4	Phytoalexin deficient 4
PAMP	Pathogen associated molecular pattern
PCR	Polymerase Chain Reaction
<i>Pi</i>	<i>Phytophthora infestans</i>
PRR	Pattern recognition receptor
<i>Ps</i>	<i>Pseudomonas syringae</i>
PTI	PAMP-triggered immunity
R	Resistance
RAC	Resistance to <i>Albugo candida</i>
RAL	Resistance to <i>Albugo laibachii</i>
RLK	Receptor like kinase
RNA	Ribonucleic acid
RPP	Resistance to <i>Peronospora parasitica</i> ( <i>Hpa</i> )
SAG101	Senescence Associated Gene 101
SID2	Salicylic acid biosynthesis deficient 2
SM	Sequencing marker
SSLP	Simple sequence length polymorphism
SSP	Short secreted protein
T3	Type 3
T3SS	Type 3 secretion system
TIR	Toll/ interleukin receptor
VCF	Variant call format
WRR	White rust resistance

## CHAPTER 1: GENERAL INTRODUCTION

Plants convert sunlight and CO<sub>2</sub> into the primary source of carbohydrate for life on the earth. They encounter biotic stress, through parasites and pests, yet plant populations in the wild are rarely wiped out. This is in part because plants have evolved a surveillance system capable of detecting pathogens and mechanisms by which detection can activate immunity. The evolution of these host resistance mechanisms is in turn countered by the evolution of pathogen mechanisms to overcome them.

In this thesis I will discuss the mechanisms by which the biotrophic pathogen *Albugo laibachii* (*Al*) parasitizes and is detected by its host *Arabidopsis thaliana* (*At*). To provide context for this discussion I will review the relevant literature on the mechanisms of plant immunity and microbial pathogenesis.

### 1.1 PATTERN-TRIGGERED IMMUNITY, THE FIRST LINE OF DEFENSE

Pattern Triggered Immunity (PTI) is the first line of active surveillance and defense against pathogenic microbes of plants. Pathogen-associated molecular patterns (PAMPs) are considered to be products produced by microbes that are essential to their competitiveness. PAMPs are recognized by both animal and plant cells. Several examples of plant-recognised PAMPs have been identified in plant pathogenic microbes. From bacteria an epitope of the flagellum, flg22, and an epitope of the elongation factor tu, elf18, are recognized by membrane bound receptors. Collectively, the receptors that recognize PAMPs are known as pattern recognition receptors (PRRs) (Jones and Dangl, 2006). Flg22 and elf18 are both recognized by plasma membrane-bound receptor-like kinases with extracellular leucine rich repeat domains (LRR-RLKs): Flagellin sensing 2 (FLS2) (Felix et al., 1999; Gomez-Gomez and Boller, 2000) and Ef-Tu receptor (EFR) (Zipfel et al., 2006), respectively. Upon ligand detection, these LRR-RLKs form a complex in the plasma membrane with BRI1-associated receptor kinase 1 (BAK1) to initiate signalling (Sun et al., 2013; Halter et al., 2014). On the other hand, the best-known eukaryotic PAMP, chitin, is recognized through the chitin octamer-induced homo-dimerization at the plasma membrane of chitin elicitor receptor kinase 1 (CERK1) proteins (Liu et al., 2012).

CERK1 contains an extracellular LysM domain that mediates the interaction with chitin (Liu et al., 2012), and an intracellular kinase domain. Additionally, in *Oryza sativa* (rice), parallel systems exist: OsCERK1 and another LysM protein, chitin-elicitor binding protein (CEBiP), recognise chitin through homo- and hetero-dimerisation (Hayafune et al., 2014) and *Xanthomonas resistance 21* (*Xa21*), though its ligand is not known, encodes an LRR-RLK (Song et al., 1995). So far in plant systems only membrane bound PRRs have been identified, in contrast to animal systems, where both membrane localised and intra-cellular PRRs have been identified (O'Neill, Golenbock and Bowie, 2013; Franchi et al., 2009).

So far the mechanism of activation has been well defined in FLS2 and CERK1. In FLS2, the flg22 peptide interacts with both FLS2 and BAK1 ectodomains (Sun et al., 2013). The interaction of these 3 components results in the phosphorylation of both FLS2 and BAK1 (Schwessinger et al., 2011). Botrytis Induced Kinase 1 (BIK1) also interacts with and is phosphorylated following the recognition of flg22 and seems to be important for PTI signal transduction (Lu et al., 2010), and was indeed shown to directly phosphorylate the NADPH oxidase RbohD and positively regulate its activity to produce reactive oxygen species (Kadota et al., 2014).

The events directly downstream of chitin-induced CERK1 homo-dimerization are less clear, but recognition does trigger the phosphorylation of the receptor proteins (Liu et al., 2012).

Downstream of PAMP perception, at least two separate MAPK (mitogen activated protein kinase) cascades are activated which mediate immune responses via the phosphorylation of WRKY transcription factors (Eulgem and Somssich, 2007). PAMP perception also triggers a calcium ( $\text{Ca}^{2+}$ ) burst that activates CDPKs (Calcium dependent protein kinases) to also activate defence-related transcription factors (Boudsocq et al., 2010) and a NADPH oxidase, *AtRBOHD* to produce reactive oxygen species (Dubiella et al., 2013). The activation of a subset of WRKY transcription factors (following flg22 addition) results in the differential transcription of ~1100 genes in *At* (Asai et al. 2002). PAMP perception also leads to callose deposition (strengthening the cell wall and limiting availability of water and nutrients) (Gómez-

Gómez et al., 1999), reactive- oxygen species (ROS) (Nurnberger et al., 2004) and nitric oxide (NO) (Asai et al., 2008) generation.

The sum of these activities generally have the phenotype of quantitatively restricting the growth of either biotrophic or hemibiotrophic pathogens (Roux et al., 2011). However the inter-family transfer of EFR from *At* to *Solanum lycopersicum* (tomato) gave a high level immunity against *Ralstonia solanacearum* (Lacombe et al., 2010). For further information, Macho and Zipfel (2014) provide an in depth review of the recognition and signaling aspects of PTI in plants.

## 1.2 PATHOGEN EVASION AND SUPPRESSION OF PTI

Once they have breached the epidermis, PTI represents the first hurdle for pathogens to overcome. Perhaps the simplest way to overcome PTI is to avoid recognition altogether. McCann et al (2012) showed, by examining several bacterial genomes, that PAMP-encoding genes are under adaptive selection. Indeed it has been shown that adaptive variation in flagellin has led to an allele that is not recognized by the known *At* FLS2 receptor, without any cost to motility (Clarke et al., 2013).

In addition, plant pathogens have evolved many mechanisms that suppress PTI. Specialized pathogens have evolved “effectors”. Effectors are generally proteins that suppress a host’s defenses or reprogram its biochemistry to favor the pathogen (Hogenhout et al., 2009). Several fungal pathogens have attenuate PTI activation in the apoplast. *Cladosporium fulvum*, *Magnaporthe oryzae* and others have evolved secreted proteins with LysM domains capable of binding chitin oligomers in the apoplast, which are able to prevent the chitin based activation of PTI (Sánchez-Vallet et al., 2013; Mentlak et al., 2012).

Pathogens have also evolved ways to inhibit PTI inside plant cells. Gram-negative bacterial pathogens produce a type-3 secretion system (T3SS) allowing them to secrete effector proteins directly to the plant cell cytoplasm (Galan and Collmer, 1999; Alfano and Collmer, 2004). Oomycetes and fungi produce structures in order to create high-surface area contacts with their host cells to allow effector uptake and nutrient acquisition. These include haustoria (rust fungi and oomycetes)

(Kemen et al., 2005; Petre and Kamoun, 2014) and invasive hyphae with so-called biotrophic interfacial complex (BIC) in ascomycete fungi such as *Magnaporthe oryzae* (Giraldo et al., 2013).

There are a multitude of effectors that interfere with PTI and other processes; I will highlight a few examples.

AvrPto is an effector of *Pseudomonas syringae* (*Ps*), secreted via the T3SS. AvrPto is a short hydrophilic protein that interferes with FLS2 association to disrupt PTI by preventing BIK1 phosphorylation (Xiang et al., 2011). It also associates with EFR, ostensibly for the same reason (Zong et al., 2008). *Ps* strains lacking *AvrPto* are less virulent on wild type *At* (Zong et al., 2008).

Another *Ps* T3-secreted effector, HopAO1, can directly target EFR. It is a tyrosine phosphatase and reduces the level of phosphorylation of EFR following elf18 perception, thus directly inhibiting the activation of PTI (Macho et al., 2014).

An RXLR effector (these will be defined later), secreted by *Phytophthora infestans* (*Pi*), called PexRD2 can also inhibit PTI. It works by interacting with a MAPK (MAPKKKε) that is induced by *Pi* culture-filtrate treatment, and perturbs this signaling pathway to make plants more susceptible (King et al., 2014). Pathogen effectors can also act against PTI by targeting components of plant signaling pathways and activating those that inhibit defense responses. For example the *Ps* T3-secreted effector HopZ1a promotes the degradation of JAZ transcription factors by acetylating them (Jiang et al., 2013). JAZ proteins are negative regulators of jasmonate (JA) induced gene-expression, which is associated with defence to necrotrophic pathogens and interferes with defences against biotroph and hemibiotrophic pathogens (Glazebrook, 2005). The *Hyaloperonospora arabidopsidis* (*Hpa*) RXLR-like effector HaRxL44 interacts with Mediator subunit 19a (MED19a), resulting in its degradation in a proteasome-dependent manner (Caillaud et al., 2013). This interference with the mediator seems to shift transcription in favor of JA-induced gene-expression.

In addition to secreting effector proteins to suppress PTI, some pathogens can also produce mimics of plant signaling molecules. The most notable example is coronatine, which is a *Ps* produced JA-isoleucine mimic (Geng et al., 2014). Other pathogens can hijack various hormone-signaling pathways (Robert-Seilaniantz, Grant and Jones, 2011).

### 1.3 EFFECTOR TRIGGERED IMMUNITY BY RESISTANCE PROTEINS

To counter effectors, plants have evolved *Resistance* genes (*R* genes). *R* gene-encoded proteins recognize pathogen effectors and/or their activity, and induce a stronger defense response known as effector triggered immunity (ETI) (Jones and Dangl, 2006). Genes encoding recognized effectors are referred to as avirulence genes (*AVR* genes), a term dating to before the molecular identification of the *AVR* gene products as secreted effector proteins. The ETI elicited by avirulence protein recognition is stronger than PTI and results in MAPK signaling, transcriptional reprogramming, release of salicylic acid (SA) and the production of ROS and NO (Jones and Dangl, 2006). In many cases this leads to programmed cell death, also known as the hypersensitive response (HR), but removing this component doesn't necessarily affect the capacity of ETI to halt pathogen growth (Coll et al., 2011).

This interaction between *R* genes and *AVR* genes was initially known as the gene-for-gene relationship (Flor, 1971), because it was possible to define single segregating loci that conferred resistance and avirulence in plants (e.g. flax) and pathogens (e.g. flax rust), respectively. Known *R* genes almost always encode NB-LRR (Nucleotide binding site, Leucine-rich repeat) proteins (Eitas and Dangl, 2010). Arabidopsis NB-LRRs are split into two classes depending on their N-terminal domain: the CC (Coiled-coil) or TIR (Toll/interleukin receptor/Resistance protein) – NB-LRRs (Meyers et al., 2003). In the reference *At* accession, Col-0, there are 53 CC-NB-LRRs and 90 TIR-NB-LRRs. There are also around 42 partial NB-LRRs, lacking one or more domains (Meyers et al., 2003). Some NB-LRRs act as pairs, for example in the well-studied *RPS4/RRS1* system (Williams et al., 2014), but most appear to function genetically independent of other NB-LRRs. Downstream signaling of NB-LRRs has been challenging to characterize (Eitas and Dangl, 2010). In *At*, some CC-NB-LRRs are dependent on NDR1 (Non-race specific disease resistance-1), an

integrin like protein involved in cell wall- plasma membrane adhesion (Knepper, Sovory and Day, 2011). TIR-NB-LRR based defense activation is dependent on the EDS1 (Enhanced Disease Susceptibility 1), PAD4 (Phytoalexin Deficient 4) and SAG101 (Senescence Associated Gene 101) which are structurally similar to lipase proteins. These proteins can homo- and heterodimerise and potentially form a ternary complex (Feys et al., 2005). They may also interact directly with TIR-NB-LRRs (Heidrich et al., 2011). The mutation *snc1-1* in the linker of the NB and LRR domains of a TIR-NB-LRR results in constitutive activation of defense. A mutant screen searching for suppressors of this phenotype, led to the discovery of *At* protein MOS7, which localizes to the nuclear envelope, and is required for NB-LRR mediated immunity and the accumulation of other defense related proteins (including EDS1) in the nucleus (Cheng et al., 2009). Interestingly, a double mutant of *sid2* (Salicylic acid biosynthesis deficient) and *eds1* is impaired in signaling for two of the CC-NB-LRRs that are not impaired by *ndr1*, *sid2* or *eds1* single mutants suggesting that the pathways are redundant in this case (Venugopal et al., 2009).

A third type of R proteins has been defined in *Solanum* species. These are structurally similar to the LRR-RLKs that are generally associated with PTI in that they encode a transmembrane domain and LRR, however they lack protein kinase domains (Wang et al., 2010). Their function seems to be to monitor the apoplast for effectors or their activities. Recent data have indicated that these proteins associate with a kinase called SOBIR1 in order to initiate defense signaling (Liebrand et al., 2013).

In terms of the physical recognition of effectors, two general mechanisms have emerged. The first is the direct interaction of R proteins with their cognate avirulence proteins. Examples include the direct interaction of the *Melampsora lini* AVR protein AvrL567 and the flax TIR-NB-LRR L6 (Dodds et al., 2006) and the recognition of the AVR protein ATR1 from *Hpa* by the TIR-NB-LRR R protein RPP1, which occurs in the LRR domain (Krasileva, Dahlbeck and Staskawicz, 2010). Due to the parallels with animal systems where TIR-NB-LRR proteins function as PRRs, it was hypothesized that effector recognition would occur in the LRR 'sensor' domain,

however in some cases the TIR- or CC- domain of NB-LRRs can be involved in direct interaction with the AVR ligand (Burch-Smith et al., 2007; Chen et al., 2012).

To explain the observation of indirect recognition of effectors by R proteins, a second model, the 'guard model' emerged (van der Biezen and Jones, 1998). In this model it is proposed that R proteins can guard important cellular targets to detect when they are targeted by pathogen effectors. It is best exemplified by RPM1 interacting protein 4 (RIN4) (Grant et al., 1995; Mackey et al., 2002). RIN4 is a protein involved in PTI that plays a role in the regulation of stomatal closure in response to PAMPs (Liu et al., 2009). RIN4 is the target of the activities of at least 4 effectors and guarded by 2 R proteins. Two *Ps* effectors, AvrB and AvrRpm1 promote the phosphorylation of RIN4 via an endogenous kinase called RIPK1 (Liu et al., 2011). This phosphorylation leads to the activation of the R protein RPM1, (Grant et al., Mackey et al., 2002). A further *Ps* effector, AvrRpt2, cleaves RIN4 and triggers ETI via RPS2 (Kim et al., 2005). Both RPM1 and RPS2 function in an NDR1 dependent manner, and NDR1 associates with RIN4 (Day, Dahlbeck and Staskawicz, 2006). In addition, another effector, HopF2 *Pto*, interacts with RIN4 to enhance susceptibility to *Ps* but doesn't trigger an immune response (Wilton et al., 2010). The guard model elegantly explains how a plant with a relatively small repertoire of NB-LRR encoding genes can defend itself against a broad range of different pathogens (Dangl and Jones, 2001).

A further conceptual development of this model is the decoy theory proposed by van der Hoorn and Kamoun (2008). Probably the most clear cut example of the decoy model is the *Bs3* R gene in pepper. *Xanthomonas* species plant pathogens have evolved effectors capable of binding to host DNA and promoting the transcription of genes that favor their life-styles (Boch and Bonas, 2010). In order to recognize one of these effectors, the *Bs3* R gene is effectively a promoter that the effector (AvrBs3) can bind, coupled to a gene encoding a flavin dependent mono-oxygenase (FMO1), that when transcriptionally activated triggers a HR (Romer et al., 2007). The plant therefore sets a trap, or 'decoy', for the pathogen to target. There are other more ambiguous cases for the decoy model, for example the TIR-NB-LRR encoding gene *RRS1* that encodes a WRKY DNA binding domain at its C-terminus. A

*Ralstonia solanacearum* T3 effector PopP2 interacts with this C-terminal domain to trigger resistance (Deslandes et al., 2003). *RRS1* is encoded in the genome in a head-to-head configuration with *RPS4*, which encodes another TIR-NB-LRR required for *RRS1* activation. This raises the question of whether *RRS1* acquired the WRKY motif as a decoy against interference with other WRKY transcription factors that are involved in defense (Eulgem and Somssich, 2007) (with *RPS4* as a guard of this decoy). Alternatively, it may be the case that *RPS4/RRS1* represents a protein complex that has the capability to recognize, signal and activate defense through the binding of its WRKY domain to DNA (Narusaka et al., 2009)

The recognition of an effector causes a NB-LRR protein to switch from an inactive to an active signaling state. In many cases the ATPase domain of the NB-LRR is required for its function (Takken and Govere, 2012). In the off state, the protein NB domain is proposed to be bound to ADP. The recognition of an effector is then hypothesized to bring about a conformational change resulting in the availability of the NB domain and the exchange of the ADP for ATP. This could bring the protein into an 'open' signaling conformation. The ATP is hydrolyzed and the R protein is returned to the ADP bound state (Takken and Govere, 2012). The oligomerisation of multiple NB-LRRs via the homo-dimerisation of the CC or TIR domains may also be important for the activation of signaling (Maekawa et al., 2011; Williams et al., 2014).

#### 1.4 POPULATION GENETICS OF GENE-FOR-GENE INTERACTIONS

Even before the discovery of the first *R* and *AVR* genes in the early 1990s, population biologists have attempted to build models of their population dynamics (Leonard, 1994). It is clear that the evolution of an *R* gene that is capable of conferring recognition and complete resistance to a pathogen strain carrying a specific effector will bring a strong selective pressure against the effector, as the *R* gene proliferates within the plant population. On the other hand, useful effectors will have fitness benefits, leading to their proliferation and maintenance within the pathogen population.

It is useful to define the various terms used when discussing genetic selection. Adaptive or positive selection, though a possible tautology, refers specifically to an enrichment of polymorphisms that confer changes that encode amino-acid changes (McDonald and Kreitman, 1991). Purifying or negative selection refers to the opposite, when a higher proportion of accumulated polymorphisms do not confer any amino acid changes (Terauchi and Yoshida, 2010). Balancing selection refers to genes where a limited number of diverse alleles appear to be maintained within a population. Balancing selection is an indicator of the action of negative frequency dependent selection (NFDS), whereby the frequency of the occurrence of an allele will determine its relative fitness benefit or cost (Brown and Tellier, 2011). Linkage disequilibrium, or the non-random distribution of allele frequency in natural populations, is also a signature of balancing selection.

There are two models proposed for population level interactions between *R*-genes, *Avr* genes and the targets of *Avr* gene products: the “arms race” model and the “trench warfare” model (Stahl et al., 1999, Stukenbrock and McDonald, 2009; Terauchi and Yoshida, 2010). The arms race model, proposes that novel adaptive mutation in any of the interactors (i.e. a gain of recognition mutation in an *R* gene) causes the corresponding interactor to be swept from the population and be replaced by a new allele that can evade the new recognition (Dawkins and Krebs, 1979; Stahl et al., 1999; Anderson et al., 2010). This model is problematic in that it seems to assume there will be no fitness cost to evolve a certain adaptive mutation.

The trench warfare model is similar except it takes into account that adaptive changes to gain or loss of recognition may have a cost in terms of fitness. This leads to NFDS (Brown and Tellier, 2011). In this scenario there remains a selective advantage for those individuals who possess the *AVR* allele so long as they do not encounter their cognate *R*-gene, and in the case of non-functional *R*-gene alleles, the avirulent pathogen. This model incorporates previously made predictions that there will be cyclical oscillation of allele frequency within the population and the maintenance of higher than average diversity and number of alleles within the cognate loci (Frank, 1992). In both models, proteins involved in host-parasite interaction phenotype are expected to show signatures of adaptive selection. At the

DNA sequence level, genes undergoing some form of selection should display deviation from the neutral theory of molecular evolution (Kimura, 1968). Genes undergoing NFDS/balancing selection or involved in population bottlenecks (eg *Avr* or *R* genes) will have high genetic diversity but a lower number of unique alleles than projected based on the neutral theory due to the balancing effect of NFDS and selection against alleles intermediate between recognition and evasion. The likelihood is that in reality a mix of “trench warfare” and “arms race” occurs, depending on the specific fitness costs and circumstances such as population size and rate of dispersal (Frank, 1992; Holub, 2001; Tellier and Brown, 2011).

Data from the sequencing of large populations of *At* genomes suggest that there is elevated diversity in the NB-LRR complement, supporting these theories (Cao et al., 2011; Bakker et al., 2006). However, the difficulty of resolving the full sequences of NB-LRRs using current short-read sequencing means that a true measure of selection has not yet been reliably reported. There is also evidence from the pathogen side. Hall et al (2009) examined 47 sequences of the *Hpa Avr* gene *ATR13* and found 15 different alleles, suggesting a high level of diversity and balancing selection. In a study of genome-wide polymorphism in 8 strains of *Colletotrichum graminicola*, signatures of adaptive and balancing selection were found in the predicted effector complement, including the 5' regulatory sequences of these genes (Rech et al., 2014). There is evidence of a useful effector undergoing a transfer via interspecific hybridization and then becoming fixed within the pathogen population (McDonald et al., 2013). Recent evidence suggests that *At* genes under balancing selection are more likely to interact with pathogen effectors (Weßling et al., 2014). Stukenbrock and McDonald (2009) and Terauchi and Yoshida (2010) provide many examples of the different types of selection operating on genes involved in host- pathogen interactions.

## 1.6 OOMYCETES, EFFECTORS AND GENOMICS

The oomycetes (or oomycota) are a diverse group of eukaryotic microorganisms. They have colonized a wide variety of host species as parasites: mammals (including humans) (Botton et al., 2011), fish (Ke et al., 2009) true fungi (Le Floch et al., 2003). ~60% of identified oomycete species, however, colonize plants (Thines and

Kamoun, 2010). Many poorly studied oomycetes also exist in nature as saprotrophs, feeding from dead and decaying matter (Thines and Kamoun, 2009).

Oomycetes have evolved from algae in a marine environment: many basal lineages are parasites of marine organisms (Thines and Kamoun, 2009). They are part of the supergroup Chromalveolata, within which they belong to the phylum heterokonta, which includes mostly diatoms and red/brown/golden algae (Adl et al., 2005). The oomycota are a class within the heterokonta, containing six orders: the lagenidiales which includes species that are pathogens of mosquitoes and dogs (Grooters, 2003), the leptomitales, the peronosporales, which includes some of the major plant pathogenic genera, including *Phytophthora* and *Hyaloperonospora*, the rhipidiales and the saprolegniales, which includes the fish pathogenic genera *Saprolegnia* (van West et al., 2010) and finally the albuginales, which includes *Albugo* spp., the white rust pathogens (Thines and Spring, 2005).

Pathogenesis of plants has evolved three times among the oomycetes; in the saprolegniales *Aphanomyces euteiches* is a legume pathogen of increasing prominence (Gaulin et al., 2008), in the peronosporales there are the Peronosporaceae/ Pythiaceae and in the albuginales three plant pathogenic genera (including *Albugo* spp.) (Thines and Kamoun, 2009).

The genus *Phytophthora* is the most intensively studied genus of oomycetes. *Phytophthora infestans* (*Pi*) is the most economically important oomycete pathogen (infecting numerous Solanaceaeous plants including potatoes and tomatoes) (Kamoun et al., 2014). Other *Phytophthora* species *P. ramorum* and *P. sojae* cause important diseases of many woody tree species (Mascheretti et al., 2009) and soybean (Tyler et al., 2006), respectively. *Phytophthora* spp. are hemibiotrophs, killing their hosts during the later stages of infection. Despite growing on *Solanaceae*, *Pi* has become a model pathogen for effector studies of eukaryotic plant- pathogenic microorganisms.

13 CC-NB-LRR encoding genes have been identified as *R* genes against *Pi* from *Solanum* species (Rodewald and Trognitz, 2013). Corresponding to these *R* genes,

several AVR genes have also been cloned. These include *Avr3a* (Armstrong et al., 2005), *AVR2* (Gilroy et al., 2011), *Avr-blb1* and *Avr-blb2* (Oh et al., 2009).

*Arabidopsis* Downy Mildew or *Hpa* is a model oomycete plant pathogen (Coates et al., 2010). *Hpa* grows on *Arabidopsis* in the wild, as an obligate biotroph (Holub, Beynon and Crute, 1994). *Hpa* is a member of the peronosporalean lineage, like *Pi*. However, *Hpa* has become adapted to the lifestyle of an obligate biotroph, losing several critical metabolic enzymes making it dependent on its host (Baxter et al., 2010).

In contrast to *Pi*, *R* genes of both the CC- and TIR-NB-LRR classes have been identified in *At* against *Hpa* (Coates et al., 2010). Again, multiple corresponding AVR genes have been identified. These are *ATR13* (Allen et al., 2004), *ATR1* (Rehmany et al., 2005) and *ATR5* (Bailey et al., 2011). The key to the discovery of these *Avr* genes was the development of a protocol to make crosses in *Hpa*, which led to their genetic mapping.

Excepting *ATR5*, all of the AVR genes cloned from either *Phytophthora* species or *Hpa* contain an RXLR motif (Arginine, any amino acid, Leucine, Arginine) in the region following secretion signal cleavage. This was first noted by Rehmany et al, (2005), following the discoveries of *ATR13*, *ATR1* and *Avr3a*. Bhattacharjee et al, (2006) noticed a similar motif in the effector proteins of the malaria parasite *Plasmodium falciparum* (known as PEXEL effectors). They showed that the RXLR motif could translocate GFP from *Plasmodium* into human erythrocyte cells, and suggested that the motif is either an ancient and conserved effector translocation mechanism, or an example of convergent evolution. Whisson et al (2007) showed that by mutating the *Avr3a* RXLR motif (RXLR –EER) to either alanines or to KMIK-DDK, the protein was no longer translocated into the host cell, nor able to activate resistance via *R3a*, its corresponding cytoplasmic resistance protein. Since that time a number of studies have tried to address the mechanistic process by which these effectors move from the extra-haustorial matrix to the cytoplasm of host cells. The mechanism remains unclear. Petre and Kamoun, (2014) provide an up to date

account of these studies. Note that Tian et al, (2011) identified a variant RXLR (a QXLR motif) effector class in the *Pseudoperonospora cubensis* genome.

Analyses of the *Phytophthora* genomes revealed 563 RXLR effector candidates in *Pi* race T30-4, and 335 and 309 in *P. sojae* and *P. ramorum*, respectively (Haas et al., 2009; Tyler et al., 2006). One of the most striking features of the *Pi* genome is the clear differential between the core genes and the effector complement. *Pi* core (conserved) genes are located in stable, gene dense regions. In contrast the RXLR effector-encoding genes reside in gene sparse, repeat rich regions (Haas et al., 2009). Raffaele et al (2010a) reported sequence data for several further species, and were able to identify genomic regions under differential selection pressure: “a two speed genome”. Regions rich in RXLR effectors showed an enhanced rate of adaptive selection. *Pi* RXLR encoding genes were also shown to be undergoing strong adaptive selection in their C-terminal “effector domain” encoding region (Win et al., 2007).

Analyses of the *Hpa* genome revealed that it has a complement of at least 134 RXLR effectors (Baxter et al., 2010). Functional analyses of 64 of these effectors revealed that many can enhance virulence of *Ps* when delivered via the T3SS (Fabro et al., 2011).

In addition to hundreds of RXLR effectors, the *Pi* and *Hpa* genomes revealed a class of putative effectors called “crinklers” (CRN) due to several members causing a cell death response in *N. benthamiana* (Torto et al., 2003). CRN effectors have a conserved LXLFLAK motif in the N-terminus and a wide array of different C-termini (Haas et al., 2009). Schornack et al (2010) showed that an N- terminus “LXLFLAK” domain is required and for translocation of CRNs, as well as other chimeric reporter C-termini to the host cell.

## 1.7 ALBUGO SPECIES

White blister rust is a disease of many dicotyledonous plant species, caused by obligate biotrophic parasites. For example, *Albugo candida* (*Ac*) infection of *Brassica juncea* (Indian mustard) can result in significant crop losses in India (Awasthi et al., 2012), Canada (Rimmer et al., 2000) and Australia (Kaur et al., 2008).

The white rusts, order Albuginales, are oomycetes but phylogenetically distant from the Peronosporales and probably represent an independent acquisition of biotrophy (Thines and Spring, 2005; Thines and Kamoun 2010). All *Albugo* species infecting the Brassicaceae were thought to be races of *Ac*, but molecular studies of isolates from various hosts and locations led to the description of specialists, for example *A. laibachii* (*Al*) on *Arabidopsis thaliana* (Figure 1.1; Thines et al., 2009; Thines 2014). *Al* can only grow on *At*, and around 15% of accessions are resistant to the two isolates characterised by Kemen et al (2011). Figure 1.1 shows typical susceptibility and resistant phenotypes. Specific *Ac* races can grow on diverse plant hosts, including Brassicaceae, Cleomaceae and Capparaceae (Thines, 2014).

*Albugo* spp. reproduce asexually via zoosporangia, which release flagellated motile zoospores upon incubation in water. On the surface of a plant leaf, zoospores settle in stomata, and each extends a germ tube into the sub-stomatal chamber (Holub et al., 1995). Coenocytic hyphae then grow intercellularly through the plant. Small globose haustoria penetrate into plant cells (Soylu et al., 2003). When an *Albugo* infection is mature, zoosporangia rupture the plant epidermis with force and enzymatic digestion (Heller and Thines, 2009). This results in characteristic “white blister” pustules. *Albugo* also has a sexual cycle, producing tough oospores that can survive difficult environmental conditions (Petrie, 1975). During systemic infection of Brassicaceae hosts, the inflorescences become misshapen, forming so-called ‘stagheads’. In addition to the white blister phenotype, the observations of Ploch and Thines, (2011) suggested that *Albugo* could be widespread as an asymptomatic endophyte.

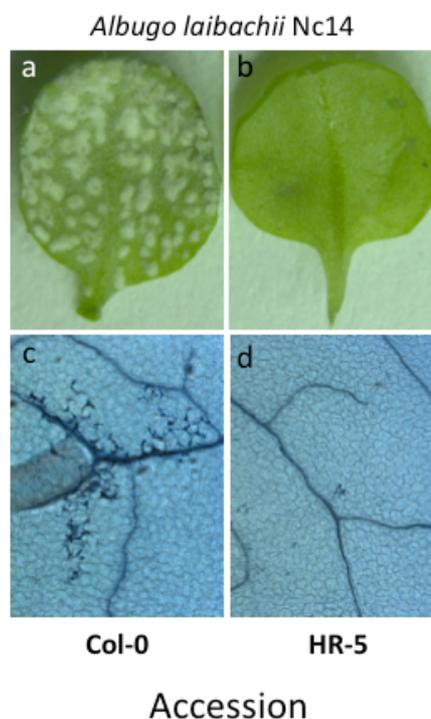
*Albugo* infection has long been associated with “green islands” where infected tissue appears healthy and senescence is delayed. Infection by *Albugo* also greatly enhances susceptibility to co-infections with downy mildews (Bains and Jhooty, 1985; Crute *et al.*, 1994). Cooper et al (2008) investigated the ability of *Al* and *Ac* to suppress host immunity. They showed that *Al* can suppress the “runaway cell death” of *Arabidopsis lsd1* mutants after inoculation with avirulent *Hpa*. Furthermore, when pre-infected with virulent *Al*, resistant *Arabidopsis* accessions were no longer resistant to avirulent *Hpa* isolates, lettuce downy mildew or

powdery mildew. Suppression was also observed on *B. juncea* with *Ac* and Brassica downy mildew (Cooper et al., 2008). These results suggest that *Albugo* is effective at broad suppression of plant immunity, including effector-triggered-immunity activated via several well-defined TIR- and CC-NB-LRR R proteins.

The first step to understanding how *Albugo* spp. impose such susceptibility is to examine their genomes. Links et al (2011) and Kemen et al, (2011) sequenced *Ac* and *Al* genomes, respectively. The genomes are around 40 Mb and compact; about 50% of the assemblies consist of coding sequences. Both genomes show adaptations to obligate biotrophy; they are missing sulfite oxidases, nitrate and nitrite reductases and in the case of *Al* the whole molybdopterin biosynthesis pathway. This implies a long evolved dependence on host metabolism. The *Ac* secretome consists of 929 proteins (without transmembrane domains) compared to 672 in *Al*, perhaps reflecting its wider host range. Within the secretomes there is no enrichment of putative RXLR effectors. Kemen et al, (2011) discovered the CHXC (cysteine, histidine, any amino acid, cysteine) motif at the N-terminus of a class of candidate effectors. The CHXC-containing N-terminus is sufficient to translocate the C-terminus of *Pi* AVR3a into host cells (Kemen et al., 2011). Studies of CHXC effectors using the effector detector vector (EDV) system developed by Sohn et al (2007) provided further evidence that these might be effector proteins, suggesting in some cases a small but significant increase in the virulence of *Ps* strains delivering them via the T3SS (Kemen et al., 2011).

Several *Ac* races can infect some but not all *At* accessions and from crosses between resistant and susceptible accessions, an *R* gene against four *Ac* races, *WRR4* (encoding a TIR-NB-LRR R protein), was identified (Borhan et al., 2008). *WRR4* can also provide resistance to *Ac* when transformed into susceptible cultivars of *B. napus* and *B. juncea* (Borhan et al., 2010). In *At*, *RAC1* (also encoding a TIR-NB-LRR R protein) confers resistance to *Al* (Borhan et al., 2004). The inheritance of avirulence of a *B. juncea* isolate (Ac2V) was studied through a cross between two *A. candida* isolates; this work predicted a single avirulence gene for the incompatibility between Ac2V and *B. rapa* (Adhikari et al., 2003).

There are open questions about *Albugo* from both fundamental and translational perspectives. Thines (2014) speculated that the broad host range of the *Ac* meta-population is maintained through frequent genetic exchange where the host range of individual isolates overlap. Comparing the genomes of multiple isolates from different hosts would test this hypothesis and build up a clear picture of population variation. This would also aid the discovery of new effector candidates through the identification of secreted proteins under strong selective pressure. More extensive phylogenetic and functional analysis of *Albugo* effectors should be carried out, for example the presence of the CHXC effectors inside host cells needs to be confirmed and the translocation mechanism elucidated. It is unclear which *Albugo* effector proteins are recognised by the few known R-proteins. *At* cannot be colonised by most *Ac* isolates. The molecular basis for this resistance could be exploited to introduce durable resistance to Brassica crops.



**Figure 1.1. Susceptible and resistant *Albugo laibachii* infection phenotypes.** *Al* isolate Nc14 was spray inoculated onto *At* accessions Col-0 (panels a and c) and HR-5 (panels b and d). Leaves were taken and photographed at 15 dpi (panels a and b) and Trypan blue stained to visualize pathogen growth (panels c and d).

## 1.8 AIMS OF THIS STUDY

The major aim of my thesis is to build up knowledge about *Al*. So far little is known about the effectors or R proteins that condition susceptibility and resistance in the *Al-At* patho-system. This work leverages natural diversity as a tool to reveal these components. Although Kemen et al, (2011) discovered the CHXC class of candidate effectors, the low number of these, and the weak evidence of their virulence activity suggests that they might not be the only *Al* effectors. Therefore I took a tried and tested approach to identify key effectors; find pathogen *Avr* genes/genes encoding recognized effectors. To do this in a system where genetics isn't practical, I hypothesized that I would be able to generate candidates based on patterns of polymorphism both in terms of association with pathogen phenotype, and by the signatures of strong adaptive and balancing selection that these genes are predicted to be under.

The scope of this work encompasses: surveying natural variation in *Al* host range (in chapter 3), genetic diversity (chapter 4), using this genetic diversity to identify recognized effectors (chapters 4 and 5) and finally developing our understanding of host resistance to *Al* (chapter 6).

## CHAPTER 2A: STANDARD MATERIALS AND METHODS

### 2A.1 MOLECULAR METHODS

#### *ISOLATION OF DNA FOR PCR*

Isolation of plant and *Albugo* DNA for general purposes such as diagnostic sequencing and cloning was achieved using the so-called Shorty method. Briefly, 1 leaf or 1 infected leaf was flash frozen in a 1.5ml Eppendorf tube and crushed using a plastic pestle. The slurry was resuspended in 0.5 µl of Shorty buffer (20% 1M Tris HCl pH 9, 20% 2M LiCl, 5% 0.5M EDTA, 10% SDS, 45% dH<sub>2</sub>O). 0.5 µl Phenol/chloroform isoamyl alcohol was added and mixed briefly by vortex. The tube was spun at 13000 g for 5 minutes and the upper aqueous phase pipetted into a fresh 1.5 ml Eppendorf. Adding 0.5 µl 100% isopropanol and spinning at 13000G for 10 minutes then precipitated the DNA. The pellet was washed with 70% ethanol before resuspension in pH 8 Tris-EDTA (TE) buffer.

#### *RAPID ISOLATION OF ARABIDOPSIS DNA FOR SCREENING*

A rapid DNA extraction technique was used to extract DNA for mapping. Briefly, 0.1cm<sup>2</sup> leaf samples were crushed using tips in PCR tubes with a 10% w/v chelex 100 solution (Bio-Rad). The samples were then boiled for 5 minutes, pelleted, and 2 µl used for PCR.

#### *ISOLATION OF ALBUGO LAIBACHII DNA FOR NEXT GENERATION SEQUENCING*

To isolate pure *Al* DNA, a homemade vacuum device was built for each isolate. These devices were attached to a vacuum generating pump and used to draw conidiospores from the air around gently agitated *Al* infected *At* plants.

Conidiospores were crushed in liquid N<sub>2</sub> and resuspended by inversion in an extraction buffer (50mM Tris pH 8, 200 mM NaCl, 0.2 mM EDTA, 0.5% w/v SDS, 100 mg/ml Proteinase K) in a 1:2 ratio. An equal volume of phenol:chloroform was added and the tube was mixed by inversion. After centrifugation the top aqueous layer was transferred to a fresh tube and 1 volume of chloroform:isoamyl alcohol (24:1) was added. After centrifugation the top aqueous layer was then transferred to a fresh tube and DNA precipitated using sodium acetate and isopropanol. DNA

was washed twice in 70% ethanol, and the dried pellet suspended in TE buffer. RNase One (Promega) was used at this point to remove RNA from the sample, according to the manufacturer's instructions (~2 hours incubation at 37°C). Following RNase treatment, DNA was again precipitated with sodium acetate and isopropanol.

#### *POLYMERASE CHAIN REACTION (PCR)*

PCR for cloning or for diagnostic sequencing purposes was carried out with proofreading Phusion taq (NEB). 25 µL reactions contained: 1 X Phusion HF buffer, 0.2 mM dNTPs, 0.002 U Phusion polymerase (NEB), and 0.5 µM of each primer. PCR cycles were optimised for primers and the projected length of amplicons. Generally, an annealing temperature of 58°C and extension of 30 sec per kb was used. PCR was performed in a thermal cycler (Peltier Thermal Cycler 225, MJ Research).

PCRs for mapping and colony PCR were performed with NEB standard Taq polymerase and buffer. Colony PCR was performed as above except that the DNA template was substituted with bacterial cells suspended in 50 µl water. PCRs for mapping were typically carried out in 10 µl reaction volume.

PCR of amplicons for USER cloning were performed using PfuTurbo C<sub>x</sub> Hotstart DNA polymerase (compatible with amplification of Uracils) (Stratagene).

#### *GEL ELECTROPHORESIS AND GEL EXTRACTION*

To estimate PCR product or genomic DNA prep size and quantity, gel electrophoresis was used. Typically, gels of 1% agarose were made in 1 % TAE buffer and 0.5 µg/ml ethidium bromide. Gels were run at 100-150V and were visualized using UV light on a Geldoc. Gel extractions were performed using the Machery-Nagel PCR clean up kit, following the manufacturer's instructions.

#### *DNA QUANTIFICATION*

Rough DNA quantification was performed with a Nanodrop (Thermo scientific). To measure accurately the DNA concentration PicoGreen dye (Invitrogen) was used. Briefly, PicoGreen dye was added to DNA samples of known and unknown concentration. The 520 nm fluorescence of these samples was measured after

excitation at 480 nm, using a Varioskan Flash plate reader (Thermo Scientific). The readings for samples of known concentration were used to produce standard curves, which were used to estimate DNA concentrations of the unknown samples.

#### *ISOLATION OF ALBUGO RNA*

Infected leaf tissue was ground in liquid N<sub>2</sub> and the powder transferred to an Eppendorf tube (pre-cooled with dry ice). 1 ml Tri Reagent (Sigma Aldrich) was added and the solution incubated for 10 minutes at room temperature. The tube was centrifuged for 20 minutes at 4°C at 12000g. The supernatant was then transferred to a fresh tube and 1 volume of isopropanol was added. After a brief incubation on ice the tube was centrifuged for 15 minutes at 4°C at 12000G. The pellet was then washed with 70% ethanol and resuspended in RNase free water.

#### *RAPID AMPLIFICATION OF CDNA ENDS-PCR*

RACE-PCR was carried out as recommended by the manufacturer of the GeneRacer kit (Invitrogen). Specific 5' and 3' nested primers were designed to facilitate this process (primers listed in appendices).

#### *GATEWAY CLONING*

Initially genes were cloned into the Gateway entry vector *pENTR-D-TOPO*, according to the manufacturer's instructions (Invitrogen). Sequences were then shuttled into Gateway compatible destination vectors using the LR clonase II mix, according to the manufacturer's instructions (Invitrogen).

#### *GOLDENGATE CLONING*

GoldenGate cloning is a system that allows the assembly of DNA modules based on the custom 4bp overhangs generated by restriction of DNA by type II endonucleases (BsaI and BbsI) and the specific ligations that this can entail (Engler et al., 2009). On the negative side, in order to use GoldenGate cloning one must remove endogenous BsaI and BbsI sites from the gene of interest. NEBcutter ([tools.neb.com/NEBcutter2](http://tools.neb.com/NEBcutter2)) was used to identify such sites and primers designed to induce synonymous mutations to remove them. This is a process known as domestication (primers used for this purpose are marked as such in the appendices). GoldenGate reactions can be carried out as a simultaneous digestion-ligation

reaction (diglig). Digligs were done in the following conditions: the PCR products (gel-purified) and destination vector were mixed at equal molar amount and added in a mix of 0.2 µl BSA, 1x T4 DNA ligase buffer, 1 µl Bsal/BbsI and 1 µl T4 DNA ligase (enzymes NEB). Digligs were carried out in a thermocycler with the following steps: 25 x (37 °C for 30 s, 37 °C for 3 min, 16 °C for 4 min), 50 °C for 5 min, 80 °C for 5 min. 2-3 µl of the reaction was used directly to transform electro-competent *E. coli* cells.

#### *USER CLONING*

USER-compatible *pICH86966* was pre-digested with *PacI* and *Nt.BbvCI*. USER overhang oligos were used to amplify products for USER cloning. USER ligation was carried out by mixing direct PCR product and vector in a 1:5 molar ratio and mixing with 0.75U USER mix (NEB). The mix was incubated for 15 minutes at 37 °C followed by 15 minutes at 25 °C and 5 µl used directly for transformation of chemically competent *E. coli*.

#### *QUANTITATIVE PCR (QPCR)*

qPCR was carried out using 20 µl reactions including 10 µl of SYBR green (Sigma Aldrich), and 10 µl of a mix of DNA and primers. Reactions were carried out in a CFX96 Real-Time System C1000 thermal cycler (Biorad). An annealing temperature of 55 °C was used. Quantitative data were analysed in MS Excel.

#### *PLASMID EXTRACTION*

*E. coli* cultures were grown for 12-14 hours at 37 °C in a shaking incubator (200 rpm). The plasmids were mini-prepped using the NucleoSpin® Plasmid kit (Macherey-Nagel) following manufacturer's instructions. Table 2a.1 shows the various plasmids used.

#### *SEQUENCING OF PCR PRODUCTS AND PLASMIDS*

In order to verify clones and PCR sequences I checked them by Sanger sequencing. This was carried out by sending a mix of DNA and primer to the GATC biotechnology company ([www.gatc-biotech.com](http://www.gatc-biotech.com)). Sequencing analysis was conducted using the LASERGENE suite (DNASStar).

#### *NEXT-GENERATION SEQUENCING*

The sequencing of the 5 *Al* isolates was carried out using the in-house Illumina GAIIx sequencer. Libraries were prepared according to the Illumina TruSeq kit specifications.

#### *LIST OF OLIGONUCLEOTIDE PRIMERS*

Primers were synthesized by Sigma Aldrich (UK) See the appendices for a list of primers used in this thesis.

<b>Name</b>	<b>Purpose</b>	<b>Selectable marker</b>
<i>pENTR-D-TOPO</i>	Entry vector	Kanamycin
<i>pK2GW7</i>	35S binary vector	Spectinomycin/Kanamycin
<i>pEDV6</i>	Effector detector vector	Gentamycin
<i>pICH86988</i>	35S binary vector	Kanamycin, blue/white
<i>pICH86966_user</i>	own promoter binary vector	Kanamycin, blue/white
<i>pRK2013</i>	<i>Pseudomonas</i> mating helper plasmid	Kanamycin

**Table 2a.1. List of plasmids used.**

## 2A.2 MICROBIAL METHODS

### *ISOLATION, PURIFICATION, PROPAGATION AND SCREENING OF ALBUGO LAIBACHII*

Field isolates of *Albugo* were collected using dry folded paper. These samples were immersed in 20 ml chilled dH<sub>2</sub>O to release conidiospores. After 1 hour the suspension was sprayed onto four *At* Col-THO plants, which were kept in the dark at 4°C overnight, then transferred to a plant growth chamber. To purify isolates, three generations of single pustule infections were carried out. Single pustule leaf infections were sampled and immersed in 20 ml chilled dH<sub>2</sub>O. Using a haemocytometer, conidiospore concentration was adjusted to 1x10<sup>3</sup> / ml, and this suspension was sprayed onto Col-THO plants. Col-THO is a Col-5 line transformed with *RPW8* making it resistant to co-infections with powdery mildews (Eric Kemen, personal communication). Sprayed plants were again kept in the dark at 4°C overnight, and transferred to a plant growth chamber. After three generations of single spore infections, the isolate can be considered pure (Kemen et al., 2011). Using the isolate differentiating molecular marker discussed in chapter 3, the isolates were checked regularly for purity. To screen *At* accessions for

resistance/susceptibility to different *AI* isolates, four plants of each accession were sprayed with a suspension of spore concentration  $1 \times 10^5$  / ml and checked periodically for 3 weeks for signs of infection. This was repeated at least twice for each isolate/accession combination.

#### *TRANSFORMATION AND GROWTH OF ESCHERICHIA COLI*

Plasmids were transformed into electro-competent *E. coli* DH10B cells using  $\sim 5 \times 10^8$  cfu cells in 0.1 cm cuvettes and a Biorad Micropulser on the standard *E. coli* setting. Following electroporation, strains were cultured in liquid lysogeny broth (LB) medium for 1 hour at 37°C. Subsequently, the culture was spread onto L plates with the appropriate antibiotic selection.

Antibiotics were added to liquid LB or L media or solid L plates as required, kanamycin (50 µg / ml), ampicillin (100 µg / ml), spectinomycin (50 µg / ml), gentamycin (10 µg / ml) and chloramphenicol (25 µg / ml).

TOP10 competent cells (Invitrogen) were used for the transformation of USER plasmids (not compatible with electroporation).

#### *TRANSFORMATION OF AGROBACTERIUM TUMEFACIENS*

Plasmids were transformed into electro-competent *A. tumefaciens* GV3101 cells using  $\sim 5 \times 10^6$  cfu cells in 0.1 cm cuvettes and a Biorad Micropulser on the standard *A. tumefaciens* setting. Following electroporation, strains were cultured in liquid LB medium for 2 hours at 28°C. Subsequently, the culture was spread onto L plates with the appropriate antibiotic selection and grown at 28°C.

#### *TRIPARENTAL MATING*

In order to transfer the *pEDV6* plasmid from the *E. coli* strains to recipient *P. syringae lux* strain, triparental mating was used. This form of conjugation requires the assistance of a third strain, an *E. coli* 'helper' strain carrying the *pRK2013* plasmid designated as strain HB101. 24 hour liquid cultures were mixed 3:3:1 (donor:recipient:helper). The mixes were spotted on non-selective L medium for 10 hours at 28°C, and then streaked onto selective media for  $\sim 48$  hours at 28°C. Positive colonies were checked using colony PCR.

### *ESTIMATION OF COLONY FORMING UNITS / ML*

The optical density at 600 nm (OD<sub>600</sub>) was used to estimate the number of colony forming units (cfu) / ml of liquid bacterial cultures. To calculate the OD<sub>600</sub> a Bio Photometer plus spectrophotometer (Eppendorf) was used.

## 2A.3 PLANT METHODS

### *ARABIDOPSIS THALIANA GROWTH*

*At* seeds were sown onto Levington F2 compost and stratified for 7 days at 5°C. They were then transferred to controlled growth chambers with 8 hours per day of light and temperature maintained at 23-24°C. After approximately 2 weeks, seedlings were transferred to pots with Levington F2 compost supplemented with grit and INTERCEPT insecticide, and maintained at the same conditions as above. For seed production, plants were transferred to growth facilities with 12 hours per day of light.

### *NICOTIANA SPP. GROWTH*

*Nicotiana benthamiana* and *N. tabacum* were treated as the *Arabidopsis* plants, but individual pots contained only F2 compost and they were grown under 10 hours per day of light.

### *AGROBACTERIUM TRANSFORMATION OF ARABIDOPSIS*

*Agrobacterium* mediated transformation was carried out using the floral dip method as described by Clough and Bent (1998). Briefly, 6 to 8 week old *At* plants were dipped in a solution of *A. tumefaciens* GV3101 at OD<sub>600</sub> of 0.5 (2.5x10<sup>8</sup>cfu /ml). Recovered T1 seeds were sterilized and selected on GM medium with the appropriate antibiotic. GM medium had the following composition for 1 l: 4.3 g MS salts, 0.1 g myoinositol, 0.59 g MES, 1 ml 1000X GM vitamin stock, 8 g Bacto agar, pH 5.7. 100 ml of 1000X GM contains 0.1 g thiamine, 0.05 g pyridoxine, 0.05 g nicotinic acid).

### *SEED STERILIZATION*

*At* seeds were sterilized by chlorine gas. 100 ml of 10% sodium hypochlorite was mixed with 3 ml of 36% hydrochloric acid and left with seeds for 3-6 hours inside of a sealed container.

#### *AGROBACTERIUM TUMEFACIENS TRANSIENT ASSAYS IN NICOTIANA SPP.*

*A. tumefaciens* strains were streaked on selective media and incubated at 28°C for 1-2 days. Single colonies were transferred to liquid L media with appropriate antibiotics, and were cultured for 1-2 days at 28°C in a shaking incubator (200 rpm). The resulting cultures were spun at 3000 rpm for 10 minutes and resuspended in 10 mM MgCl<sub>2</sub>, 10 mM MES with 20 µM acetosyringone. The OD<sub>600</sub> was adjusted to 0.5 (2.5x10<sup>8</sup> cfu/ml) for each strain or mixture of strains. The abaxial surface of *Nicotiana benthamiana* or *Nicotiana tabacum* leaves were infiltrated with a needleless 1 ml syringe. Leaves were checked for HR and imaged 3 or 4 dpi.

#### *PSEUDOMONAS SYRINGAE INFILTRATION AND GROWTH ASSAYS*

*P. syringae* strains were streaked onto selective media and incubated at 28°C for 1-2 days. The resulting mass of colonies was scraped from the plate and resuspended in 10 mM MgCl<sub>2</sub>. The OD<sub>600</sub> was adjusted to 0.2 (2x10<sup>8</sup> cfu / ml). 0.02% v/v Silwet-77 was added to the solution and 4 4-5 week old *At* plants were sprayed per experimental group. The plants were covered with a transparent lid. 3-4 dpi, a cork borer was used to take 3 punches (total 1cm<sup>2</sup>) from six leaves per group of four plants. The punches were added to an Eppendorf tube and crushed with a pestle in 1 ml dH<sub>2</sub>O. Serial dilutions (10<sup>-2</sup>, 10<sup>-3</sup>, 10<sup>-4</sup>, 10<sup>-5</sup>) were spotted on selective media, and bacterial colonies counted after 2 days of incubation at 28°C.

#### *TRANSIENT EXPRESSION WITH BIOLISTICS*

Plasmids were prepared to at least 500 ng/µl using the Qiagen Plasmid Midi kit. 1 µm gold microcarriers (particles) were prepared at 30 mg/ml in 50% glycerol. To prepare for bombardment, 50 µl aliquots of microcarriers were made. 1µg of *pK2GW7:GUS* (encoding β-glucuronidase) and 1µg of the experimental plasmid were mixed to a volume of 5 µl. This DNA, 50 µl of CaCl<sub>2</sub> (2.5M) and 20 µl spermidine (0.1M) were added in succession as the mixture was gently vortexed. The DNA coated microcarriers were pelleted and washed 3 times with ethanol. They were

then resuspended in 50 µl ethanol and coated onto 6 macrocarriers (plastic disks). Bombardment was carried out using a BioRad PDS-1000 (He) delivery system per the manufacturer's instructions. During the bombardment, two *At* leaves were laid side by side (one experimental, one control) on a Petrie dish with 1% agarose and sterile filter paper. 6 repeats were carried out per group. Afterwards, the petrie dishes were sealed with parafilm and incubated for 24 hours at 25°C.

#### *GUS STAINING AND QUANTIFICATION*

Bombarded leaves were individually stained for four hours at 37°C with 5-bromo-4-chloro-3-indolyl glucuronide. They were then cleared with 100% propanol for 3-4 days. GUS expression was measured by imaging the leaves with a Leica DMR. The number of GUS spots was counted by eye using the cell counter plugin for ImageJ. The author was unaware of which leaves he was counting to make the process unbiased. The non-parametric Wilcoxon Rank-sum test (Mann and Whitney, 1947) was used in MS Excel to establish the statistical significance of results.

#### *TRYPAN BLUE STAINING OF ALBUGO STRUCTURES*

*Albugo* infected leaves were transferred to 10 ml universal tubes and ~4 ml of 1% trypan blue w/v in lactophenol (NBS Biological Limited) mixed 1:1 with 100% ethanol. The leaves were boiled for 4 minutes, the trypan blue discarded and the leaves destained in chloralhydrate (2.5 g/ml; Sigma Aldrich). The chloralhydrate solution was changed twice over a period of 3-10 days. To visualise, the leaves were mounted in 60% glycerol.

#### *CONFOCAL MICROSCOPY*

Confocal microscopy was used to determine the sub-cellular localization of several GFP-tagged proteins. *N. benthamiana* leaves were infiltrated with *A. tumefaciens* strains as described above. 2-3 dpi, sections of the infiltrated zone of the leaves were fixed on glass microscopy slides with water. The GFP-tagged proteins were visualized using a Leica DM6000B/TCS SP5 (Leica Microsystems). GFP was excited using a wavelength of 488 nm.

#### *LIST OF PLANT ACCESSIONS AND MUTANTS*

The following accessions were obtained from lab stocks: Col-0, Col-THO, Ws-2, As-77, BAT1, CIBC-5, Ei-2, EkN 3, Fly2-1, Fly2-2, Fri2, Ge-0, GrA-5, Hov1-7, HR-10, HR-5,

Kin-0, Kni-1, Knox-18, Ksk-1, NFA-10, Pna-17, Ren-11, Rev-3, RRS-7, S294BeL4, San-2, Sf-2, Sq-1, T1010, T1160, T450, T800, T860, TDr9, Ts-1, Ts-5, Udul 1-34, Uk-1, UllA-1 and UllA-2. The mutants: *Ws-eds1-1* (Parker et al., 1996) and *Col-0-ndr1-1* (Century et al., 1997) were also obtained from within the lab.

## 2A.4 BIOINFORMATICS METHODS

### *QUALITY ASSESSMENT OF ILLUMINA READS*

Read quality was assessed using the FASTX toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)).

### *READ ALIGNMENT*

Reads were aligned to reference genomes using the Burrows Wheeler Aligner (BWA) version 0.6.1 using the default settings (including minimum seed length of 19) (Li and Durbin, 2010). I also included the soft trim option, `-q 20` on the `aln` command step. This trims any bases with quality of less than 20 from the end of the aligned reads.

### *VARIANT CALLING*

Variants were called using the SAMTools version 0.1.8 package, using default parameters (Li et al., 2009).

### *GENOME ALIGNMENT VISUALIZATION*

Genome alignments were visualized using the Integrated Genome Viewer (Thorvaldsdóttir, Robinson and Mesirov, 2013).

### *PREDICTION OF THE EFFECTS OF POLYMORPHISMS*

SNPEff version 3.2 was used to predict the effects of polymorphisms detected by the above methods (Cingolani et al., 2012).

### *THE VARITALE PIPELINE*

The Varitale pipeline was principally designed by Naveed Ishaque. The pipeline is an integrated suite of 3 Perl scripts which are integrated with PAML 4 (Yang, 2007), PHASE (Stevens and Scheet, 2005), and DNAsp (Librado et al., 2009) to calculate population genetic statistics as reported in chapter 4 for populations of genome-

sequencing data. These scripts were utilized using the default parameters. Under default parameters alleles with less than 0.5x average coverage and greater than 2x average coverage are excluded.

#### *IDENTIFICATION OF RECOMBINATION EVENTS*

A custom Perl script was written to produce contigs “corrected” with the SNPs from each sequenced *AI* isolate. Contigs longer than 10 kb were analysed using recombination detection program (RDP) version 3 (Martin and Ribicki, 2000). To map the distance from each gene to its nearest recombination events, a custom Perl script in combination with SNPEff was used.

#### *BAYES-EMPIRICAL-BAYES ANALYSIS*

Bayes-empirical-bayes analysis was performed at <http://selecton.tau.ac.il> using the M8 beta + w >= 1 model (Stern et al., 2007). Plots were produced in MS Excel.

#### *GENE-DENSITY PLOTTING IN R*

Gene density plots were generated using the protocol described by Saunders et al, (2014).

#### *GENOME ASSEMBLY*

Genome assembly using Illumina 76 bp and 100 bp paired end data were made using Velvet version 1.2.08 (Zerbino and Birney, 2008). For each assembly several logical kmer lengths were used to produce an optimum assembly. Assemblies using Illumina 300 bp paired end data were made using SPADes version 3.1.0 (Bankevich et al., 2012) in the careful mode. SPADes incorporates a kmer scanning feature allowing the optimum assemblies from multiple kmers to be found automatically.

#### *BLAST*

BLAST searches (blastn, megablast, tblastn, tblastx, blastp etc) were either carried out at blast.ncbi.nlm.nih.gov, or locally using BLAST+ version 2.2.29+.

#### *GENOME COVERAGE STATISTICS*

BEDtools version 2.11.2 (Quinlan and Hall, 2010) was used to compute the per base coverage values in Illumina based genome alignments.

#### *ALIGNMENTS AND PHYLOGENETIC TREES*

Nucleotide and protein alignments were made using Clustal Omega (Sievers et al., 2011). Neighbour-joining phylogenetic trees were generated using either Clustal Omega (Sievers et al., 2011) or the *CLC Main Workbench* version 5.7.2 from CLC Bio. Radial trees were manipulated using Figtree (<http://tree.bio.ed.ac.uk/software/figtree/>).

## CHAPTER 2B: DEVELOPMENT OF NOVEL GENOMICS TECHNIQUES

### 2B.1 ASSOCIATION GENOMICS IN PLANT PATHOGENS

#### *INTRODUCTION*

The improvement of DNA sequencing technologies in the past decade has made whole genome sequencing of multiple individuals a practicality (Ong et al., 2013; Thudi et al., 2012). In the human disease and cancer fields, genome-wide association studies involving hundreds of individual genomes are commonplace (Mooney et al., 2014; van der Sijde, Ng and Fu, 2014). These studies have resulted in a wide range of success, mainly dependent on whether the genetics underlying a particular trait is polygenic. These studies are mainly hampered by the fact that most human disorders can be linked to multiple possible allelic variations at multiple loci. This combined with a high degree of background mutation, means that even large populations can only give results that often only partially explain the observed phenotypic variation. (Stranger, Stahl and Raj, 2011; van der Sijde, Ng and Fu, 2014).

Plant pathogens potentially present an opportunity for association genomics to work well. It is well established, and described in chapter 1, that the outcome of interactions between specialised pathogens and their plant hosts can come down to the interaction of a single *AVR-R*-gene pair or host-specific toxin/ receptor pair. This leads to excess variation relative to the rest of the genome at these loci, as they under strong selective pressures to avoid/gain recognition (Stuckenbrock and McDonald, 2009).

One study has already proven the efficacy of “association genomics” to identify an *AVR*-gene. de Jonge et al (2012) sequenced the genomes of 10 strains of *Verticillium dahliae* and discovered a 50 kb region of sequence specific to strains avirulent on *Ve1* plants. Within this 50 kb region, a secreted protein recognised by *Ve1* was identified.

Other studies have sequenced inferred effector repertoires using lower-throughput technologies and associated polymorphisms, thus identifying *AVR*-genes with associated polymorphisms (Armstrong et al., 2005; Yoshida et al., 2009).

My objective for this section was to develop a bioinformatics pipeline that can utilise high-throughput sequencing data from multiple pathogen strains to make predictions of potential causal gene alleles associated with specific phenotypes. Although multiple programs exist for this kind of analysis, for example GENECLUSTER (Su et al., 2009), SNPTEST (Marchini et al., 2007) and PLINK (Ferreira and Purcell, 2009) these programs are primarily designed for large SNP-only datasets and do not compute amino-acid level changes (Galesloot et al., 2014). For practical reasons I needed to develop a system that could take advantage of a small number of samples to the maximum effect, examining non-synonymous polymorphisms.

#### *METHODOLOGY*

I developed this method over several iterations and much trial and error.

As described in chapter 2a, I used a method for analysing Illumina data that involves alignment of short read sequences to a reference genome using BWA (Li and Durbin, 2010) and then Samtools (Li et al., 2009) to predict SNPs and Indels compared to the reference genome. The standard output of polymorphisms is the variant-call-format (VCF). A VCF file is a tab-delimited-text file containing the following information at each location where a polymorphism has been detected (an abbreviated example below):

Chromosome/Contig	Genomic coordinate	Ref base	New base	Read depth	...
500	561467	T	G	25	...

There are multiple further fields containing information such as the mapping quality, and the bases encoded by each of the reads aligning at this coordinate. These allow the polymorphisms to be filtered based on quality and depth. Indel polymorphisms are also represented, using characters such as -A to signify a deletion or +A for an insertion.

The program PileLine developed by Glez-Peña et al (2011) is designed to work with this format of data, and I used the 'nsmc' command to compile data from multiple

sequenced samples into a single table, with simple YES and NO to indicate whether each sample has the mutation, for example:

Chromosome/Contig	Genomic coordinate	Ref base	New base	Sample 1	Sample 2	Sample 3	Sample 4	Samples	# YES
500	61467	T	G	YES	YES	NO	YES	4	3

Where there are multiple different polymorphisms at the same site, additional lines are created. Thus such a table, although generally very large, can be searched for specific patterns of polymorphisms. For this purpose I wrote custom Perl scripts, splitting the lines within the file based on tabs and conditionally printing lines based on the pattern of YES and NO. An example of this code is pasted below:

```
my $table = $ARGV[0];

open(IN, "<$table") or die "error opening $table for reading";

while (my $line = <IN>) {

    my @cells = split(/\t/, $line);

    if ($cells[4] eq "NO" && $cells[5] eq "NO" && $cells[6] eq "NO" && $cells[7] eq
"YES" && $cells[8] eq "NO"){

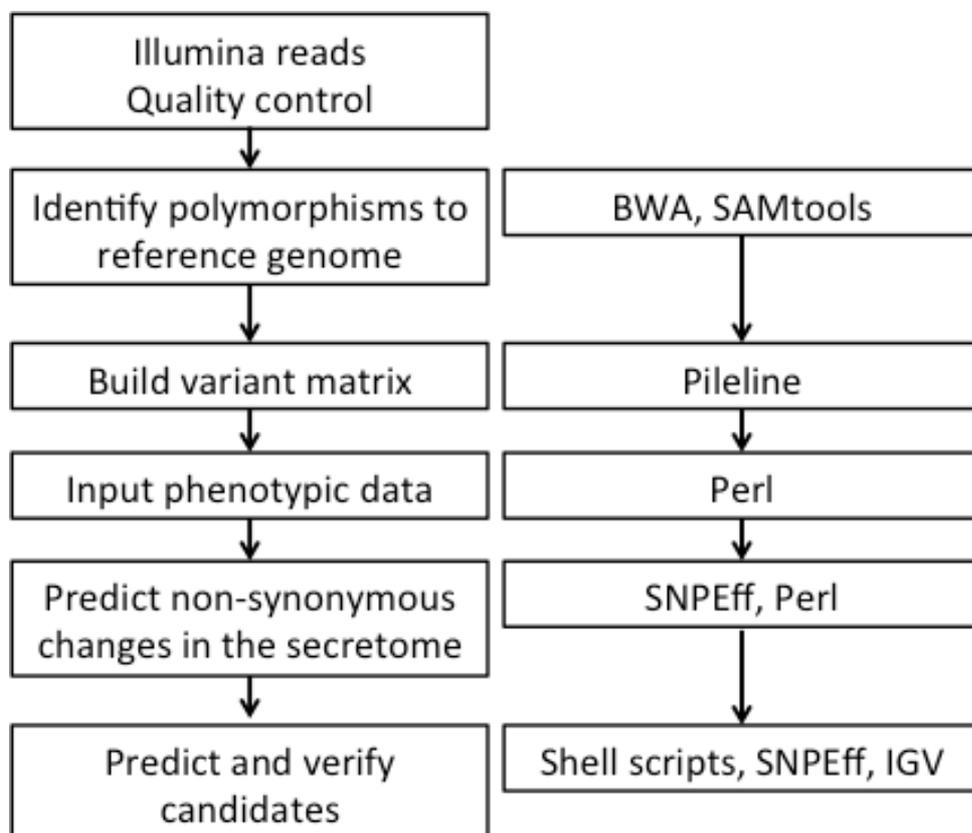
        print $line;}

}

close IN;
```

Therefore, polymorphisms associated with a specific phenotype can be extracted. However, I wanted to find non-synonymous polymorphisms within secreted proteins. To do this, my pipeline uses a program called SNPEff (Cingolani et al., 2012). SNPEff can accept VCF files as input and, provided with a set of gene models, make predictions of the effects of SNPs and Indels. It also annotates the gene within which each polymorphism occurs and has a detailed description of the effect of each polymorphism.

I then use two further Perl scripts to extract i) non-synonymous polymorphisms and ii) those within the list of predicted secreted proteins. It's then possible to sort potential candidates by the number of associated non-synonymous changes that they have. This can then provide the basis for a prediction of candidate genes. Once predictions are made, and a limited number of candidates are identified, the manual verification of the polymorphisms within a gene is advised, using a genome browser such as the integrated genome viewer (IGV) (Thorvaldsdóttir et al., 2013). Additional statistics, such as the non-synonymous/synonymous polymorphism ratio and those based on linkage disequilibrium should be used to help remove false-positives and narrow down candidates.



**Figure 2b.1. Detailed outline of the association genomics pipeline.** AVR gene candidates are predicted using association genomics with Illumina sequencing data.

*HYALOPERONOSPORA ARABIDOPSISIS AS A TEST CASE*

In order to test my association genomics method, I used *Hpa* genome data from the Jones lab. Within the lab, we previously obtained Illumina sequencing data for

seven *Hpa* races (Ishaque, Furzer et al., *in preparation*). There are several characterised *AVR*-genes in the *Hpa* system, known as *ATR* (*Arabidopsis thaliana* recognised) genes (Allen et al., 2004; Rehmany et al., 2005; Bailey et al., 2011; Goritschnig et al., 2012). Some *Arabidopsis* accessions that recognise these genes' products are known. However, there are other accessions where it is unknown which specific *ATR*-genes are recognised, as described in Krasileva et al (2011). Indeed Krasileva et al (2011) tested 83 *At* accessions at adult leaf stage with five *Hpa* races (Emoy2, Maks9, Emco5, Cala2 and Emwa1) (Table 2b.1). Out of those 83 accessions, 53 had unambiguous phenotypes. Therefore it is possible to use these data as a test set for the association genomics pipeline. From this collection six accessions that recognise  $ATR1^{Emoy2}$  are known (Nd-1, Zdr-1, Ws-0, Pu2-23, Est-1 and Ws-2) and three that recognise  $ATR13^{Emoy2}$  (Nd-1, Nok-3, N13) (Krasileva et al., 2011). Additionally Ler-1 can recognise  $ATR5$  (Bailey et al., 2011) and Wei-0  $ATR39$  (Goritschnig et al., 2012).

Group	Emoy2	Maks9	Emco5	Cala2	Emwa1	Accessions
1	NO	YES	NO	NO	NO	Ag-0, Bor-4, CS22491, Lz-0
2	NO	YES	YES	NO	NO	An-1
3	NO	NO	NO	NO	NO	Bor-1, Est-0, Ms-0, Mt-0, Pu2-7, RRS-7, Shahdara, Spr1-6, Wa-1, Zdr-1, Zdr-6
4	NO	YES	NO	NO	NO	Bur-0, Gu-0, Pro-0
5	NO	YES	NO	YES	YES	Cvi-0
6	NO	NO	NO	YES	NO	Ei-2, Se-0, Tamm-2,
7	NO	NO	NO	NO	YES	Fei-0
8	NO	NO	YES	YES	NO	Gy-0, Lov-5, Rmx-A180, Sorbo
9	NO	NO	YES	YES	NO	HR-10, Pu2-23,
10	NO	NO	YES	NO	NO	Knox-10, Knox-18, Pna-10, RRS-10,
11	NO	NO	NO	YES	NO	Kin-0, Kz-9, Ler-1, Nok-3, Var2-6
12	NO	NO	YES	NO	YES	Lov-1,
13	NO	YES	YES	NO	YES	Mz-0
14	NO	NO	NO	YES	YES	Nd-1, Ts-1
15	NO	YES	NO	YES	NO	Ren-1, Uod-7
16	NO	YES	NO	YES	NO	Wei-0
17	NO	NO	YES	NO	YES	Ws-0, Ws-2

**Table 2b.1. Interpolated phenotyping data from 5 *Hpa* races on 53 *At* accessions.** The accessions have been grouped depending on their phenotype; the colours signify resistant and susceptible (Red = resistant, yellow = susceptible). The YES and NO indicate the presence of the hypothetical (+/-) *AVR* mutations, which must be different from Emoy2, the reference race. Phenotyping data from Krasileva et al, (2012).

I applied the pipeline to predict candidates for *AVR*-genes for each of the 17 phenotypic groups. The results for selected groups are shown in table 2b.2. *ATR1* was the clear top candidate for group 14, containing Nd-1, which is known to

recognise it. *ATR5* was one of the top three candidates for recognition by Ler-1, where it is recognised. No significant candidates were identified for the Ws-0 and Ws-2 avirulence however. There are several strong candidates in accessions where no AVR-genes have been previously identified, for example *HpRXLR121* in accession group 1 and *HaRxLL38* in accession group 10.

Group	Accessions		Group	Accessions
14	Nd-1, Ts-1		11	Kin-0, Kz-9, Ler-1, Nok-3, Var2-6
<b>NS polymorphisms</b>	<b>Genes</b>		<b>NS polymorphisms</b>	<b>Genes</b>
31	<b><i>ATR1</i></b>		14	<i>HaRxLL426</i>
8	<i>HaRxLL48</i>		12	<i>HaRxL76</i>
6	<i>HaRxL73</i>		11	<b><i>ATR5 (HaRxLL424)</i></b>
			10	<i>HaRxLL425</i>
Group	Accessions		Group	Accessions
1	Ag-0, Bor-4, CS22491, Lz-0		15	Ren-1, Uod-7
<b>NS polymorphisms</b>	<b>Genes</b>		<b>NS polymorphisms</b>	<b>Genes</b>
54	<i>Emoy2cDNA_HpRXLR121</i>		19	<i>HaRxL95</i>
26	<i>HaRxLL40</i>		13	<i>HaRxLL425</i>
25	<i>HaRxLL425</i>		12	<i>HaRxL93</i>
24	<i>810723(RXLR)</i>		10	<i>HaRxL92</i>
20	<i>HaRxLL29</i>			
19	<i>HaRxL95</i>			
Group	Accessions		Group	Accessions
10	Knox-10, Knox-18, Pna-10, RRS-10,		8	Gy-0, Lov-5, Rmx-A180, Sorbo
<b>NS polymorphisms</b>	<b>Genes</b>		<b>NS polymorphisms</b>	<b>Genes</b>
26	<i>HaRxLL38</i>		18	<i>HaRxL128</i>
15	<i>ATR13</i>		17	<i>HaRxL21</i>
14	<i>HaRxLL80</i>		10	<i>HaRxLL445</i>
14	<i>ATR1</i>		9	<i>RXLR19</i>
12	<i>HaRxL53</i>		6	<i>HaRxL62</i>
			6	<i>RXLR2</i>

**Table 2b.2. Selected *Hpa* avirulence gene candidates.** AVR gene candidates were predicted using association genomics with Illumina sequencing data and disease phenotype data. Known AVRs *ATR1* and *ATR5* are highlighted.

### CONCLUSIONS

Results from testing of the association genomics pipeline were consistent with its efficacy; *ATR1* was predicted as the strongest candidate for recognition by Nd-1 and *ATR5* was one of the best candidates for recognition by Ler-1. Nevertheless, the test also exposed some weaknesses of the system. *ATR1* was not predicted for recognition by Ws-0 or Ws-2, even though these accessions recognise *ATR1*<sup>Emoy2</sup>. However, they are resistant to Cala2 but do not recognise the Cala2 allele of *ATR1*, suggesting that they could recognise another effector from Cala2 (Volkan Cevik,

personal communication). Such complex genetics is a weakness of this association-based system; with such a low number of samples, absolute association is a necessity, but the underlying genetics may be more complicated. This kind of complication is also probably behind the failure to predict *ATR13*: in Nd-1 *ATR1* is also recognised meaning again that the association breaks down. The mantra of bioinformatics is very much in evidence: '*garbage in, garbage out*'. One way to avoid such complications is to work with accessions where the host genetics is established, particularly if a single gene is known to confer recognition.

In addition to proving the efficacy of the approach, this test may provide some useful information and could perhaps guide the discovery of new *Hpa AVR*-genes in previously untested *At* accessions. There are numerous candidates, almost entirely RxLR or RxLR-like effectors, with high association scores for different accession groups. Some of these, when assessed in conjunction with other data such as pN/pS, could merit testing for recognition.

## 2B.2 BRASSICACEAE *R*-GENE ENRICHMENT SEQUENCING

### *INTRODUCTION*

As alluded to in chapter 2b.1, advances in sequencing technology have made whole genome sequencing of multiple individuals a practicality. Consortium projects, for example the 1001 Arabidopsis genomes project (Cao et al., 2011), have now obtained and published online Illumina sequencing reads of up to 100 bp paired end for hundreds of *At* accessions. Analyses of these data in *At* have revealed that the NB-LRR class of genes, to which most known *R*-genes in *At* belong, are the most variable class of genes (Gan et al., 2011; Cao et al., 2011). They also belong to large multi-gene families (Meyers et al., 2003). Altogether this means that in *NB-LRR* rich regions of the genome, alignment of short reads becomes either impossible or unreliable due to the high level of polymorphism compared to the reference genome. Because of these factors, reliable *de novo* assembly of NB-LRR encoding genes is also difficult using short-read data. One solution to this issue is to obtain longer read data. Current generation Illumina MiSeq can produce up to 300 bp paired end reads in a relatively high-throughput manner, and PacBio can produce a variety of read lengths up to 10 kb, although at less read depth. Although whole genome shotgun sequencing of more samples with longer read technology is feasible in *At*, when one is interested in specifically the 'NB-LRRome' it is potentially wasteful to sequence the rest of the genome, particularly when much of it is already available for most accessions through the 1001 genomes project. Therefore I led the development and testing of a system to enrich NB-LRR sequences from DNA samples from plants of the *Brassicaceae* family

Capture array technology is designed so that only a desired fraction of a genome or transcriptome is captured and sequenced. Both solid- and liquid-phase synthetic RNA bait methods have been developed (Gnirke et al., 2009; Mamanova et al., 2010). The principle of these methods is the hybridization of adaptor-ligated DNA to an array of RNA bait oligos, capture of matching or similar sequences and the washing away of the remaining DNA. Jupe et al (2013) developed a protocol for *R*-gene enrichment sequencing (RenSeq). They used NB-LRR sequences from the assembled genomes of potato, tomato and pepper to build an Agilent *SureSelect*

*Target Enrichment System* in-solution library of 44 549 unique biotinylated 120mers. This was used to enrich 500 nt *NB-LRR* fragments from various *Solanum* species successfully, the study showing that around 80% bait-target identity is sufficient to anneal to a given sequence.

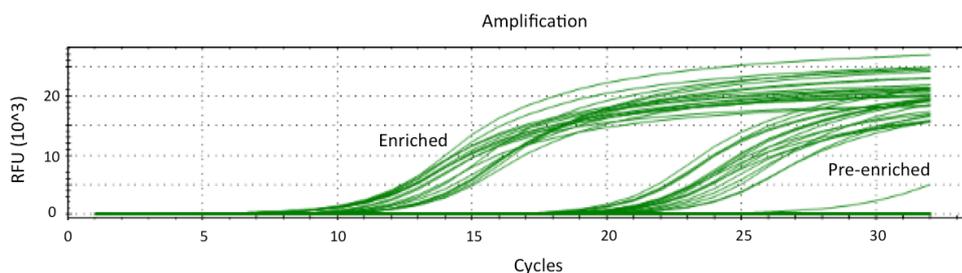
#### *METHODOLOGY*

With assistance from Johannes Hofberger (Wageningen University) a library of baits for Brassicaceae RenSeq was constructed *in silico*. We extracted 657 NB-LRR coding sequences from four species: *Arabidopsis thaliana* (including sequences from two accessions, Col-0 and Ler-0), *Aethionema arabicum*, *Brassica rapa* and *Eutrema parvulum* (the latter three as provided by Johannes Hofberger). These sequences were converted into 120 mers, with 30 bp tiling (overlap between consecutive baits). Any identical sequences were removed, and the library was pruned down to 20000 baits based on pairwise identity (the most identical pairs of baits having one of the pair removed). The sequences were then synthesised by MYcroarray (mycroarray.com) as a '*MYbaits: Custom Bait Library for Sequence Capture in Targeted Sequencing*'. The 120 mers are provided as a kit containing approximately  $6 \times 10^{12}$  biotinylated baits per microliter.

The enrichment of NB-LRR sequences from Brassicaceae for the reasons outlined above proved to be an approach that several Jones lab collaborators could apply to their projects. For the first experiment we devised a strategy whereby we would enrich 23 samples after the application of unique barcodes to each sample. The 23 samples were various accessions of *At*, *Brassica rapa*, *Brassica juncea*, *Capsella orientalis* and several F<sub>2</sub> bulk susceptible populations (I won't fully disclose the details of the samples from collaborators). I prepared high-quality, high molecular weight DNA from *At* Col-0, HR-5, Ksk-1, Ws-2 and a bulked HR-5 x Ws-2 F<sub>2</sub> population, and received the remaining samples in either liquid or pellet form from the various collaborators. Dr Florian Jupe provided high-quality *Solanum lycopersicum* DNA as a control to test the performance of the array against distantly related sequences. I sheared each sample with the Illumina covaris to a fragment size of approximately 800 nt. I used the New England Biosciences NEBNext Ultra kit to add Illumina adaptors to each sample, and then amplified with 23 different

multiplexing oligos to generate 23 separate libraries for the enrichment. The MYBaits protocol recommends enriching each library sample separately. However this would not be cost effective as there would be a vast excess of RNA baits compared to compatible DNA fragments. Although it is theoretically possible to combine 23 samples into one pool prior to enrichment, I opted to create 6 pools of adaptor ligated DNA samples for the enrichment procedure. This was carried out per the manufacturer's instructions, including a 24-hour hybridisation under paraffin oil at 65°C in a thermocycler. Post-hybridisation, the biotinylated baits were recovered on magnetic Dynabeads® MyOne™ streptavidin C1 and PCR was carried out for between 9 and 14 cycles on- bead with the Illumina outer adaptor primers P5 and P7 to generate enough DNA for sequencing. These libraries were quantified and were used for a qPCR (in triplicates) to determine the level of enrichment. qPCR was carried out on the same quantity of DNA from pre- and post enriched library mixes with primers amplifying the *At* TIR-NB-LRR encoding *RRS1* gene. There was at least one *At* sample in each of the six mixes. The qPCR revealed an average of ~3000 fold enrichment for *RRS1* in the enriched libraries compared with non-enriched (figure 2b.2).

Finally, the six pools were quantified using both agarose gel electrophoresis and picogreen based quantification, and mixed into an equimolar pool (Double the quantity of the *Brassica* samples was added to account for their larger genomes). Invitrogen AMPure magnetic beads were used to remove primer-dimers and other impurities from the final DNA pool. Sequencing was carried out by TGAC, Norwich, on an Illumina MiSeq generating 300 bp paired-end reads. De-multiplexing of the read data was also carried out by TGAC.



**Figure 2b.2. qPCR amplification curves of enriched and pre-enriched DNA libraries with *RRS1* primers.** A pair of primers amplifying a 250 bp *RRS1* product were used to assess the relative level of *RRS1* specific DNA in libraries pre- and post- *R*-gene enrichment. Most of the negative controls did not amplify, but one might have had a very small amount of contamination and began to amplify after 26 cycles.

## RESULTS

Following the de-multiplexing of the Illumina reads, a basic assessment of quality and quantity was made. Table 2b.4 summarises these data. Although the samples were adjusted as far as possible to take account of the estimated molecular weight and quantity of each library, there is significant variation among the read number achieved between different samples. The highest number of reads was achieved from the HR-5 x Ws-2 F2 bulk ( $3.3 \times 10^6$ ) and the lowest from the *Capsella orientalis* sample ( $1.6 \times 10^5$ ). This may be due to factors such as the differential clustering and ligation efficiency of different length fragments on an Illumina chip, and possible problems for the Illumina laser to make base calls if clusters are too long or too short compared to the mean. Nevertheless, according to projections based on an *At* 'NB-LRRome' of ~150 kb of DNA sequence, even the lowest yielding sample should have around 100 deep coverage of its NB-LRRome. The *Solanum lycopersicum* sample had a reasonable amount of reads ( $5 \times 10^5$ ) but had poor quality scores.

Sample	# Paired Reads	R1 Q30 to Base	R2 Q30 to Base	# Bases sequenced	Projected depth on 1 <i>At</i> NB-LRRome
<i>Arabidopsis</i> accession Col-0	1,866,229	249	149	933,114,500	1,244
<i>Arabidopsis</i> accession Ws-2	858,433	249	149	429,216,500	572
<i>Arabidopsis</i> accession HR-5	964,135	249	199	482,067,500	643
HR-5 x Ws-2 pool F2 suscept (n~400)	3,348,182	199	199	1,674,091,000	2,232
<i>Arabidopsis</i> accession Ksk-1	396,927	249	199	198,463,500	265
<i>Arabidopsis</i> F2 pool	1,442,531	249	199	721,265,500	962
<i>Arabidopsis</i> accession	1,686,166	299	199	843,083,000	1,124
<i>Solanum lycopersicum</i>	599,901	99	69	299,950,500	400
<i>Arabidopsis</i> accession	689,410	249	149	344,705,000	460
<i>Brassica rapa</i> accession Chiifu	594,582	199	149	297,291,000	396
<i>Conringia orientalis</i>	161,666	199	149	80,833,000	108
<i>Arabidopsis</i> accession	201,214	249	199	100,607,000	134
<i>Arabidopsis</i> accession	266,459	249	199	133,229,500	178
<i>Arabidopsis</i> accession	1,125,001	249	199	562,500,500	750
<i>Arabidopsis</i> accession	1,570,384	249	199	785,192,000	1,047
<i>Arabidopsis</i> accession	1,505,052	199	199	752,526,000	1,003
<i>Arabidopsis</i> accession	1,136,029	249	149	568,014,500	757
<i>Arabidopsis</i> F2 pool	929,064	199	149	464,532,000	619
<i>Brassica oleracea</i> accession A12	292,699	249	199	146,349,500	195
<i>Brassica oleracea</i> accession	1,151,197	299	199	575,598,500	767
<i>Brassica oleracea</i> accession	775,358	249	199	387,679,000	517
<i>Arabidopsis</i> accession	1,338,534	249	149	669,267,000	892
<i>Arabidopsis</i> accession	363,891	199	149	181,945,500	243
Totals	23,263,044			11,631,522,000	

**Table 2b.3. Read statistics.** The details of some samples are obscured to respect the collaborators who sent them. The R1 or R2 Q30 to base refers to the number of bases in the forward and reverse reads that have at least a quality score of 30 (on a scale of 0 to 40). The projected depth on 1 *At* NB-LRRome was calculated using the projected number of bases in one *A. thaliana* NB-LRR encoding gene complement (750 kb).

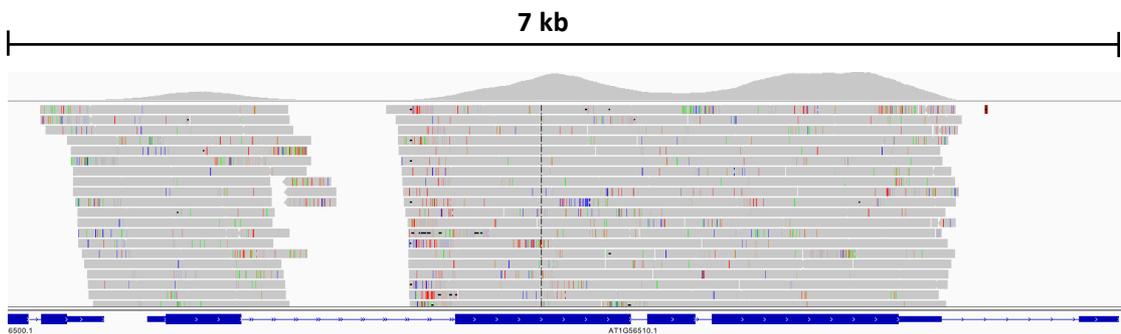
The reads from Col-0, Ws-2, HR-5, Ksk-1 and the HR-5 x Ws-2 F<sub>2</sub> bulk susceptible pool were aligned to the Col-0 reference genome with BWA. As the positive control, I will present the results from Col-0 here. I will discuss in chapter 6 the results from HR-5, Ws-2 and the F<sub>2</sub> bulk. I used a custom Perl script to extract the depth and breadth of each *At* TAIR10 gene models. According to this analysis, each of the 140 NB-LRRs used in the array was covered 100% by reads, and the lowest in terms of read coverage was *AT4G16900*, covered to 522 deep (on average over its length). The average coverage for genes included in the array was 951 deep, in comparison to 2.1 deep for genes not included in the baits. However, many genes encoding partial NB-LRRs were enriched, and several full length NB-LRRs weren't included in the bait library design because their NB or LRR domain annotation predictions did not meet the minimum confidence threshold. Table 2b.5 shows a summary of all of the NB-LRRs from *At* Col-0 chromosome 1 and whether they are covered by the RenSeq reads and to what depth over their length. It also shows the partial NB-LRRs, as annotated by Meyers et al, (2003) and/or Nandety et al (2013).

Gene	Mean read coverage	Breadth of coverage	Protein Class	Annotation	In bait library?
AT1G10920	977.59	100	CC-NBS-LRR	LOV1	NO
AT1G12210	919.28	100	CC-NBS-LRR	RFL1	YES
AT1G12220	1028.48	100	CC-NBS-LRR	RPS5	YES
AT1G12280	1039.91	100	CC-NBS-LRR		YES
AT1G12290	1016.68	100	CC-NBS-LRR		YES
AT1G15890	995.81	100	CC-NBS-LRR		YES
AT1G17600	1084.46	100	TIR-NBS-LRR		YES
AT1G17610	85.66	100	TIR-NBS		NO
AT1G17615	1.69	74	TIR-NBS		NO
AT1G27170	876.92	100	TIR-NBS-LRR		YES
AT1G27180	1008.12	100	TIR-NBS-LRR		YES
AT1G31540	687.51	100	TIR-NBS-LRR	RAC1	YES
AT1G33560	1079.88	100	CC-NBS-LRR	ADR1	YES
AT1G47370	0.71	51	TIR-X		NO
AT1G50180	1.79	71	CC-NBS-LRR		NO
AT1G51270	49.66	60	X-TIR		NO
AT1G51480	729.22	100	CC-NBS-LRR		YES
AT1G52660	1360.8	100	CC-NBS		YES
AT1G52900	0.4	40	TIR-X		NO
AT1G53350	1039.01	100	CC-NBS-LRR		YES
AT1G56510	816.17	100	TIR-NBS-LRR	WRR4	YES
AT1G56520	733.24	100	TIR-NBS-LRR		YES
AT1G56540	874.7	100	TIR-NBS-LRR		YES
AT1G57630	790	100	TIR-NBS-LRR		NO
AT1G57670	1.46	80	TIR-X		NO
AT1G57830	7.03	100	TIR-X		NO
AT1G57850	2.67	100	TIR-X		NO
AT1G58390	809.54	100	CC-NBS-LRR		NO
AT1G58400	679.63	100	CC-NBS-LRR		YES
AT1G58410	173.18	100	CC-NBS-LRR		NO
AT1G58602	294.85	100	CC-NBS-LRR		NO
AT1G58807	608.77	100	CC-NBS-LRR		NO
AT1G58848	458.6	100	CC-NBS-LRR		NO
AT1G59124	641.35	100	CC-NBS-LRR		YES
AT1G59218	458.98	100	CC-NBS-LRR		NO
AT1G59620	958.9	100	CC-NBS-LRR	CW9	YES
AT1G59780	781.71	100	CC-NBS-LRR		YES
AT1G60320	3.3	100	TIR-X		NO
AT1G61105	0.86	79	TIR-X		NO
AT1G61180	1123.11	100	CC-NBS-LRR		YES
AT1G61190	1139.55	100	CC-NBS-LRR		YES
AT1G61300	1182.53	100	NBS-LRR		YES
AT1G61310	1185.77	100	CC-NBS-LRR		YES
AT1G62630	1036.75	100	CC-NBS-LRR		YES
AT1G63350	929.47	100	CC-NBS-LRR		YES
AT1G63360	1245.79	100	CC-NBS-LRR		YES
AT1G63730	730.59	100	TIR-NBS-LRR		YES
AT1G63740	950.93	100	TIR-NBS-LRR		YES
AT1G63750	1026.01	100	TIR-NBS-LRR		YES
AT1G63860	856.3	100	TIR-NBS-LRR		YES
AT1G63870	752.63	100	TIR-NBS-LRR		YES
AT1G63880	878.59	100	TIR-NBS-LRR		YES
AT1G64070	750.39	100	TIR-NBS-LRR	RLM1	YES
AT1G65390	1.85	97	TIR-X		NO
AT1G65850	706.26	100	TIR-NBS-LRR		YES
AT1G66090	161.55	100	TIR-NBS		NO
AT1G69550	808.42	100	TIR-NBS-LRR		YES
AT1G72840	1349.63	100	TIR-NBS-LRR		YES
AT1G72850	40.5	100	TIR-NBS		NO
AT1G72860	1188.81	100	TIR-NBS-LRR		YES
AT1G72870	1019.15	100	TIR-NBS		YES
AT1G72890	17.98	94	TIR-NBS		NO
AT1G72900	1.03	34	TIR-NBS		NO
AT1G72910	0.34	34	TIR-NBS		NO
AT1G72940	1.37	82	TIR-NBS		NO
AT1G72950	0.92	45	TIR-NBS		NO

**Table 2b.4. RenSeq coverage of chromosome 1 NB-LRRs and partial NB-LRRs.** The depth and breadth of Illumina read coverage over NB-LRRs included in the oligo bait library, and those genes annotated as possible NB-LRR-related fragments. Mean read coverage is the average depth of coverage across the length of the gene model. Breadth of coverage is the percentage of the gene model that is covered by at least one read. The protein class column is a composite of annotations made by Meyers et al, (2003) and/or Nandety et al (2013). Certain possible NB-LRRs or partial NB-LRRs were not included in the bait library. The last column indicates if each gene was included.

## CONCLUSIONS

This analysis of the Col-0 reads revealed that the bait library was functional in enriching the DNA samples for the NB-LRR encoding genes as designed. It also contained sequences similar enough to enrich for many NB-LRRs that were not included in the design, and some of the partial NB-LRR encoding genes showed good coverage. However, a substantial number of genes annotated as containing NB-LRR like domains had a very low depth and breadth of coverage suggesting that their DNA sequences are less than 80% similar to those included in the bait library. One issue not raised in the results was the introns. Because the bait library was constructed using CDSs, the baits do not contain introns. Although there is some capture of sequences on the borders of NB-LRR CDSs, for some genes there are noticeable gaps in the introns (figure 2b.3). This could pose a challenge when *de novo* assembly of these reads is attempted, as different exons of a single gene could be assembled into different contigs if the reads do not overlap in the introns. For a future bait library design it might be wise to include intron sequences.



**Figure 2b.3. RenSeq coverage of *At1g56510* (*WRR4*).** The exons of TIR-NB-LRR encoding *WRR4* are deeply covered by Illumina reads, but they do not cover the long first introns. Col-0 RenSeq reads were aligned to the Col-0 genome using BWA and SAMTools and visualized in IGV. In the upper panel, an overview of read coverage is displayed. In the lower panel, individual Illumina reads are shown. Grey bars represent individual Illumina reads; the small colored dashes within show discrepancies to the reference in individual reads. The blue bars below represent the annotated gene models, the dotted line showing introns and the blocks representing exons.

## CHAPTER 3: COLLECTION, IDENTIFICATION AND PHENOTYPING OF NEW *ALBUGO LAIBACHII* ISOLATES

### 3.1 INTRODUCTION

As described in the introduction, *Hpa* has become an important and well-characterised model biotroph pathogen (Coates et al., 2010). The success of *Hpa* research rested on collections of natural isolates from the field in the late 1980s and 1990s by Paul H Williams and Eric Holub (Holub, Beynon and Crute, 1994; Holub, 2008). In particular, the discovery of the differential recognition of different isolates, and the subsequent genetics carried out with these isolates led to the discovery of the *ATR* genes, contributing to the discovery of the RXLR motif, and shaped the way we think about the evolution and population genetics of effectors in filamentous biotrophic pathogens (Allen et al., 2004; Rehmany et al., 2005).

Also during the early 1990s, *Al* isolates were sampled from wild *At* populations. Holub et al (1995) screened a collection of *At* ecotypes for resistance against two isolates of *Al*, Alem1 (*Al* East Malling 1) and Acks1 (*Ac* Keswick 1). It is unknown whether Acks1 is *Al* or *Ac*, but based on its host range it is most likely *Al*. They found several apparent differential resistances and that reported ~15% of the panel of *At* accessions was resistant to at least one isolate. In contrast 50%-70% of *At* accessions are resistant to at least one *Hpa* race (Holub, Beynon and Crute, 1994; Nemri et al., 2010). It was not until Thines et al, (2009) that another *Al* isolate entered the literature (Nc14). In this paper *Al* was defined as a distinct species based on the polymorphism within its ITS1 and *cox2* sequences and differences in oospore morphology compared to *Ac*. Kemen et al (2011) later sequenced the genomes of Nc14 and Em1, and established that there were only around 15 000 polymorphism between the two isolates, which were collected in different locations and around 15 years apart.

Thus there are several important questions that can be answered by collecting new isolates. How diverse is the *Al* population in terms of nucleotide polymorphism and host range on *At* accessions? Both *Al* and *Ac* are capable of parasitizing *At* (Thines et al., 2009) and are indistinguishable to the naked eye; which is the most prevalent in

the field? To what extent does sexual recombination occur in the field? And finally, can the natural diversity of *Al-At* interactions lead to the discovery of novel components responsible for virulence and resistance in host and pathogen?

In this short first results chapter I will describe the collection, identification and phenotyping of new *Al* isolates.

## 3.2 RESULTS

### 3.2.1 COLLECTION OF NEW ISOLATES

In the period 2010-2011 8 *Albugo*-infected *At* field isolates were collected from various locations in the UK and Germany. Each isolate was germinated on *At* line Col-THO (Col-THO is Col-5 transformed with *RPW8*, an R gene against powdery mildew). I purified the isolates where possible by single spore propagation for three generations. The 4 successfully purified isolates were named: Ash4 (Ashbrittle, Somerset, UK, collected with assistance of Dr Nicola Perera), Abo1 (Ashbourne, Derbyshire, UK, collected by Dr Eric Kemen and Dr Ariane Kemen), Sua1 (Stratford-upon-Avon, Warwickshire, UK, collected by Professor Ian Crute) and Went1 (Tübingen, Baden-Württemberg, Germany collected by Dr Eric Kemen) (figure 3.1). One of the apparently *Albugo* infected samples, collected in Peterborough, did not produce a successful reinfection.



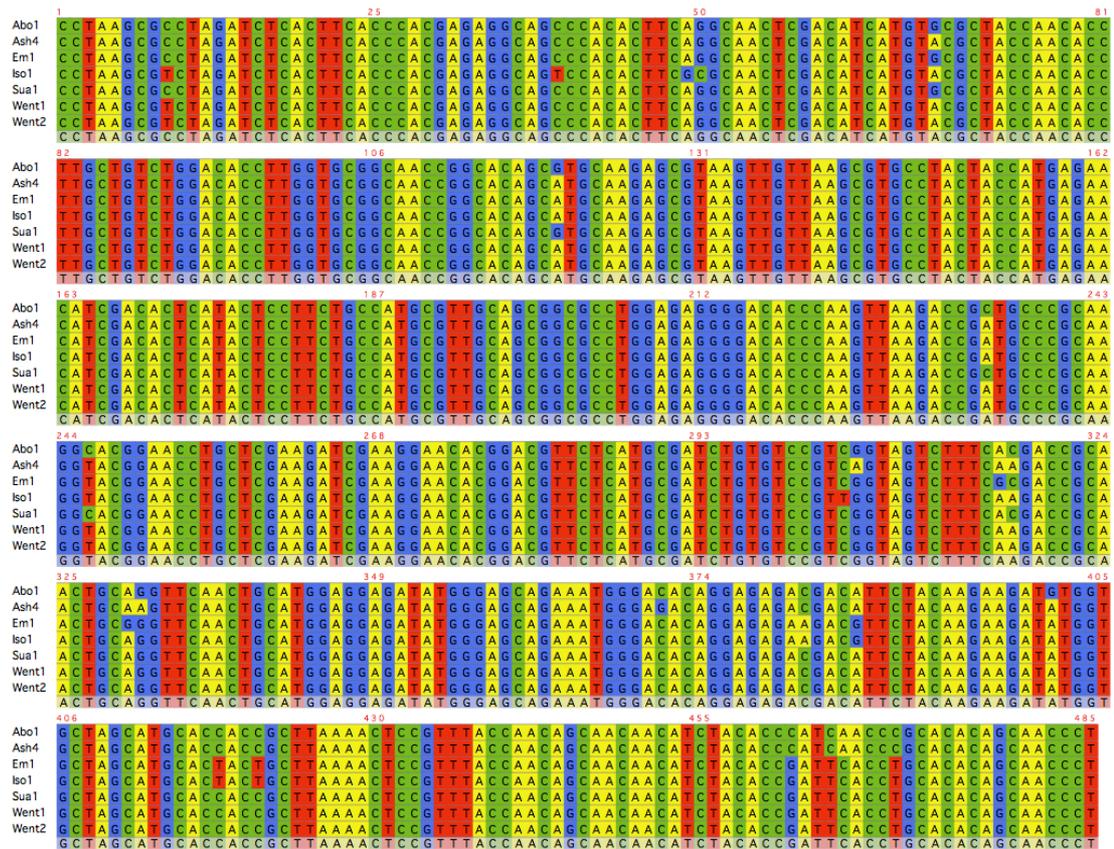
**Figure 3.1. *Albugo* sampling locations.** Arabidopsis plants displaying white rust symptoms were collected from the field. (a) Norwich, isolate Nc14. (b) East Malling, isolate Em1. (c) Ashbrittle, isolate Ash4. (d) Stratford upon-Avon, isolate Sua1. (e) Ashbourne, isolate Abo1. (f) Tuebingen region, isolates Iso1, Went1 and Went2.

### 3.2.2 MOLECULAR IDENTIFICATION OF NEW ISOLATES

To ascertain the species that each isolate belonged to I sequenced part of the rDNA internal transcribed spacer 1 (ITS1), previously used to distinguish *A/* and *Ac* (Thines et al., 2009). I found that each field isolate that was able to propagate (8 in total), has a 100% identical ITS1 sequence to the *A/* isolates Nc14 and Em1.

Next, to establish if each isolate was different from the others, I developed a new polymorphic DNA sequence marker. I found that the region of the genome harboring a gene called *RXLR10* in Nc14 is highly polymorphic between Nc14 and Em1, according to the data generated by Kemen et al (2011). It is part of a repetitive region of the genome. I designed primers to amplify part of this region, and sequenced each new isolate in turn as I purified it. Every new isolate contained novel polymorphism within this region. Therefore I could conclude that each isolate is unique, with one exception. The *At* plant from which the isolate Went1 was isolated was very heavily infected. "Isolate" Went2 was derived from the

sporulation on the upper tissues of the infected plant, and Went1 from the rosette leaves. In the first propagated generation both leaf and aerial derived infections had the same *RXLR10* sequence, indicating that they are probably the same. Therefore only Went1 was carried forward for purification. This marker was also used later to check the purity of each isolate when required. Figure 3.2 shows the nucleotide sequence of part of the *RXLR10* region in several isolates.



**Figure 3.2.** The nucleotide sequence of the *RXLR10* region in 7 isolates. The *RXLR10* region was amplified from each isolate and the product sequenced. There are unique polymorphisms in each isolate, except Went1 and Went2 which were derived from the same infected plant. Note that the Nc14 sequence is not in this alignment because the sequence contains a large indel polymorphism relative to the isolates shown.

### 3.2.3 PHENOTYPING OF NEW ISOLATES

I tested a panel of *At* accessions to investigate the natural diversity in the virulence of each of the newly collected isolates. I selected accessions on the basis that they were reported as resistant to Nc14 and/or Em1 in Kemen et al, (2011). 38 accessions were tested twice with all 4 of the 4 new purified *A/* isolates. Each accession was also re-tested with Nc14 and Em1. I found a diverse set of differential

virulence specificities (table 3.1). Only two accessions were resistant to each isolate (Sf-2 and Ts-1). The commonly used accessions Col-0, Ws-0 and Ws-2 are susceptible to all 6 isolates.

Isolate/Accession	<i>Nc14</i>	<i>Em1</i>	<i>Abo1</i>	<i>Sua1</i>	<i>Went1</i>	<i>Ash4</i>
As-77	R	S	S	S	S	S
BAT1	R	S	S	S	S	S
CIBC-5	S	R	S	R	R	R
Ei-2	S	R	R	R	S	S
Ekn 3	R	S	S	S	S	S
Fly2-1	R	S	S	S	R	S
Fly2-2	R	S	S	S	R	S
Fri2	R	S	S	S	S	S
Ge-0	S	R	S	R	R	R
GrA-5	R	S	S	S	S	S
Hov1-7	R	S	S	S	R	S
HR-10	R	R	S	S	S	R
HR-5	R	S	S	S	S	S
Kin-0	S	R	S	S	S	S
Kni-1	R	S	S	S	R	S
Knox-18	R	R	S	S	R	S
Ksk-1	R	R	S	R	R	R
NFA-10	R	R	R	R	S	R
Pna-17	S	R	S	S	R	S
Ren-11	R	S	S	S	R	S
Rev-3	R	S	S	S	R	S
RRS-7	S	R	R	R	R	S
S294BeL4	R	R	S	S	S	S
San-2	R	R	S	S	R	S
Sf-2	R	R	R	R	R	R
Sq-1	R	R	S	S	R	S
T1010	R	R	S	S	R	S
T1160	R	S	S	S	R	S
T450	R	R	S	S	S	S
T800	R	S	S	S	R	S
T860	R	S	S	S	R	S
TDr9	R	R	S	S	S	S
Ts-1	R	R	R	R	R	R
Ts-5	S	R	R	S	S	S
Udul 1-34	R	S	S	S	R	R
Uk-1	S	R	R	R	R	R
UIIA-1	R	S	S	S	R	S
UIIA-2	R	R	S	S	R	S

**Table 3.1. Phenotypic characterization of 38 *At* accessions with six *Al* isolates.** 5

week old *At* plants were inoculated with each isolate and scored resistant/susceptible (R/S) depending on pustule formation at 14 dpi.

### 3.3 DISCUSSION

Thines et al, (2009) established the distinction between *Al* and *Ac*, although examples of both species have been isolated from wild *At* plants. Given that *Al* is apparently a specialist on *At*, I hypothesized that a higher frequency of wild infections would be caused by *Al*. I purified 6 field isolates of *Albugo*-infected *At*

plants from geographically distant sites around the UK and one site in Germany. The sequences of 5 isolates' ITS regions revealed that they were all *AI* infections. This suggests that the specialist *AI* is common on *At* in the field. I cannot make any conclusion about the presence of *Ac* in the field isolates; *At* line Col-TH0 used for propagation contains a broad spectrum *Ac* resistance (*WRR4*; Borhan et al., 2008) so would have selected against most *Ac* isolates.

I discovered that the highly polymorphic *RXLR10* region could be used as a diagnostic marker to differentiate any of the various isolates. This region is highly repetitive and contains a large indel in Em1 (Kemen et al., 2011) and every other isolate relative to Nc14. Exactly why this region is so highly polymorphic has not been established but it is useful as a marker.

Previously it was established that out of 143 *At* accessions, 14% were resistant to Em1 and 10% resistant to Nc14 (7% to both) (Kemen et al., 2011). Thus the majority of accessions (~80%) are susceptible to both Nc14 and Em1. To save time and resources, accessions known to be resistant to either Nc14 or Em1 were used as the basis for the screening of my 4 new isolates. After this screen, it was revealed that the new isolates frequently overcame resistance and 14 unique differential groups of accessions were identified. This indicates that there are many different *Avr/R*-gene relationships underlying these different phenotypes. The two accessions resistant to all 6 isolates, Sf-2 and Ts-1, both originate from the San-Feliu region in eastern Spain (Horton et al., 2012). *RAC1* and *RAC3* are known to confer resistance to Em1 in the accession Ksk-1 (Borhan et al., 2004). Uniquely the isolate Abo1 can grow on Ksk-1, so it must overcome both of these resistances.

This collection and phenotyping of a new set of 4 *AI* isolates provides the basis for the exploration of the genetic components of virulence and resistance that will be discussed in the next 3 chapters.

## CHAPTER 4: COMPARATIVE GENOMIC ANALYSIS OF SIX *ALBUGO LAIBACHII* ISOLATES AND PREDICTION OF RECOGNISED EFFECTORS

### 4.1 INTRODUCTION

Prior to this study, no recognised effectors or *AVR* genes had been identified from *Al*. The predicted secretomes of the two sequenced *Albugo* sp. do not encode an elevated proportion of RxLR or Crinkler effectors (Kemen et al., 2011; Links et al., 2011). Further analyses did reveal a new class of over-represented potential effectors. These have a N-terminal Cys-His-x-Cys (CHxC) motif (Kemen et al., 2011) and a conserved Glycine. The genomes of the Peronosporalean oomycetes encode between 150 and 700 RxLR effectors but the *Al* genome encodes only 35 CHxC class proteins (Baxter et al., 2010; Raffaele et al., 2010a; Kemen et al., 2011). Considering that *Al* has ~900 proteins with a predicted secretion signal, it is likely that it has further effectors for which an N-terminal translocation motif is unknown.

The recognised effectors, that are also *AVR* genes, *At* recognised 1 (*ATR1*), *ATR13* and *ATR5* from *Hpa* were isolated using classical genetic approaches (Allen et al., 2004; Rentel et al., 2008; Bailey et al., 2011). The *P infestans* recognised effector *Avr3a* and several recognised effectors from *Magnaporthe oryzae* (rice blast fungus) were identified using association genetics (Armstrong et al., 2005; Yoshida et al., 2009). Recently the recognised effector *Ave1* from *Verticillium dahliae* (vascular wilt fungus) was identified using association genomics (de Jonge et al., 2012). It is therefore possible that, using genome sequence information from several isolates with differential virulence phenotypes, potential recognised effectors could be identified.

It is generally accepted that antagonistic co-evolution accelerates the rate of evolution of molecules determining the outcome of host-parasite interactions (Paterson et al., 2010). There are two (overlapping) models proposed for interactions between *R*-genes, *Avr* genes and the virulence targets of *Avr* gene products: the “arms race” model and the “trench warfare” model (Stukenbrock and McDonald, 2009). In the arms race model, novel adaptive mutation in any of the cognate interactors (i.e. a gain of recognition mutation in an *R*-gene) causes the

corresponding interactor to be “swept” from the population and be replaced by a new allele that can evade the new recognition. The trench warfare model is similar except it takes into account that adaptive changes to gain or loss of recognition may have a cost in terms of fitness. This leads to frequency-dependent selection (FDS) (Brown and Tellier, 2011). Sequences undergoing trench warfare-like selection typically show signatures of adaptive selection (enrichment of non-synonymous mutations).

DNA sequences undergoing trench warfare-like selection may display deviation from the neutral theory of molecular evolution (Kimura, 1968), depending on both the state of the population at the time of sampling and on how representative the sample is of diversity at a given locus. I hypothesise that genes undergoing FDS (eg *Avr* genes) will have high genetic diversity but a lower number of unique alleles than projected based on the neutral theory due to the balancing effect of FDS and selection against alleles intermediate between recognition (*AVR*) and evasion (*avr*). A number of tests to identify deviation from the neutral model using population samples have been devised. These include Tajima’s D and Fu’s  $F_s$  (Tajima, 1989; Fu, 1997). The allelic diversity of the recognised effectors from *Hpa* ATR13 and ATR39 is consistent with my hypothesis; they consist of a relatively small number (compared to the number of isolates) of highly divergent alleles within the known *Hpa* isolates (Hall et al., 2009; Goritschnig et al., 2012). There are other signatures that may be used to identify possible effectors, mostly discovered through the analysis of *Phytophthora* species. These include enrichment for non-synonymous changes in the C-terminus (Win et al., 2007) and the presence of many effectors in gene-sparse regions (Raffaele et al., 2010b), although *Al* appears to lack gene sparse regions (Kemen et al 2011).

In this chapter I set out to identify potential *AVR* genes in *Al*. I took a population genomics approach, comparing the genomes of six *Al* isolates, four of which were collected and purified in chapter 2. Using disease phenotype information and our expectations of the allelic diversity of *AVR* genes I made predictions of possible *AVR* genes to be tested in chapter 5. I also report additional findings made during the

analysis of these genomes, pertaining to potential effector evolution and sexual recombination between isolates.

## 4.2 RESULTS

### 4.2.1 GENOME SEQUENCING OF SIX *ALBUGO LAIBACHII* ISOLATES

In chapter 3 I described the collection, purification, identification and phenotyping of *Al* isolates collected from infected plants in the wild. The next step in my study was to fully genotype this population of *Al* isolates. I prepared high quality, high molecular weight genomic DNA specifically from the spores of each isolate. The four new isolates (Abo1, Ash4, Sua1, and Went1) were sequenced with Illumina technology to produce 76 bp paired end reads. Em1 that had previously been sequenced with 36 bp reads (Kemen et al., 2011) was re-sequenced with paired end 76 bp reads. For Nc14 I used the 76 bp reads generated by Kemen et al (2011) for my analysis. The reads were aligned to the published Nc14 genome. I assessed the alignments and concluded that sequencing was successful for each isolate and that the majority of reads were derived from *Al* DNA (table 4.1). Average sequencing depth of at least 60 was achieved for each isolate. Unaligned reads were extracted and used for de-novo assemblies for each isolate. The bulk of these contigs were clearly derived from *Arabidopsis thaliana*, with some contamination from *Pseudomonas* and *Xanthomonas* species.

Isolate	Total # reads	% Reads aligned	Avg depth of coverage	% genome >5 deep	% genome >1 deep	% genome >2x avg depth
Nc14*	7.06E+07	85.44	140.03	99.70	99.95	1.83
Em1**	3.57E+07	73.37	60.88	99.46	99.70	1.94
Abo1	5.45E+07	86.65	109.68	99.33	99.52	1.46
Ash4	5.41E+07	86.87	109.06	99.34	99.54	1.20
Sua1	7.00E+07	84.42	137.10	99.19	99.36	1.32
Went1	5.62E+07	92.71	120.99	99.43	99.63	0.95

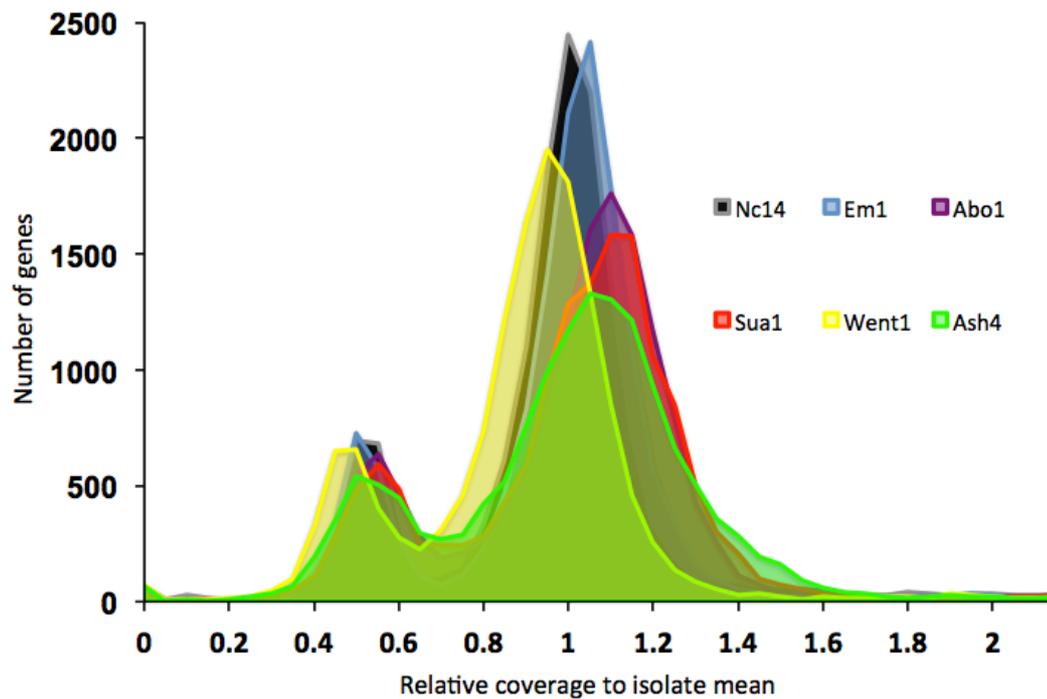
**Table 4.1 Summary of alignments of Illumina-generated 76 bp paired-end reads from each *Albugo laibachii* isolate against the Nc14 reference genome.** Total # reads indicates the raw output from the Illumina sequencing. These reads were aligned to the Nc14 genome, the % of which successfully aligned is shown the second column. The average depth of coverage, and the proportions of the genome with greater than 1 and 5 deep read coverage were calculated using custom Perl scripts. The % of the genome >2x avg depth was calculated to highlight potential copy number variation. \*The Nc14 reads were generated by Kemen et al (2011). \*\*The DNA and Illumina library for Em1 were generated by Kemen et al (2011) for 36 bp sequencing. In this study the same library was sequenced again using 76 bp technology.

#### 4.2.2 POLYMORPHISM IN SIX *ALBUGO LAIBACHII* ISOLATES

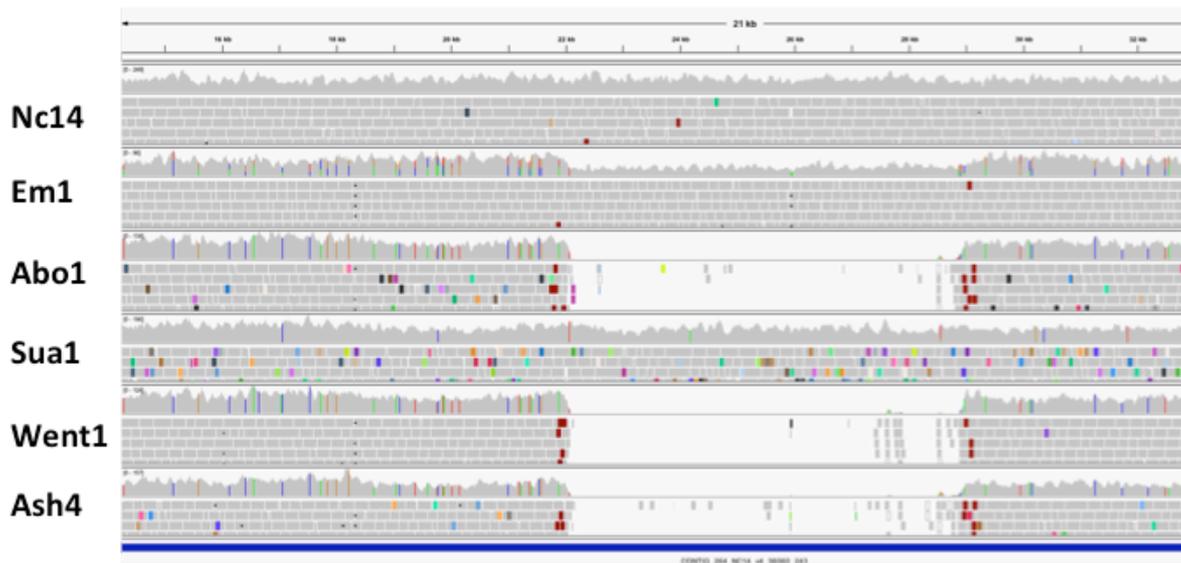
Using the alignments to the Nc14 reference, I predicted polymorphisms, including single nucleotide polymorphisms (SNPs) and small insertion/deletions (Indels). Overall, I detected a relatively low level of polymorphism between these isolates (table 4.2). For example Sua1 had the most polymorphisms compared to Nc14: 41772 polymorphic positions or 1.25 SNPs/kb. I analysed the overall similarity between the polymorphisms predicted for each isolate, in order to assess the overall similarity of each isolate to each other (table 4.3). From this analysis it seems that Sua1 and Went1 share the most polymorphisms. Em1 seems to share the least number of polymorphisms with the other isolates. Note that Nc14 cannot be meaningfully compared because it of course has very few homozygous polymorphisms to itself (these are caused by minor errors in the original assembly). I also examined the alignments for large indels and for copy-number variation (CNV). To summarise, the vast majority of genes show the same coverage in each isolate (figure 4.1). This includes a substantial number of genes (~1000) located in apparently hemizygous regions (containing only one copy as opposed to the two expected in a diploid) of the genome. However there are several exceptions, detailed in Table 4.3. Of note, there is an ~6.5 kb segment of contig 264 containing three genes that has normal coverage in Nc14 and Sua1, no coverage in Abo1, Went1 and Ash4 and ~50% coverage in Em1 (figure 4.2). The polymorphisms bordering this region appear to be heterozygous only in Em1 (figure 4.2). My analysis also revealed many short (mostly <1 kb) contigs apparently unique to Nc14. Em1 has a higher proportion of heterozygous to homozygous polymorphisms than other isolates.

Isolate	Polymorphisms	Homozygous polymorphisms	Heterozygous polymorphisms	Single nucleotide polymorphisms	Small indel polymorphisms
Nc14	21545	6850	14695	16179	5366
Em1	36029	13726	<b>22303</b>	32415	3614
Abo1	39112	24051	15061	34177	4935
Ash4	40759	24415	16344	35390	5369
Sua1	<b>41772</b>	<b>25995</b>	15777	35898	5874
Went1	39360	24300	15060	34646	4714

**Table 4.2 Summary of polymorphism across 6 isolates of *Albugo laibachii*.** Polymorphisms were predicted based on the alignment of Illumina reads from each isolate to the Nc14 reference sequence as described in Chapter 2. Homozygous, heterozygous, single nucleotide and small indels were differentiated using custom Perl scripts.



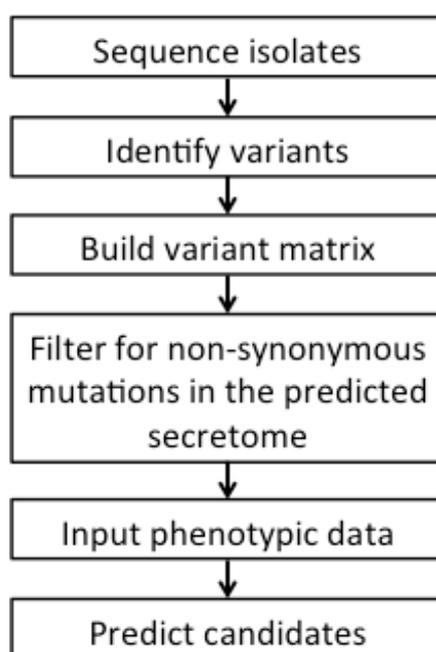
**Figure 4.1. Distribution of the average coverage of genes in six *AI* isolates.** The relative depth of coverage for each gene in each isolate was extracted and the distribution plotted with a bin size of 0.05. In each isolate, ~700 genes occur with half coverage (0.4 - 0.6 x average coverage).



**Figure 4.2. Visualization of the reads from the six *AI* isolates aligned to the Nc14 genome.** Alignments of the Illumina reads from each isolate against the Nc14 reference are visualized in the IGV genome browser. The panel corresponding to each isolate is split into two parts, the upper showing the overall coverage at each position and the lower showing actual Illumina reads. Colored vertical lines represent discrepancies to the reference (ie SNPs). A 21 kb region of contig 264 is shown, which displays an apparent deletion in three isolates and 50% coverage in another (Em1).

#### 4.2.3 ASSOCIATION GENOMICS IN SIX *ALBUGO LAIBACHII* ISOLATES

To identify AVR gene candidates from the population-genomics data, I designed a computational pipeline to correlate non-synonymous polymorphisms across the secretomes with the virulence differentials reported in chapter 3. The principle of the pipeline is summarised in figure 4.3, and detailed in chapter 2b. The outcome is a list of candidate secreted proteins ranked by the number of differential associated non-synonymous polymorphisms including pseudogenisations and predicted changes to intron/exon structure, for each unique differential pattern. Appendix table 4A1 is a re-organised table structured around the fourteen observed differential virulence groups.



**Figure 4.3. A simplified representation of the pipeline to associate non-synonymous polymorphisms with isolate virulence differentials.** The development of the pipeline is discussed in detail in chapter 2b.

I applied the association genomics pipeline to the polymorphisms predicted from the six *Al* isolates, producing a list of candidate genes based around the premise that there should be a pattern of mutation consistent with the recognition of certain alleles within the population. Some of these data are presented in table 4.4, and fully in appendix table 4A2. At least one candidate was predicted for each group of accessions. In some cases, genes were found that had mutations encoding up to 33 non-synonymous mutations fully correlated with a particular phenotype, for

example AINc14C169G7963 (table 4.4) that encodes a short (less than 500 amino acids) secreted protein (abbreviated to SSP17), possibly recognised by accession Ksk-1.

HR-10			
Gene	Corelated NS polymorphisms	Annotation (automatic)	Notes
AINc14C103G6107	26	unknown	hemizygous, SSP
AINc14C65G4610	24	unknown	SSP24
AINc14C303G10412	3	unknown	RxL14
AINc14C43G3600	3	SSP6	

HR-5, Ren-11, As-77, BAT1, Fri2, EkN 3, GrA-5			
Gene	Corelated NS polymorphisms	Annotation (automatic)	Notes
AINc14C28G2718	6	unknown	SSP16
AINc14C260G9793	4	unknown	
AINc14C365G11050	4	unknown	
AINc14C142G7291	3	unknown	SSP27

Ksk-1			
Gene	Corelated NS polymorphisms	Annotation (automatic)	Notes
AINc14C169G7963	33	unknown	SSP17
AINc14C56G4264	22	unknown	SSP18
AINc14C35G3154	14	unknown	SSP21
AINc14C163G7833	5	unknown	
AINc14C56G4265	5	unknown	
AINc14C64G4566	4	unknown	elicitin-like

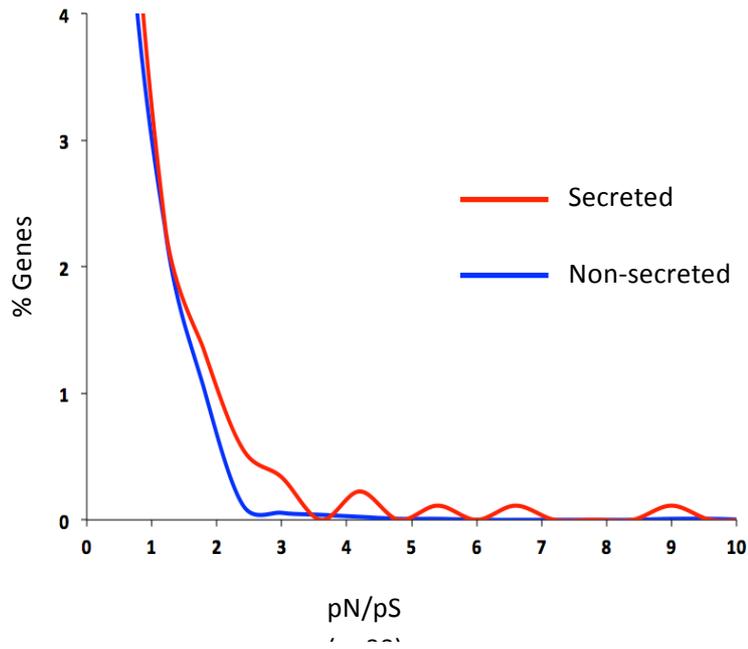
**Table 4.4 Some of the predictions from the association genomics pipeline.** The annotation (automatic) refers to the annotation assigned to these genes by Kemen et al (2011) by methods described in that paper. During this study, the sequences of many potential candidates were examined manually; the outcome of these searches is included in the notes column. SSPs are genes encoding a predicted signal peptide and less than 500 amino acids.

#### 4.2.4 POPULATION-GENETIC ANALYSIS OF SIX ALBUGO LAIBACHII ISOLATES

As described in the introduction, true recognised effectors are postulated to show signatures of selection such as adaptive selection (enrichment for non-synonymous polymorphisms) and balancing selection (selection to maintain divergent alleles). Because of a potential concern that false-positives and artefacts might populate the lists generated by the association genomic pipeline, and in order to produce candidates worthy of testing for recognition, statistical values for these two types of selection were calculated for all of the genes in *Al* using the polymorphism data from the six sequenced isolates. Using the PAML suite and the yn00 method (Yang and Nielsen, 2000), pN/pS values (proportion non-synonymous to synonymous polymorphisms; adaptive selection) were calculated. A distribution of the yn00

pN/pS scores is shown in figure 4.4. The predicted secretome shows a shift towards higher pN/pS values in comparison with the genes that do not encode proteins with predicted signal peptides, which is statistically significant (unpaired Student's t-test,  $p < 0.01$ ). I extracted the top 20 secreted and non-secreted protein-encoding genes, tables 4.5 and 4A4 respectively. To measure balancing selection, two similar statistical tests were used; Fu's  $F_s$  (Fu, 1997) and Tajima's  $D$  (Tajima, 1989). These methods test for balancing selection by assessing the polymorphism at a locus, generating an expected allele number based on this polymorphism and comparing it to the observed number of alleles. Genes with a lower than expected allele number, despite high levels of polymorphism, will thus have a higher than average Tajima's  $D$  Fu's  $F_s$  score. Fu's  $F_s$  is predicted to be a more sensitive test. Fu's  $F_s$  and Tajima's  $D$  were calculated for all *Al* genes using DNAsp (Librado et al., 2009). The values for Fu's  $F_s$  are plotted as a distribution in figure 4.5 and the top 20 secreted and non-secreted protein encoding genes, are in tables 4.6 and 4A5 respectively. As with the pN/pS, the secretome has a significantly different distribution of Fu's  $F_s$ ; there is a noticeable shoulder present only in the secretome towards positive Fu's  $F_s$  scores (figure 4.5.). The Tajima's  $D$  statistic did not show the same trend as pN/pS and Fu's  $F_s$ . The graph and tables for Tajima's  $D$  can be found in appendix figure 4A1 and table 4A3 respectively.

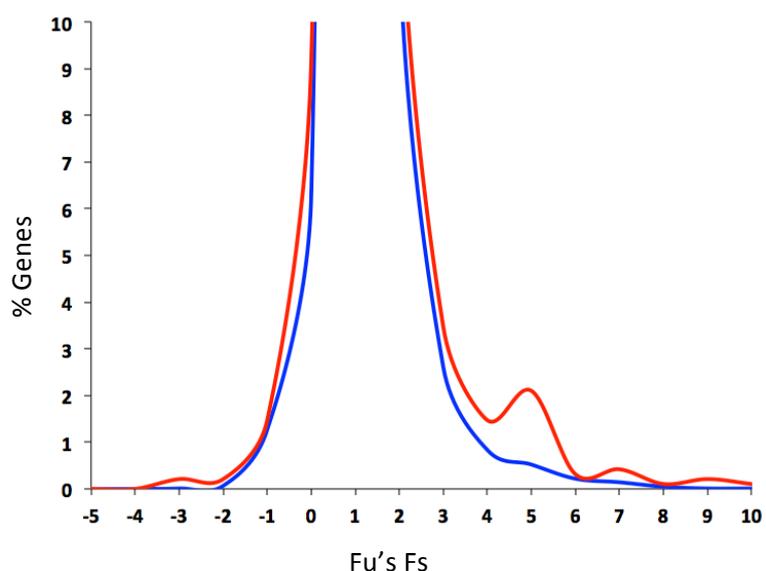
Since those genes with significant pN/pS and Fu's  $F_s$  are hypothetically more likely to be recognised effectors, it was possible to use this information to prioritise the list of candidate genes from the association genomics pipeline. Tables were created to incorporate each statistic, and the resulting top candidates are shown in table 4.7.



**Figure 4.4. Distribution of pN/pS for all AI genes, generated from the analysis of the 6 isolates.** pN/pS (yn00 max method) was calculated for each gene based on the Illumina data from six isolates as described in Chapter 2. Genes encoding proteins with a predicted signal peptide are in red, and those without in blue.

Contig	Gene	pN/pS yn00 max	Annotation
28	2718	15.4579	SSP16
35	3154	8.6876	SSP21
69	4807	6.0658	CHXC2
169	7963	5.0858	SSP17
61	4450	3.9738	glycoside hydrolase
325	10635	3.7633	SSP19
6	824	2.8225	SSP20
28	2684	2.755	glycoside hydrolase
177	8157	2.6702	kazal protease inhibitor putative
236	9385	2.1855	DEAD BOX RNA helicase
56	4265	2.1205	similar to SSP18
125	6784	2.0969	conserved hypothetical
212	8939	2.0854	serine protease putative
163	7833	1.8977	unknown
65	4610	1.7618	SSP24
31	2884	1.7399	conserved hypothetical protein
301	10374	1.6445	PREDICTED: hypothetical protein
73	4960	1.627	unknown
412	11464	1.6133	unknown
176	8132	1.5392	unknown

**Table 4.5. The top 20 genes encoding secreted proteins by pN/pS (yn00) statistic.** pN/pS (yn00 max method) was calculated for each gene based on the Illumina data from six isolates as described in Chapter 2.



**Figure 4.5. Distribution of Fu's Fs for all AI genes, generated from the analysis of the 6 isolates.** Fu's Fs was calculated for each gene using the VariTale method as described in Chapter 2. Genes encoding proteins with a predicted signal peptide are in red, and those without in blue.

Contig	Gene	Fu's Fs	Annotation
28	2718	11.275	unknown SSP16
18	1839	9.949	SSP24
35	3154	8.872	kazal protease inhibitor putative
169	7963	8.463	unknown
56	4264	7.059	unknown
177	8157	6.991	unknown
205	8798	6.991	conserved hypothetical protein
361	10994	6.991	unknown
69	4807	6.215	CHXC2
176	8132	5.498	unknown
313	13412	5.498	CHXC29
28	2684	5.16	glycoside hydrolase putative
205	8792	4.647	hypothetical protein UM06115.1
453	11745	4.647	nd
61	4472	4.548	beta-glucan synthesis-associated protein putative
65	4610	4.351	unknown
46	3742	4.326	inositol-3 putative
264	9852	4.278	unknown
143	7313	4.24	unknown
262	9825	4.23	unknown

**Table 4.6 The top 20 genes encoding secreted proteins by Fu's Fs statistic.**

Fu's Fs was calculated for each gene using the VariTale method as described in Chapter 2.

Best candidates (arbitrary order)

Gene	Corellated NS polymorphisms	Accession(s)	Fu's Fs	pN/pS	Annotation (automatic)	Notes
AINc14C169G7963	33	Ksk-1	8.463	5.0858	unknown	SSP17
AINc14C56G4264	22	Ksk-1	7.059	infinite	unknown	SSP18
AINc14C35G3154	14	Ksk-1	8.872	8.6876	unknown	SSP21
AINc14C65G4610	24	HR-10	4.351	1.7618	unknown	SSP24
AINc14C28G2718	6	HR-5, Ren-11, As-77, BAT1, Fri2, EKN 3, GrA-5	11.275	15.4579	unknown SSP16	SSP16
AINc14C28G2718	21	Knox-18, San-2, T1010, ULLA-2	11.275	15.4579	unknown SSP16	SSP16
AINc14C18G1839	12	Knox-18, San-2, T1010, ULLA-2	9.949	infinite	unknown	possible aspartate protease
AINc14C361G10994	7	Knox-18, San-2, T1010, ULLA-2	6.991	infinite	unknown	CxHC2

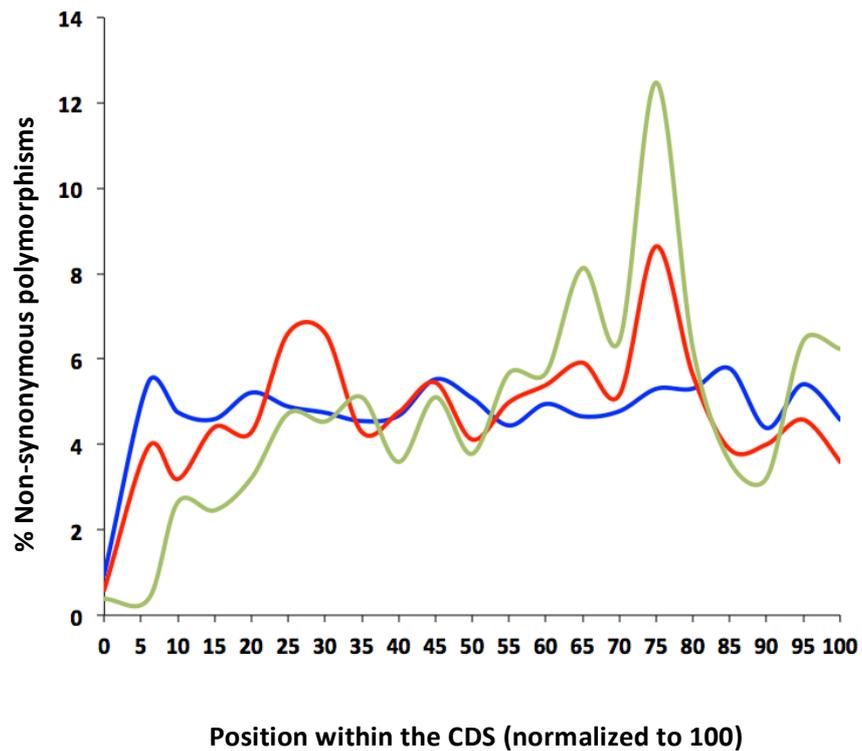
**Table 4.7 The best candidates for recognized effectors after considering association genomics, Fu's Fs and pN/pS.**

In addition to using the polymorphism data to prioritise recognised effector candidates, I also thought it pertinent to examine several other aspects of the polymorphism data, for example to check if there is any enrichment for non-synonymous polymorphism in a particular region of secreted protein encoding genes, or if the genomic context of a gene could effect its rate of adaptive evolution.

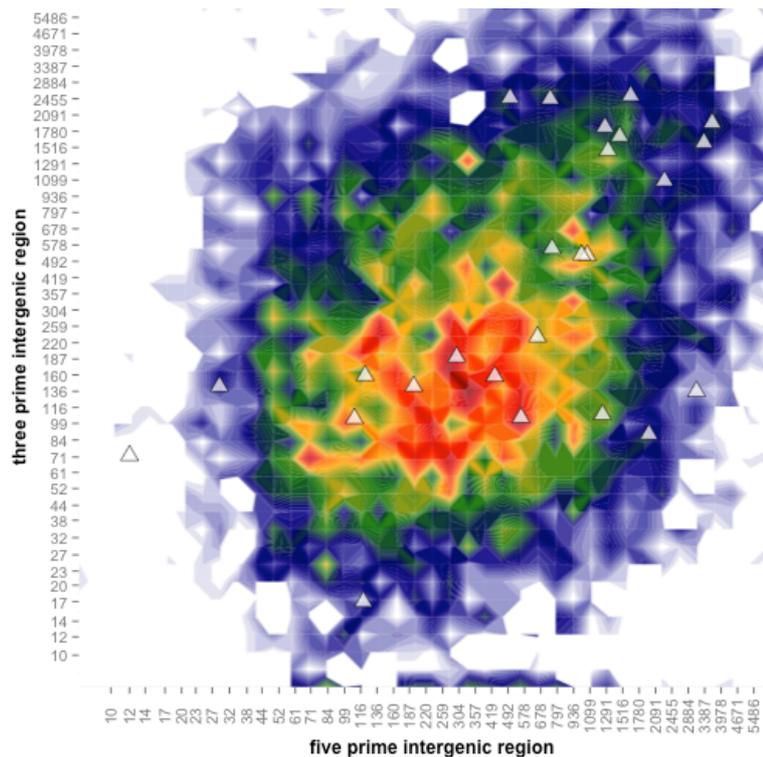
By extracting the coordinate of each non-synonymous polymorphism within its respective CDS, and normalising this to the CDS length, I was able to generate figure 4.6, summarising this information for three groups: secreted, non-secreted and finally those genes with secreted products that have both high Fu's Fs and pN/pS ('rapidly evolving' genes; present in tables 4.5 and 4.6). The distribution of mutations in the C-terminus of both the secreted and especially the 'rapidly evolving' group of secreted protein encoding genes suggests that this region is undergoing adaptive selection at an enhanced rate. These genes also seem to have less non-synonymous polymorphisms than average in the first 15-20% of the protein (figure 4.6). Of 1725 unique non-synonymous polymorphisms detected in the secretome, 529 of them occur in these 31 genes.

In order to assess the effect of genomic context, defined by the length of a gene's 5' and 3' intergenic distances to neighbouring ORFs, on the rate of evolution in genes I employed a density plot method described by Saunders et al (2014). Figure 4.7 shows that the vast majority of genes lie in gene-dense regions with intergenic distances between 100 and 1000 bp. The white triangles represent the 31

rapidly evolving secreted protein encoding genes, where they could be calculated (some genes could not be plotted because they are at contig ends). Although a number of them are found in the 'normal' regions, one-third have longer than average 5' and 3' distances.



**Figure 4.6. The frequency of non-synonymous polymorphisms across the normalized CDSs of AI genes.** Genes encoding proteins with a predicted signal peptide are in red, those without in blue, and those in pastel yellow are 'rapidly evolving' genes with a predicted signal peptide.



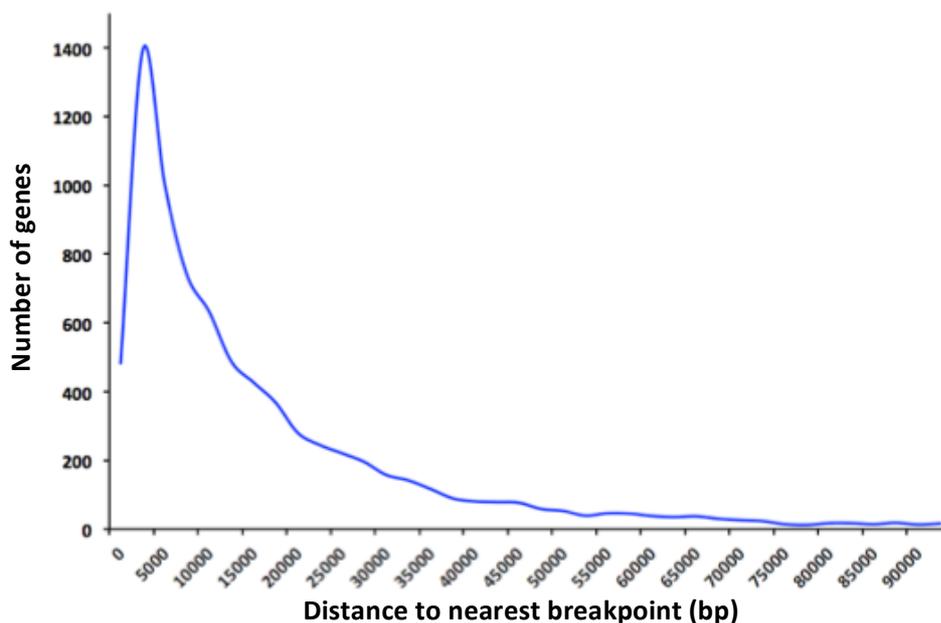
**Figure 4.7. A density plot of 5' and 3' intergenic distances, with rapidly evolving secreted protein-encoding genes highlighted with white triangles.** The axes show the 5' and 3' distance in bp from the start and stop codon of each gene respectively. The density of genes is represented by the shade of colour, pale blue being the least and red the most dense (ie the most genes). The 31 rapidly evolving genes were mapped onto the plot, indicated by the white triangles.

#### 4.2.5 ANALYSIS OF RECOMBINATION IN SIX *ALBUGO LAIBACHII* ISOLATES

In order to assess the likelihood that *Al* isolates reproduce sexually in nature, I identified putative sexual recombination events in the six sequenced isolates. To do this, I reconstructed genomes for each of the isolates, using only contigs greater than 10 kb in length and masking the first and last 500 bp from these contigs. With the assistance of Dr Mark McMullan I then used RDP (Martin and Rybicki, 2000) to predict recombination events. RDP uses multiple algorithms to scan aligned sequences from each isolate to identify blocks of common polymorphic sites between 2 or more isolates that may represent DNA introgressed via recombination. We found 474 putative events that met the threshold of the various algorithms employed. Table 4.8 shows the number of these events that occurred in each isolate. I also assessed the distribution of the distances of genes from recombination breakpoints (most commonly ~5 kb) shown in figure 4.8.

Isolate	# Predicted introgressions
Ash4	139
Sua1	130
Abo1	126
Went1	119
Nc14	114
Em1	111

**Table 4.8. The number of predicted introgressions detected in each isolate.** Contigs greater than 10 kb were reconstructed for each isolate, aligned and submitted to RDP (Martin and Rybicki, 2000).



**Figure 4.8. The distribution of gene distances to their nearest predicted recombination breakpoint.** The distance from each gene (either from 5' or 3' end) to its closest introgressed region was calculated and a distribution plotted.

To test if recombination event detection is an artefact of more polymorphic regions or related to adaptive evolution, I made plots of these values and the distance of each individual gene to its closest recombination breakpoint (appendix figures 4A3 and 4A4). However, I was unable to find any significant correlation, positive or negative with  $pN/pS$ , raw polymorphism or indeed  $F_u$ 's  $F_s$ .

#### 4.3 DISCUSSION

Comparison of the six *Al* isolates reveals a striking lack of nucleotide diversity. With an average of ~0.1% overall polymorphism between the genomes of any pair of isolates, these isolates are very similar to each other. For comparison, a similar study carried out on seven *Hpa* isolates found that they were each around 0.25% polymorphic (Ishaque, Furzer et al., *in preparation*) and on several isolates of the closely related *Ac*, 1% polymorphism was observed, though these were isolates seemingly specifically adapted to different host-ranges (Gardiner et al., *in submission*). While this level of polymorphism may be lower than within some fungal 'clonal lineages' (e.g. Cantu et al., 2013), it seems that each isolate studied here has both its own specific pattern of differentials (chapter 3) and pattern of unique polymorphisms and in particular, a unique mosaic of putative recombination events. Therefore it is likely that these are not clonal lineages but the product of sexual recombination occurring between closely related isolates in nature.

On the other hand, low diversity should mean there will be less background when attempting to identify *AVR* genes through association of phenotype with genetic polymorphism. Using a simple pipeline, *AVR* gene candidates were identified on the basis of the phenotypes observed in chapter 3. I demonstrated in chapter 2b how it was possible to use this pipeline to re-identify *ATR1* from *Hpa* using similar genotype/phenotype data. For some accessions with differential recognition I could not predict any strong *AVR* candidates (table 4AT3). I hypothesise that in these cases there may be either confounding mutations or expression level polymorphisms (Qutob et al., 2013) that allow isolates to gain virulence. It is also likely that in some cases the presence of multiple resistance genes in different accessions could confound my analysis. Other confounding effects, for example a tri-allelic *AVR* with multiple different alleles that are able to evade recognition, could cause problems for my association genomic system. Such confounding mutations are common in other association genomic experiments, which have become advanced in the human field, and underline the importance of the parallel population genomic analyses (Mathieson and McVean, 2012).

Therefore I examined the gene complement through the lens of several population genetics statistics. Using these data I categorised 31 genes that had both

a high pN/pS (adaptive selection) and Fu's  $F_s$  (balancing selection) as 'rapidly evolving' genes (genes present in both table 4.5 and 4.6). Previous analyses revealed that many rapidly evolving RXLR effectors from *Phytophthora* spp. had particularly rapidly evolving C-termini (Win et al., 2007) and typically appeared in gene sparse regions (Raffaele et al., 2010b). Examining both where predicted non-synonymous polymorphisms are located within these rapidly evolving genes (figure 4.6) and the genomic context of these genes (figure 4.7) strengthened the argument that some of them may be effectors and/or AVR genes, since some of them occur in gene sparse regions, and as a group they show a stronger tendency to have non-synonymous polymorphisms towards the C-terminus than the secretome as a whole.

Using both the association genetics data and population analyses, I drew up a short list of candidates to test for recognition in the laboratory (table 4.7). This list contains six SSPs and one CHXC-type effector candidate (though interestingly in this case the position of the H has moved over; the motif still fits with the generalised CHXC motif). Although all of these genes are considered equal in their probability of being AVRs, the CHXC protein, the secreted aspartate protease and AINc14C28G2718 (*SSP16*) (possibly recognised in multiple accessions) seem intriguing candidates. There are also three strong candidates for recognition in Ksk-1, an accession that harbours two known resistances: the cloned *RAC1* gene and the *RAC3* locus (Borhan et al., 2004). I will explore if any of these candidates is recognised in chapter 5.

Additionally I discovered many potential recombination events might have occurred in the lineages of these isolates. It is likely that the number of recombination events is under-estimated due to the low level of polymorphism between the isolates; more polymorphism would increase the resolution of the software. Nevertheless, this finding is consistent with the findings of Adhikari et al (2003) in *Ac* and the fact that infected plants produce many oospores (Thines et al., 2009). The finding highlighted in figure 4.2 is a second data point that suggests sexual recombination is occurring. One interpretation of the finding that Went1, isolated in Germany, has an approximately equal number of predicted

recombination events (and is generally as similar as any of the other isolates in terms of polymorphism) is that the spores of *Al* are extremely mobile and probably capable of crossing seas and continents fairly quickly. Alternatively, the rate of mutation and recombination might just be very low within this species. A more wide ranging collection of isolates is needed to test either of these hypotheses. It has been noted, for example in *Drosophila*, that recombination has a positive influence on diversity (McGaugh et al., 2012). The role of recombination in this case was not clear however, as no statistically significant link between my calculations of selection and the putative recombination events could be made. However, conceivably further analysis looking at the data from a different angle (ie recombination event density rather than proximity) could reveal a significant correlation between recombination events and an increased rate of evolution.

## CHAPTER 5: TESTING AND CHARACTERISATION OF POTENTIAL RECOGNISED EFFECTORS

### 5.1 INTRODUCTION

In order to parasitize their hosts, biotrophic filamentous plant pathogens secrete an arsenal of effector molecules from the specialised host-cell interface structure called the haustorium (Kemen et al., 2005; Whisson et al., 2007). Some of these effectors pass through the host cell's membrane from the extra-haustorial matrix to alter host cellular processes to favour the pathogen (Schornack et al., 2009). Many effectors were initially identified through their property of avirulence: the triggering of strong host immunity via resistance (*R*)- gene products. Recognition of avirulence (*Avr*) gene products, and therefore resistance to oomycete plant pathogens, is typically conferred by either TIR- or CC–NB-LRR class receptor proteins and is known as effector triggered immunity (ETI) (Jones and Dangl, 2006). ETI is often manifested as a localised hypersensitive response (HR) and cell death at the site of the infection.

In order to further understanding of the *Al-At* interaction I set out to identify recognised effectors, or *AVR* genes. In chapter 4 I presented the sequencing and comparison of six isolates of *Al*. Using disease phenotype information and expectations of the allelic diversity of *AVR* genes I produced a list of candidate effectors to test for recognition by specific *At* accessions.

Such testing requires a robust phenotyping assay. Previous studies have used high-throughput screens in surrogate systems such as *Agrobacterium*-mediated *Nicotiana benthamiana* transient transformation (Oh et al., 2009). This however requires both prior knowledge of the *R* gene and an *R* gene that functions in the surrogate system.

The “GUS-eclipse” assay is based on the detection of HR through the lack of GUS ( $\beta$ -glucuronidase) activity dependent staining subsequent to transient biolistic co-expression of GUS with an *Avr* or *R*-gene in a plant containing the cognate protein. Thus it is a system that can be employed in resistant accessions without *a priori* knowledge of a specific *R* gene. The method was pioneered in the context of plant

immunity by Mindrinos et al, (1994) and later used to detect HR caused by the expression of *Hpa AVR* genes ATR13 and ATR1 in resistant *At* accessions (Allen et al., 2004; Rehmany et al., 2005), as a means to identify these recognized effectors.

In this chapter I present my results from the testing of candidate effectors from chapter 4 for recognition in certain *At* accessions. I report that two candidates are recognised in specific *At* genotypes, and provide some details of the diversity within *A. laibachii* of these candidate effector proteins, their homologs and paralogs and characterise their sub-cellular localisation.

## 5.2 RESULTS

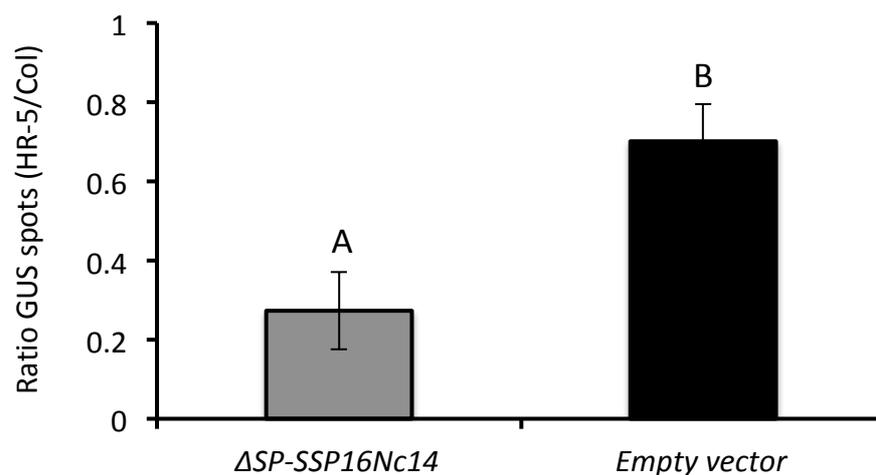
### 5.2.1 *SSP16<sup>Nc14</sup>* IS RECOGNISED BY HR-5

To test if defined candidates are recognised by specific *At* accessions, I used the “GUS eclipse” assay based on transient expression via DNA coated gold particle bombardment of detached *At* leaves (Mindrinos et al., 1994). By bombarding two *At* leaves, one susceptible accession and one resistant accession, simultaneously with gold particles coated with *GUS* and *Candidate gene driven by 35S promoter*, it is possible to screen for recognition through reduction or eclipse of GUS staining in a specific accession.

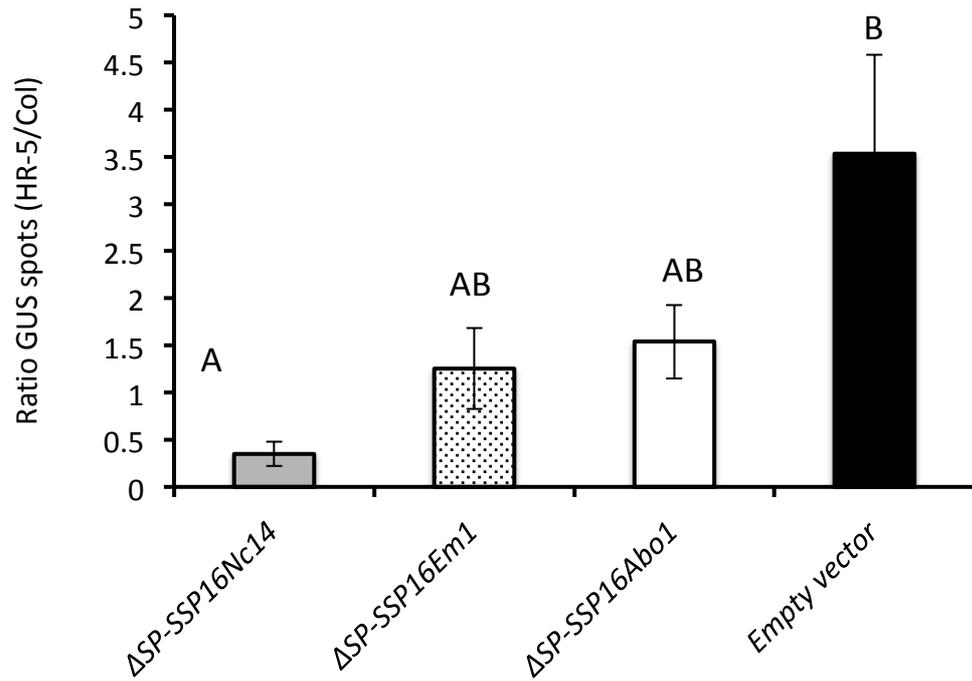
*AINc14C28G2718*, *SSP16<sup>Nc14</sup>*, was one of the most highly ranked candidates from my analyses in chapter 4. It is predicted to be recognised by the accessions HR-5, Ren-11 and Knox-18. It also has the highest Fu’s Fs and the highest (scorable) pN/pS of all genes encoding predicted secreted proteins (11.3 and 15.5 respectively). These characteristics meant that it was a high priority to be tested in a transient assay.  $\Delta SP$ -*SSP16<sup>Nc14</sup>* was cloned from Nc14 cDNA, and its UTRs confirmed by rapid amplification of cDNA ends (RACE)-PCR (Borson, Salo and Drewes, 1992). I bombarded the  $\Delta SP$ -*SSP16<sup>Nc14</sup>* (secretion signal truncated) allele under the 35S promoter into HR-5, with an empty-vector (EV) control. This combination gives a significant reduction of GUS spot ratio (figure 5.1) (Wilcoxon rank-sum test;  $p < 0.05$ ). I then cloned two further *SSP16* alleles, those from Em1 and Abo1 (the allelic differences are reported in section 5.2.3), predicted to evade recognition by HR-5. The *SSP16* alleles from Nc14, Em1 and Abo1 were then

bombarded within the same experiment, including an EV control. The two ‘virulent’ alleles gave an intermediate GUS spot ratio, whilst  $\Delta SP-SSP16^{Nc14}$  caused a significant reduction in ratio, compared to the empty vector (EV) (figure 5.2) (Wilcoxon rank-sum test;  $p < 0.05$ ).

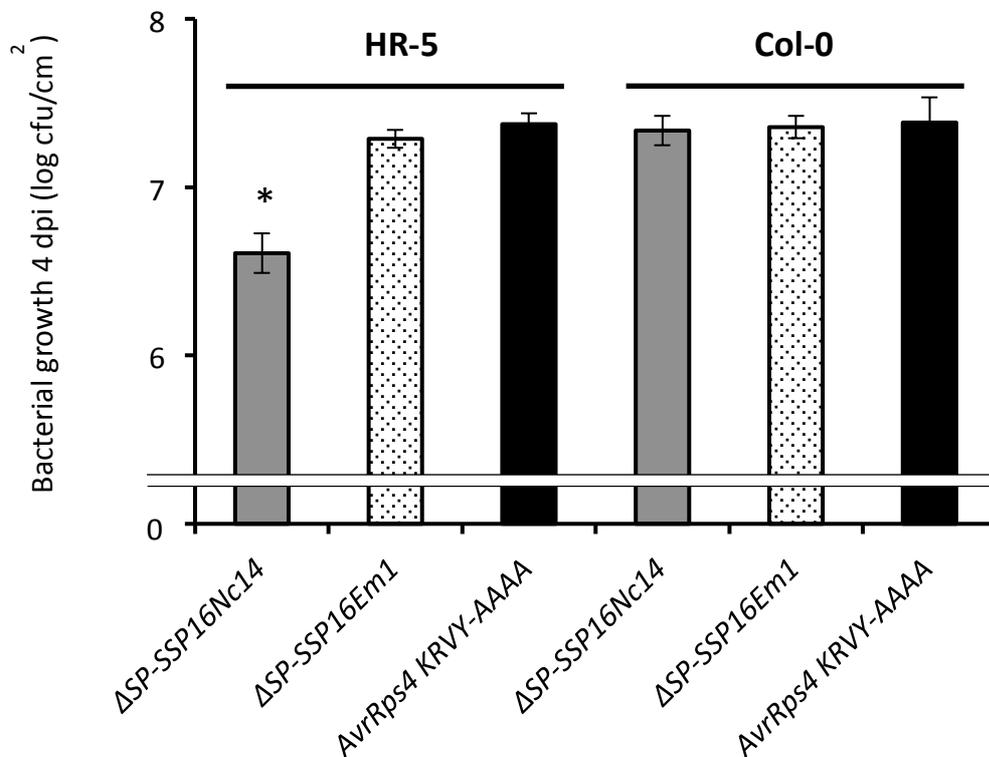
As a second assay to test if this apparent recognition is authentic, I employed the effector-detector-vector system (EDV) (Sohn et al., 2007) that uses a delivery sequence from AvrRps4. The secretion-signal truncated Nc14 and Abo1 alleles of *SSP16* were cloned into the EDV6 vector, and these were mated into *Pseudomonas syringae* DC3000 lux strain (Fan et al., 2008). I spray-inoculated HR-5 and Col-0 (susceptible) plants and assessed the bacterial growth 4 days post-inoculation (dpi) through colony counts. Exclusively in the HR-5 accession, the *SSP16^{Nc14}* allele showed significantly less growth (figure 5.3) (Student’s t-test on mean cfu/cm<sup>2</sup>,  $p < 0.05$ ) of between 0.5 and 1 log cfu/cm<sup>2</sup> (over 3 replicate experiments). The *SSP16^{Abo1}* allele was indistinguishable from the *AvrRps4^{KRVY-AAAA}* control (a non-functional effector mutant; Sohn et al., 2009) in both HR-5 plants and Col-0, while the *SSP16^{Nc14}* allele was also indistinguishable from both in Col-0. There was no detectable positive change in the growth of either allele in the susceptible Col-0 plants.



**Figure 5.1.  $\Delta SP-SSP16^{Nc14}$  is recognised by accession HR-5; GUS spot ratios for the bombardment of  $\Delta SP-SSP16^{Nc14}$  and empty vector in HR-5.** 6 pairs of leaves were bombarded with each construct. The letters signify the significance of the differentiation of the groups according to whether  $p < 0.05$  in a two-tailed Wilcoxon rank-sum test. The error bars represent standard error.



**Figure 5.2. GUS spot ratios for the bombardment of various SSP16 alleles in HR-5.** 6 pairs of leaves were bombarded with each construct. The letters signify the significance of the differentiation of the groups according to whether  $p < 0.05$  in a two-tailed Wilcoxon rank-sum test. The error bars represent standard error.

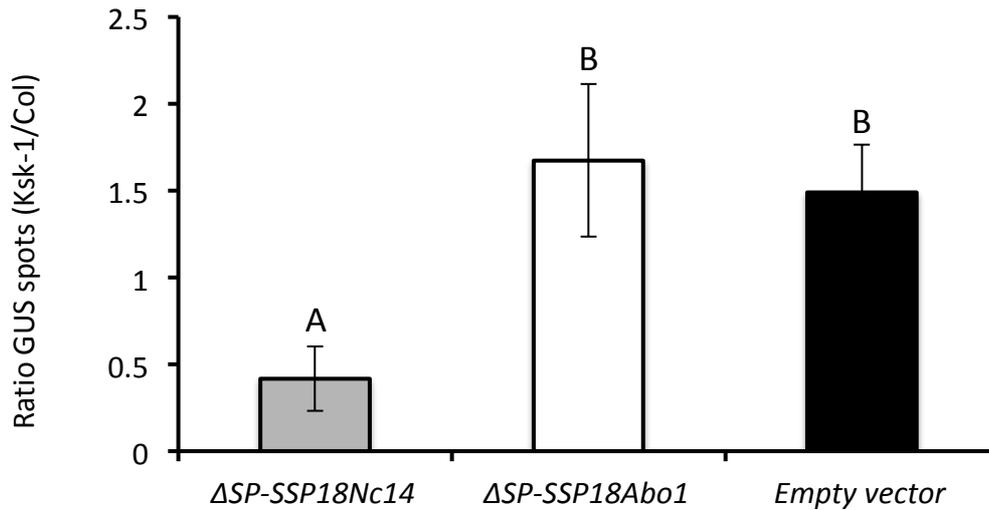


**Figure 5.3. SSP16Nc14 is recognised specifically by accession HR-5 when delivered by T3SS; Pst DC3000 lux growth curves.** Pst DC3000 lux was transformed with various EDV6 constructs and the strains were assessed for growth in planta at 4 dpi in 6 leaf discs per strain from HR-5 and Col-0. Significance indicated by pairwise Student's t-test  $p < 0.05$ .

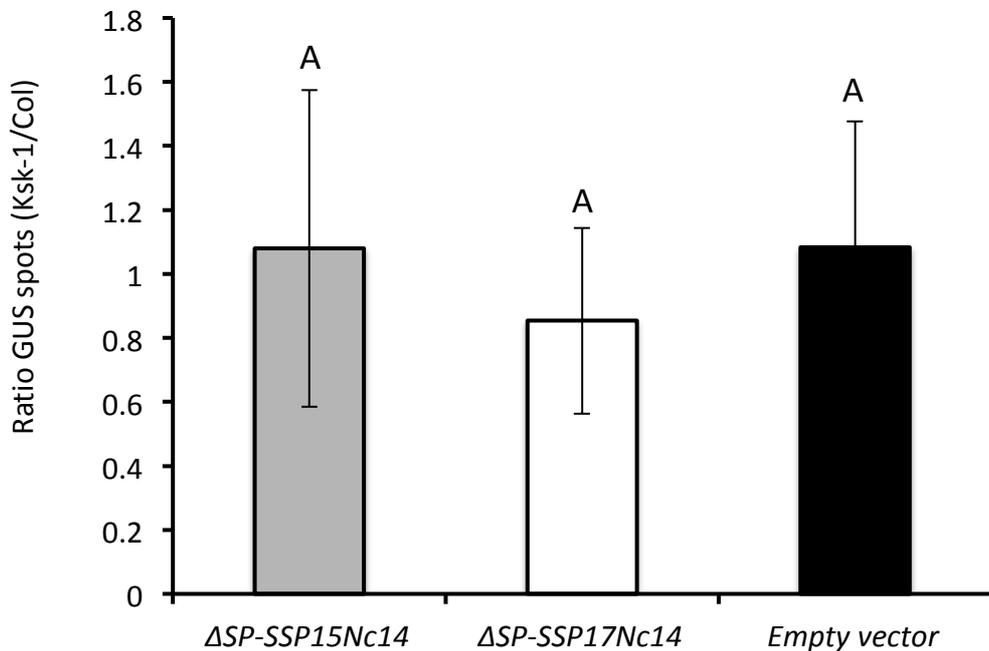
### 5.2.2 *SSP18<sup>Nc14</sup>* IS RECOGNISED BY KSK-1

There were three highly ranked candidates for Ksk-1 recognition from my analysis in chapter 4. *SSP17*, *SSP18* and *SSP21*. Each has high Fu's Fs and pN/pS scores. Each was cloned from Nc14 RACE cDNA. I bombarded the  $\Delta SP$ -*SSP18<sup>Nc14</sup>* and  $\Delta SP$ -*SSP18<sup>Abo1</sup>* alleles under 35S promoter into Ksk-1 and Col-0. The *SSP18<sup>Nc14</sup>* gave a significantly lower GUS spot ratio compared to the EV (Wilcoxon rank-sum test;  $p < 0.05$ ), whereas the virulent *SSP18<sup>Abo1</sup>* allele did not affect the ratio (figure 5.4). I also bombarded  $\Delta SP$ -*SSP15<sup>Nc14</sup>* (another potential candidate, though later excluded by association genomics) and  $\Delta SP$ -*SSP17<sup>Nc14</sup>* under 35S promoter into Ksk-1, with an empty-vector (EV) control. These combinations gave no significant reduction of GUS spot ratio (figure 5.5) (Wilcoxon rank-sum test;  $p > 0.05$ ).

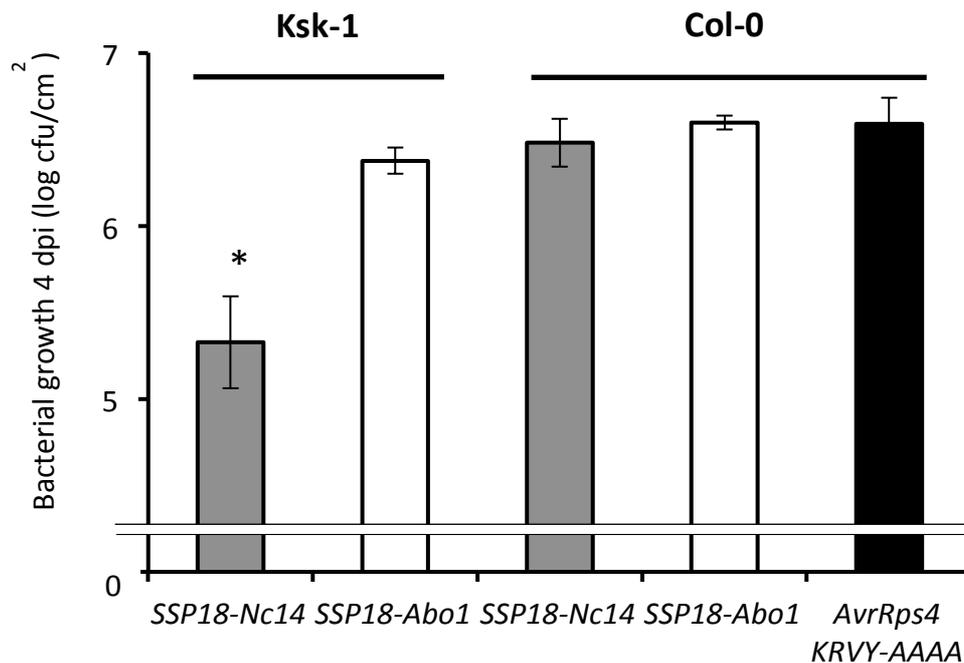
I again employed the effector-detector-vector system (EDV) (Sohn et al., 2007). The secretion-signal truncated Nc14 and Abo1 alleles of *SSP18* were cloned into the EDV6 vector, and these were mated into *Pseudomonas syringae* DC3000 lux strain (Fan et al., 2008). I spray-inoculated Ksk-1 and Col-0 (susceptible) plants and assessed the bacterial growth 4 days post-inoculation through colony counts. I used manual colony counting as opposed to the luciferase-based method as it is considered to be more reliable within our lab. Exclusively in the Ksk-1 accession, the *SSP18<sup>Nc14</sup>* allele showed significantly less growth (figure 5.6) (Student's t-test on mean cfu/cm<sup>2</sup>,  $p < 0.05$ ) of approximately 1 log cfu/cm<sup>2</sup> (over 3 repeat experiments). The *SSP18<sup>Abo1</sup>* allele was indistinguishable from the *AvrRps4<sup>KRVY-AAAA</sup>* control in both Ksk-1 plants and Col-0, where the *SSP18<sup>Nc14</sup>* allele was also indistinguishable from both in Col-0. There was no detectable positive change in the growth of either allele in the susceptible Col-0 plants.



**Figure 5.4.** *SSP18<sup>Nc14</sup>* is recognized by accession Ksk-1; GUS spot ratios for the bombardment of two SSP18 alleles and empty vector in Ksk-1. 6 pairs of leaves were bombarded with each construct. The letters signify the significance of the differentiation of the groups according to whether  $p < 0.05$  in a two-tailed Wilcoxon rank-sum test. The error bars represent standard error.



**Figure 5.5.** *SSP15<sup>Nc14</sup>* and *SSP17<sup>Nc14</sup>* are not recognized by accession Ksk-1; GUS spot ratios for the bombardment of these genes in Ksk-1. 6 pairs of leaves were bombarded with each construct. The letters signify the significance of the differentiation of the groups according to whether  $p > 0.05$  in a two-tailed Wilcoxon rank-sum test. The error bars represent standard error.



**Figure 5.6. *SSP18<sup>Nc14</sup>* is recognized specifically by accession Ksk-1 when delivered by T3SS; Pst DC3000 lux growth curves.** Pst DC3000 lux was transformed with various EDV6 constructs and the strains were assessed for growth *in planta* at 4 dpi in 6 leaf discs per strain from Ksk-1 and Col-0. The error bars represent standard error.

### 5.2.3 ALLELIC DIVERSITY OF *SSP16* AND *SSP18*

#### ***SSP16* allelic diversity**

*SSP16* is one of the most polymorphic loci in *AI*. In the Illumina-sequenced isolates there are in total 62 unique non-synonymous encoding polymorphisms, all of which are homozygous. With the assistance of Ms Agathe Jouet, the sequences of three impure field isolates' *SSP16* alleles were also obtained. Two of these contained further unique mutations, however they were still around 99% similar to previously identified alleles. Table 5.1 shows the nucleotide identity of all *SSP16* alleles to each other as calculated by Clustal Omega (Sievers et al., 2011).

To visualise the diversity of *SSP16* I constructed unrooted phylogenetic trees with several methods. Figure 5.7 shows two representative trees: a) was constructed using the Clustal Omega and neighbour joining method with 1000

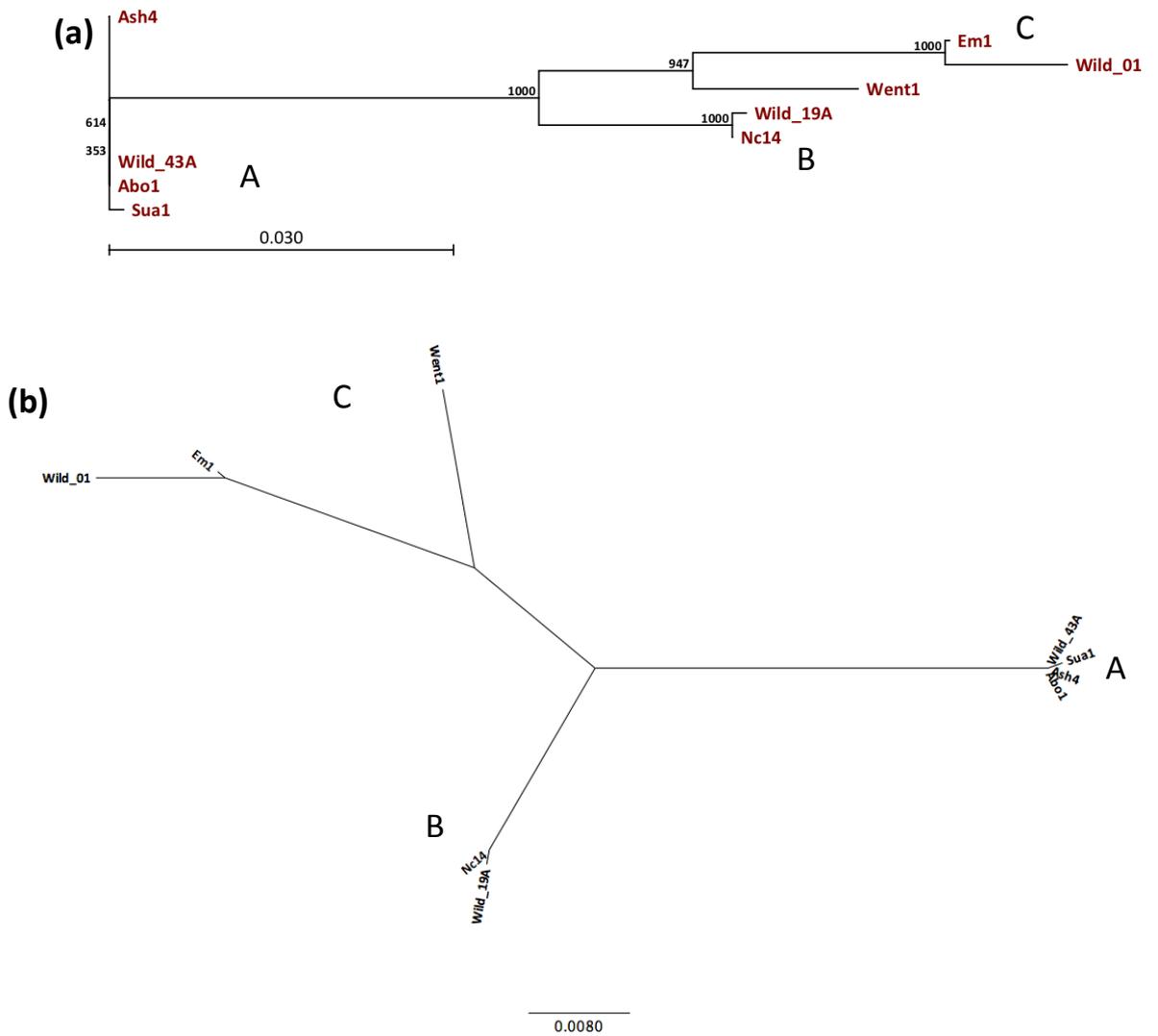
bootstraps in the cladogram layout and b) the same in the radial layout. These show that there are three distinct clades of *SSP16* alleles.

To detect which amino acid residues of *SSP16* might be under the strongest selection I used Bayes Empirical Bayes (BEB) (model M8) analysis to assign ‘positive selection’ values to each residue based on all of the available polymorphism data. This analysis revealed that statistically there are 59 residues under positive selection ( $P > 0.5$ ), and 26 of these very significantly ( $P > 0.99$ ). Figure 5.3 shows a function of the Ka/Ks ratio and probability across the length of the *SSP16* protein. The highly significant sites cluster mainly at the C-terminus of the protein, although there are several in the post signal peptide cleavage region (signal peptide cleavage predicted to be at residues 24/25). Table 5.2 shows the positions of the highly significant positively selected residues and the various amino acids present at these positions.

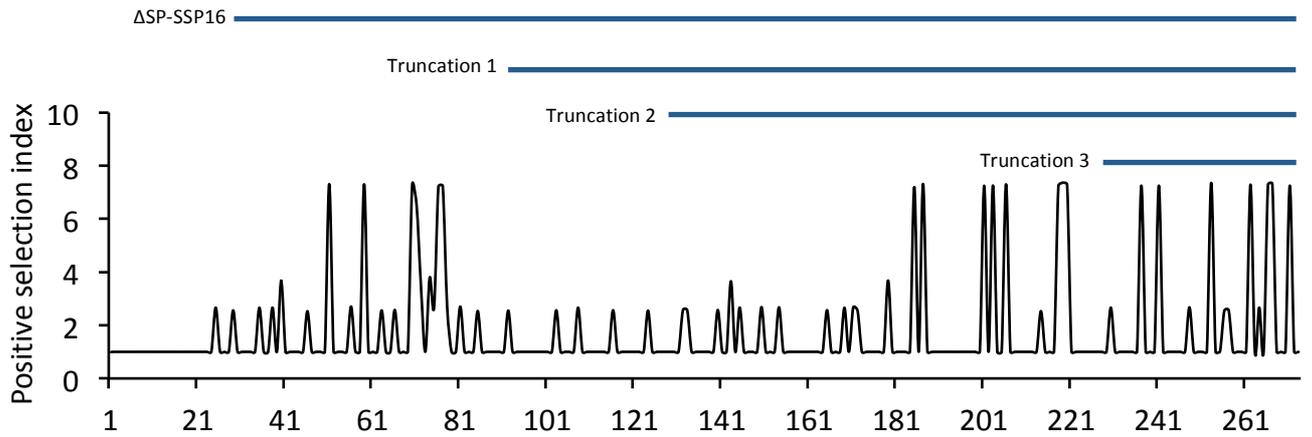
With a standard Pfam search (Finn et al., 2014), no functional annotation of *SSP16* could be made. However a Prosite analysis (Sigrist et al., 2010) revealed a p-loop domain in the C-terminus of *SSP16* clade A proteins. This motif (GEMTAGKT), located in residues 236 – 243, overlaps with two of the highly significantly positively selected residues (237-K/E and 241- D/G).

	Ash4	Abo1	Wild_43A	Sua1	Nc14	Wild_19A	Went1	Wild_01	Em1
Ash4	100	100	100	99.88	94.77	94.65	93.67	92.21	93.07
Abo1	100	100	100	99.88	94.77	94.65	93.67	92.21	93.07
Wild_43A	100	100	100	99.88	94.77	94.65	93.67	92.21	93.07
Sua1	99.88	99.88	99.88	100	94.65	94.53	93.55	92.09	92.94
Nc14	94.77	94.77	94.77	94.65	100	99.88	95.99	93.55	94.65
Wild_19A	94.65	94.65	94.65	94.53	99.88	100	95.86	93.55	94.53
Went1	93.67	93.67	93.67	93.55	95.99	95.86	100	95.38	96.47
Wild_01	92.21	92.21	92.21	92.09	93.55	93.55	95.38	100	98.91
Em1	93.07	93.07	93.07	92.94	94.65	94.53	96.47	98.91	100

**Table 5.1. *SSP16* allelic diversity.** A table of percent similarity was generated from an alignment of the DNA sequences of all of the *SSP16* alleles using Clustal Omega (Sievers et al., 2011).



**Figure 5.7. SSP16 allelic diversity. (a)** A neighbor-joining tree constructed with a Clustal Omega alignment of the DNA sequences of all SSP16 alleles. The numbers at nodes indicate their bootstrap support (#/1000 bootstraps). **(b)** The same tree in radial layout. Based on their position within the trees, I assigned alleles to clades A, B and C.



**Figure 5.8. *SSP16* positive selection analysis with PAML M8 model and Bayes-empirical-Bayes algorithm.** Codons are assessed both for their  $K_a/K_s$  value and the probability that they are undergoing positive selection. The y axis ‘positive selection index’ was calculated as a function of both values. The blue bars above show the portions of the protein used during the truncation experiments in section 5.2.5.

Residue number	Amino acid variants
40	T/P/I
51	H/D/Q
59	A/T/V
70	K/T
71	R/V
72	N/Q
74	D/S
76	K/N
77	K/E/N
143	R/S/H
179	K/E/N
185	N/D
187	H/R
201	K/E
203	T/A
206	E/Q
218	E/K
219	R/P
220	I/F
237	K/E
241	D/G
253	L/S
262	E/A
266	D/G
267	K/Q
271	E/K

**Table 5.2. Residues of *SSP16* under strong positive selection.** As defined by the PAML M8 model and Bayes-Empirical-Bayes algorithm.

### ***SSP18 allelic diversity***

*SSP18* is the same in each of the sequenced isolates except in Abo1, in which it is highly polymorphic. Compared to *SSP18<sup>Nc14</sup>*, *SSP18<sup>Abo1</sup>* has 22 non-synonymous encoding polymorphisms and an in-frame 18 bp insertion that introduces 6 amino acids to the middle of the protein (see figure 5.9). Remarkably, all of the polymorphisms in *SSP18<sup>Abo</sup>* encode non-synonymous changes, resulting in the failure of PAML to assign a pN/pS value. Since there are only two known alleles of *SSP18*, there was no need to run the BEB analysis. There are no Pfam domains in *SSP18* and a Prosite scan revealed only several possible secondary modification sites. A putative mono-partite nuclear localisation signal (NLS) was detected at the C-terminus; it is highlighted in figure 5.9.

```
SSP18_Nc14      MLSPPVLLLLSVVALRIDHVESRNALRIESETATNAYVESPDLTGKSHLGFMTTLSSRK
SSP18_Abo1     MLSPPVLLLLSVVALRIDHVESRNALRIESETATNAYVESPDLTGKSHLGFMTTLPSRK
*****

SSP18_Nc14      SNAHTEHSSRDSKHGSVNALAFERVLTEPDTWTAAEAIAR-----AEEIVQFHNNKY
SSP18_Abo1     SNAHRTEHSSRDSKHGSVNALAFERVLTEPNEDWTAAEAIARAQELVELHRNDF
****:*:*****:*****:*****:*****:*****:*****:*****:*****:

SSP18_Nc14      GIPYPTSVPTTFQEHQRLATLDQQQMLAASTIRAHERESGPSTSKPPRPVKKRSKK
SSP18_Abo1     GIPFPTAERIFQEQORLAALDQQQMLATSTIRAHDRSGPSTSKPPRPVKKRSRK
***:**:  ***:****:*****:*****:*****:*****:*****:*****:*
```

**Figure 5.9. Protein alignment of *SSP18<sup>Nc14</sup>* and *SSP18<sup>Abo1</sup>*.** Produced using Clustal Omega. The putative mono-partite NLS is highlighted in red.

#### 5.2.4 PARALOGS AND HOMOLOGS OF SSP16 AND SSP18

I made a comprehensive search within the *AI* genome to identify paralogs or pseudogenised relatives of *SSP16* and *SSP18*. I used the programs megablast, blastn, blastp and tblastn to make this search. There is no similarity between *SSP16* and *SSP18*.

##### **SSP16 Paralogs and Homologs**

At the nucleotide level, *SSP16* itself was the only significant hit. However at the protein level, one significant hit was identified. *AINc14C28G2719* is adjacent to *SSP16*, 2530 bp from its 3' end and convergently transcribed. *AINc14C28G2719* is 29% identical to *SSP16*, and doesn't have a secretion signal (SignalP 4.1 D=0.111). I examined the natural diversity of *AINc14C28G2719* in the various Illumina sequenced isolates and found that in Em1 and Sua1 there is a homozygous A>G mutation 14 before its start codon. The effect of the mutation would be to change an in-frame TAG (stop codon) to TGG (Tryptophan). There is an in-frame ATG (methionine; start codon) a further 18 bp upstream from the TAG>TGG that could potentially be a start codon for *AINc14C28G2719*<sup>Em1/Sua1</sup>. I constructed this hypothetical ORF *in silico* and found that the protein product has a predicted secretion signal (SignalP 4.1 D = 0.852). An alignment of *SSP16*<sup>Nc14</sup> and *AINc14C28G2719*<sup>Em1/Sua1</sup> is shown in figure 5.10, they are 35.2% identical (Clustal Omega alignment).

In terms of further *SSP16*-like genes or pseudogenes in *AI*, there are two possible pseudogenes revealed by a tblastn search of the genome, with 26% and 23% identity to *SSP16*. However they cannot be resolved into ORFs (there no start codons and multiple stop codons within the aligned regions).

Outside of *AI*, there appears to be a family of *SSP16* related genes in *Albugo candida*. A blastp search of the *Ac* race Nc2 predicted proteome (Gardiner et al., *submitted*) revealed 5 hits with an E-value of  $1 \times 10^{-4}$  or lower, the best of which (annotated in AcNc2 as Gg3270) is shown in alignment against *SSP16* in figure 5.11. Gg3270 is predicted to be secreted (SignalP D= 0.783) and may have a nuclear localisation signal (NLS). Gg3270 shows a high level of polymorphism in *Ac* race 2v

(Links et al., 2011) compared with race Nc2 (Gardiner et al., *submitted*); there are 24 non-synonymous change-encoding polymorphisms between these two races in this gene.

Beyond *Ac*, no further sequences of similarity to *SSP16* in NCBI databases could be identified.

```

AlNc14C28G2718_SSP16      MGFKKSQSSLVILNLFMILNSVWSACFIKKEATCEIVQSTGNNVYMLGLQHSRDSQVKAY
AlNc14C28G2719_hypothetical  -----MSISLVVLHLCMISSFVWSECFMKEKVKCSYKTSA--TPFWFRLNRRRSAPSKGI
      . ***:*.* ** . *** **:::..*. * : : : *:. *.: *.
      TAG in Nc14 ▲ ▲ AINc14C28G2719 translation start in Nc14
AlNc14C28G2718_SSP16      KCKPEDCLSKRNFDTKKQYGRLLSPNGVVGASCNNFHYDLFEDKLMLSQSPDITFVTELDK
AlNc14C28G2719_hypothetical  ECSPERVLLLEK-----ALSPNGPGKSCQWFRYDFFGNKLELSPHHHNKMWEIFDME
      :*.* * ::*          ***** * **: *::*: * :** * ..:* :*:
AlNc14C28G2718_SSP16      AE-----SKPTWEVVFVKVTVNGGVYPRATTFKLFQFGQSTRLLMP-----RWAETHLNV
AlNc14C28G2719_hypothetical  ARRDATRKE DATWKLMDAYYINQSTVEHPATVEFQLVQQTSLFPVKSFEDSWRSKTLKA
      * . . . *:::* :* .. : :*:::*: *.* * :* * .. *:.
AlNc14C28G2718_SSP16      AGKVNFEFYFKYSEVINQHTIVSTLMISTKHVKYTDSEDWVLEIGDRIPERIRISLRKSNS
AlNc14C28G2719_hypothetical  GGIF-----
      .* .
AlNc14C28G2718_SSP16      KYRVVFGKMTADKTLSSKAKNVFLFGSSSSRAKEETKDKIHLELD
AlNc14C28G2719_hypothetical  -----

```

**Figure 5.10. Protein alignment of *SSP16*<sup>Nc14</sup> and the hypothetical *AINc14C28G2719*<sup>Em1</sup>.** *AINc14C28G2719* is the neighbor of *SSP16* and is the closest paralog. The arrows indicate where a mutation occurs in *Em1* that encodes a change to W from a stop codon, and where the predicted start codon in *Nc14* is. Alignment constructed with Clustal Omega.

```

AlNc14C28G2718_SSP16      MGFKKSQSSLVILNLFMILN--SVWSACFIKKEATCEIVQSTGNNVYMLGLQH--SRDSQ
Gg3270                    MCTLLRHFFLVLLQVLLYSSLLGKCKCFIGNDAQTRIKPSDASFFKFEKLSLNKTHG
      * : ***:~:~:~ . . *** :~* * :~: : .. : : : ..
AlNc14C28G2718_SSP16      VKAYKCKPEDCLSKRNF-DTKKQYGRLLSPNGVVGASCNNFHYDLFEDKLMLSQSPDITFVT
Gg3270                    FQKYFCDHYFWWQVASSIKQDTILNLGRNKLKDKCDTFRYTFYEDKLSLSTTYEPDFHS
      .: * *. .: .:~ .*. * :* .*:~*: :~*** ** : : * :
AlNc14C28G2718_SSP16      ELDMKAESKPTWEVVFVKVTVNGGVYPRATTFKLFQFGQSTRLLMPRWAETHLNVAGKVNE
Gg3270                    TIALKASASKNWAVKFDNVTVNGARTLNYSLEFRLGDKTQLLTPSLPRHRRQKNKH---
      : :~*~. .~* * :~***~. . *~:~:~:~:~:~*~* * . : : :
AlNc14C28G2718_SSP16      FYFKYSEVINQHTIVSTLMISTKHVKYTDSEDWVLEIGDRIPERIRISLRKSNSKYRVVF
Gg3270                    -----LVSFNLRQARAIERGDSQNWILVIGPELSEPLTIAFLKKGKSFVHI
      :~* : .: :~ :~ ***:~*~*~* .: * : *~:~* . .~:~* :
AlNc14C28G2718_SSP16      GKMTADKTLSSKAK-----NVFLFGSSSSRAKE-----ETKDKIHLELD
Gg3270                    GNNTFEDKVALASTEQKKKKTTAFFSWIRNKISGSSFNHTVPRYVFLKV-
      * : * ~:~: .: .:~ * * * * . . :~:~:

```

**Figure 5.11. Protein alignment of *SSP16*<sup>Nc14</sup> and *Gg3270*.** *Gg3270* is a predicted protein from *Ac* race *Nc2* and is the most similar protein to *SSP16* outside of *Al*. Alignment produced using Clustal Omega.

## SSP18 Paralogs and Homologs

Blastp and blastn searches revealed that SSP18 is part of a small gene family whose members share some protein level identity, particularly towards the N-termini. However these proteins share only a low level of identity with SSP18, the best being AlNc14C273G9989 with 31% identity (figure 5.12).

There are also three potentially *SSP18* related pseudogenes in the *Al* genome, sharing at least 30% identity.

In other sequences in the NCBI databases, including *Ac*, I could not identify any nucleotide, protein, or translated nucleotide sequences of significant similarity to *SSP18*.

```
AlNc14C56G4264      ----MLSPPVLLLLSVVALRIDHVESRNALRIESETATNAYVESPDLTGKSHLGFMTTL
AlNc14C273G9989    MRTPMHLSPAVLLLLSAVALRMDQVEPHDILQSESKTRKSSRLDSPDVTTGKSNLRTTKHP
                    *  *.:****.****:*:**  :: *: **:*  ..: :.***:****.*  .

AlNc14C56G4264      SSRKSNAHHTHESSRDSKHGSVNALAFERVLTEPDTDWTAEEA--IARAE-----
AlNc14C273G9989    VEV-----RAQQTLSNCKDGDVNAMASGRALMD--NNWQERYNELLDPNKKYFAHLDVIK
                    .      :::::  :.*.*.***:*  *.*  :  .**  :  ::

AlNc14C56G4264      -IVQFHNNKYGIPYPTSVPTFQ-EHQRLATLDQQQLAASTIRAHERESGPSTSKPPRPP
AlNc14C273G9989    GSHSIPSDRYKASNPSSFPHYPYHSHSPV-----KESVHGAHYRHHGKKNKKVREC
                    .:  .:**  *:*.*  .*  :      *  **  *  *  ...*  *

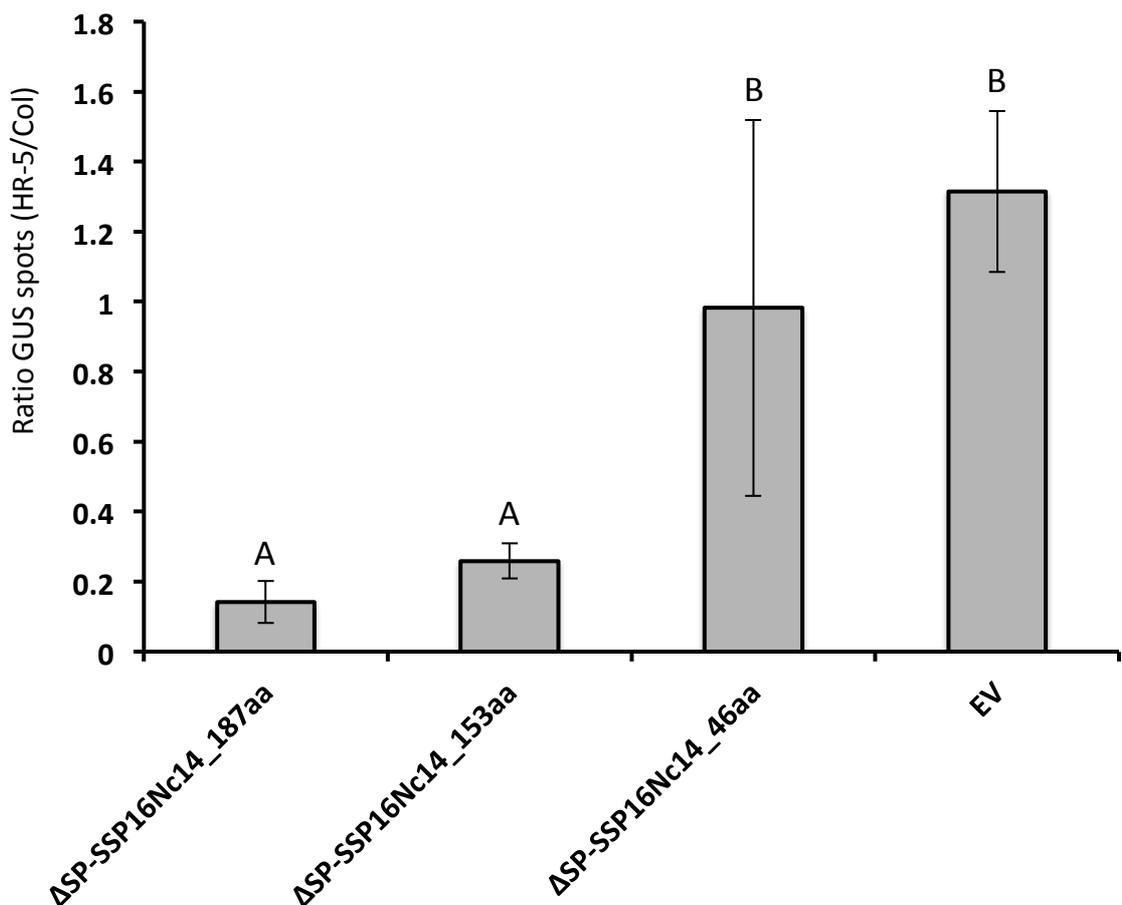
AlNc14C56G4264      VKKRSKK
AlNc14C273G9989    HV-----
```

**Figure 5.12. Protein alignment of SSP18<sup>Nc14</sup> (AlNc14C56G4264) and AlNc14C273G9989.** AlNc14C273G9989 is the closest paralog to SSP18<sup>Nc14</sup>. The red box highlights the shared predicted secretion signal cleavage region. Alignment produced using Clustal Omega.

### 5.2.5 THE C-TERMINUS OF SSP16<sup>Nc14</sup> IS SUFFICIENT FOR RECOGNITION

In order to ascertain which part of SSP16<sup>Nc14</sup> is sufficient for recognition by HR-5, a series of truncated versions were cloned and tested using the GUS-eclipse assay. Using the protein's predicted secondary structure as a guide I designed three truncations from the N-terminus. The secretion-signal truncated product that was bombarded in 5.2.1 was 249 amino acids, and I designed truncated versions of 187, 153 and 43 amino acids (see figure 5.6).

As before, I bombarded 35S-promoter constructs of each of these into HR-5 and Col-0 with *p35S:GUS*. I found that the 187 and 153 amino acid peptides caused a significant reduction in the GUS expression level in HR-5 compared to Col-0, but the C-terminal 46 amino acid peptide did not (figure 5.13). This suggests that the peptide consisting of amino acids 96-249 is sufficient for recognition, but the short 204-249 peptide is not.



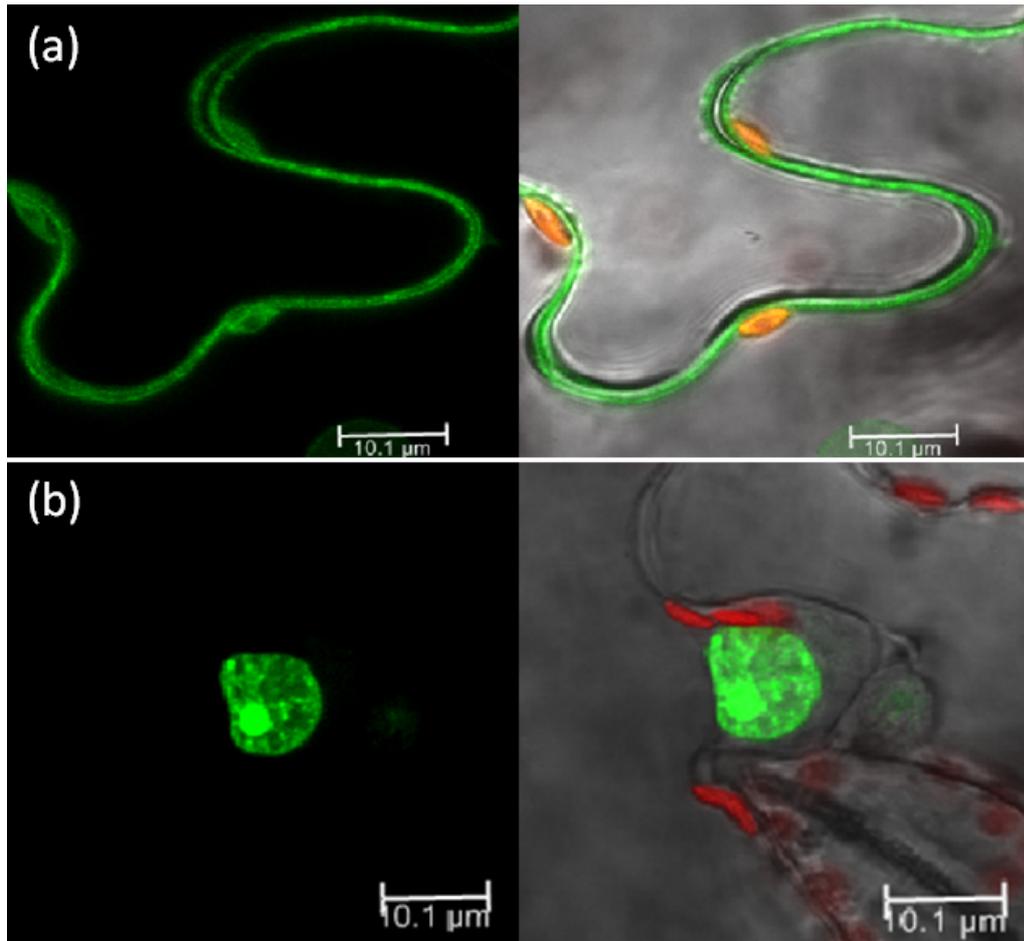
**Figure 5.13. GUS spot ratios for the bombardment of SSP16 truncations in HR-5.**

Truncations 1 (187aa) and 2 (153aa) appear to be recognized, but not truncation 3 (46aa). 6 pairs of leaves were bombarded with each construct. The letters signify the significance of the differentiation of the groups according to whether  $p < 0.05$  in a two-tailed Wilcoxon rank-sum test. The error bars represent standard error.

### 5.2.6 THE SUB-CELLULAR LOCALISATIONS OF SSP16 AND SSP18 IN *N. BENTHAMIANA*

The sub-cellular localisation of the signal peptide- truncated forms of SSP16<sup>Nc14</sup> and SSP18<sup>Nc14</sup> were predicted using WoLF PSORT (Horton et al., 2007). The prediction for  $\Delta$ SP-SSP16<sup>Nc14</sup> is unclear; the software found similarity in terms of sorting motifs and other features with 5 chloroplastic proteins, 3 nuclear proteins, 1 secreted protein and 1 vacuolar protein.  $\Delta$ SP-SSP18<sup>Nc14</sup> however had a clear prediction: the most similar proteins were 13 nuclear-localised proteins. The amino acid sequence of SSP18 contains a lysine rich region at the C-terminus (highlighted in figure 5.4), characteristic of nuclear-localisation motifs (Kosugi et al., 2009).

To ascertain the sub-cellular localisation of  $\Delta$ SP-SSP16<sup>Nc14</sup> and  $\Delta$ SP-SSP18<sup>Nc14</sup> *in planta* I constructed 35S promoter N-terminal Green Fluorescent Protein (GFP)-fusion constructs for plant transformation. *p35S::GFP- $\Delta$ SP-SSP16<sup>Nc14</sup>* and *p35S::GFP- $\Delta$ SP-SSP18<sup>Nc14</sup>* were transformed into *Agrobacterium tumefaciens*, which were then used to transiently express these genes in *N. benthamiana* leaves. I examined cells of the infiltrated leaves using a confocal microscope and found that GFP- $\Delta$ SP-SSP16<sup>Nc14</sup> seems to localise to the cell membrane (figure 5.14a) and GFP- $\Delta$ SP-SSP18<sup>Nc14</sup> localises to the nucleus/nucleolus and seems to cause some form of aggregation in the nucleoplasm (figure 5.14b). Note that although GFP- $\Delta$ SP-SSP18<sup>Nc14</sup> gave a strong signal, the GFP- $\Delta$ SP-SSP16<sup>Nc14</sup> signal was quite weak necessitating a high gain setting. I also transformed the *p35S::GFP- $\Delta$ SP-SSP16<sup>Nc14</sup>* and *p35S::GFP- $\Delta$ SP-SSP18<sup>Nc14</sup>* constructs into *At* accession Col-0. In future work I will check if the sub-cellular location of these proteins is the same in Arabidopsis as in *N. benthamiana*.



**Figure 5.14. The sub-cellular localisations of SSP16 and SSP18 in *N. benthamiana*.** *Agrobacterium* carrying (a)  $35S::GFP-\Delta SP-SSP16^{Nc14}$  and (b)  $35S::GFP-\Delta SP-SSP18^{Nc14}$  Ti plasmids were infiltrated into *N. benthamiana* leaves and checked with a confocal microscope 3 dpi. Images on the left show the GFP channel and on the right the bright-field merged with the GFP channel.

### 5.3 DISCUSSION

I found that one candidate, SSP16<sup>Nc14</sup>, is recognised by the *At* accession HR-5, and another, SSP18<sup>Nc14</sup>, is recognised by the accession Ksk-1. My data show that while SSP16<sup>Nc14</sup> triggers a GUS eclipse, the two tested 'virulent' alleles from virulent isolates Em1 and Abo1 consistently do not have as high GUS expression as the empty-vector control, and are therefore statistically indistinguishable from either the control or the Nc14 allele. This suggests that these alleles do not fully evade recognition and perhaps this weaker recognition is slow enough or of low enough magnitude for *AI* to suppress it with other effectors. On the other hand, there is a clear difference between SSP18<sup>Nc14</sup> recognition and the virulent SSP18<sup>Abo1</sup> variant that seems not to reduce GUS activity.

Unlike the RxLR and Crinkler genes that are part of expanded multigene families with N-terminal similarity but divergent C-termini (Schornack et al., 2009) neither SSP16 nor SSP18 is a member of a large multigene family. Indeed, neither SSP16 nor SSP18 has a known N-terminal host membrane translocation motif. Given that the recognition by specific *At* accessions that I observe occurs within the plant cell (expression of signal-peptide truncated versions with bombardment or bacterial type 3 secretion), in the native system the protein must enter the host via an unknown mechanism. Previous studies on RXLR effectors have suggested that when bombarded full length (with secretion signal), they could be secreted by the plant cell and then re-enter from the apoplast following secretion signal cleavage (Dou et al., 2008). However, since this conclusion is based around recognition and HR, it is unclear whether the recognition of that effector could occur prior to secretion. One further experiment that could be performed to check if the N-terminal, post secretion signal, region of SSP16 or SSP18 can enhance cell uptake would be to construct Avr3a fusions and transform them into *Phytophthora capsici* and check for Avr3a recognition (Schornack et al., 2010).

The BEB analysis of SSP16 variation indicated that its C-terminal part might be important for recognition. Assays of truncated forms suggest that a 153 amino acid portion of the C-terminus of SSP16 is sufficient for recognition. However there are also several positively selected residues in the post signal-peptide cleavage

region. The biological relevance of the non-synonymous mutations in these residues is unclear. It is possible that they have some relevance for either the virulence function of the protein, or for translocation into the plant cell. The discovery of three further alleles from wild samples, which each fell into one of the three different *SSP16* clades lends further support the hypothesis proposed in chapters 1 and 4, that recognised effectors should be under balancing selection. Indeed the discovery of the Wild\_19A allele of *SSP16* that is almost identical to the Nc14 allele suggests that there may be some fitness benefit for retaining this recognised effector. Similar to *SSP16*, there are at least 15 alleles of *ATR13*. Different accessions have differential capabilities to recognise these different alleles and I predict that various different accessions could recognise different *SSP16* alleles (Hall et al., 2009). Indeed the partial increase of GUS spots observed in HR-5 with the *SSP16<sup>Em1</sup>* and *SSP16<sup>Abo1</sup>* alleles is reminiscent of the 'intermediate' recognition phenotype observed with some *ATR13* alleles in some *At* accessions (Hall et al., 2009). Nevertheless, this intermediate recognition is completely overcome by the *Al* isolates virulent on HR-5, where there is no evidence of recognition.

It is remarkable given the overall level of polymorphism within both *SSP16* and *SSP18* that in both cases there are no polymorphisms within the secretion signal-encoding region. This suggests that the region could be under purifying selection to preserve its function. Since it is hypothetically cleaved within the pathogen prior to secretion, it should play no role in recognition or effector activity, and it makes sense that it would be disconnected in terms of selective forces from the rest of the gene. This strengthens the case that these are secreted and recognised effectors.

In order to confirm the apparent recognitions of *SSP16<sup>Nc14</sup>* and *SSP18<sup>Nc14</sup>* I employed the EDV system (Sohn et al 2007). I could confirm that specifically when *SSP16<sup>Nc14</sup>* is delivered by *Pseudomonas syringae* to HR-5, and *SSP18<sup>Nc14</sup>* to Ksk-1, growth is restricted 10-fold 4 days post-inoculation. However I did not find that either *Abo1* or *Em1* allele of *SSP16*, or the *SSP18<sup>Abo1</sup>* allele enhance bacterial growth significantly in the susceptible accession Col-0. This suggests that the virulence function of these effectors are either redundant with the functions of the

effectors already secreted by *Pto* DC3000, or that their function is somehow specific to the virulence strategy of *AI*.

I was able to identify a possible homolog of *SSP16* in the related species *A. candida*. Interestingly this gene shows a high degree of polymorphisms within *Ac* races and is probably substantially diverged from the presumed common ancestor of it and *SSP16*. This is concurrent with the model that effectors undergo co-evolution with their host targets (Dong et al., 2014), and this would appear to be an ongoing process in both *AI* and *Ac SSP16* genes.

At least in *N. benthamiana*, GFP- $\Delta$ SP-SSP16<sup>Nc14</sup> localises to the cell membrane, and GFP- $\Delta$ SP-SSP18<sup>Nc14</sup> to the nucleus. Caillaud et al, (2012) showed RXLR effectors from *Hpa* that localised to both of these compartments, and found that the *Hpa* effector repertoire is enriched for nuclear-localised effectors. Indeed the nuclear 'speckles' caused by GFP- $\Delta$ SP-SSP18<sup>Nc14</sup> are similar to some of those caused by *Hpa* RXLR effectors (Caillaud et al., 2012). It is premature however to speculate possible functions for these effectors based on these localisations.

## CHAPTER 6: IDENTIFICATION AND CHARACTERISATION OF THE *RAL4* LOCUS

### 6.1 INTRODUCTION

In chapter 5, I showed that two secreted proteins from *Al* are recognised by certain *At* accessions, using two different assays. Recognition of *Avr* gene products is typically conferred in host plants by either TIR- or CC-NB-LRR class receptor proteins known as R proteins.

Two *R* genes effective against *Albugo* species have previously been cloned: *RAC1*, encoding a TIR-NB-LRR, confers resistance against *Al*/Em1 (Borhan et al., 2004) and *WRR4*, also encoding a TIR-NB-LRR, confers resistance to several *Ac* races (Borhan et al., 2008). Against *Hpa*, a pathogen with a similar life-style to *Al* on *At*, numerous *R* genes have been identified. These include several CC-NB-LRR encoding genes: *RPP7*, *RPP8* and *RPP13*, and several TIR-NB-LRR encoding genes: *RPP1*, *RPP2A/B*, *RPP4* and *RPP5* (Eulgem et al., 2007; McDowell et al., 1998; Bittner-Eddy et al., 2000; Botella et al., 1998; Sinapidou et al., 2004; Parker et al., 1997). Additionally a CC-NB-LRR encoding gene called *RPP39* was identified that confers recognition of an effector but not pathogen resistance (Goritschnig et al., 2012).

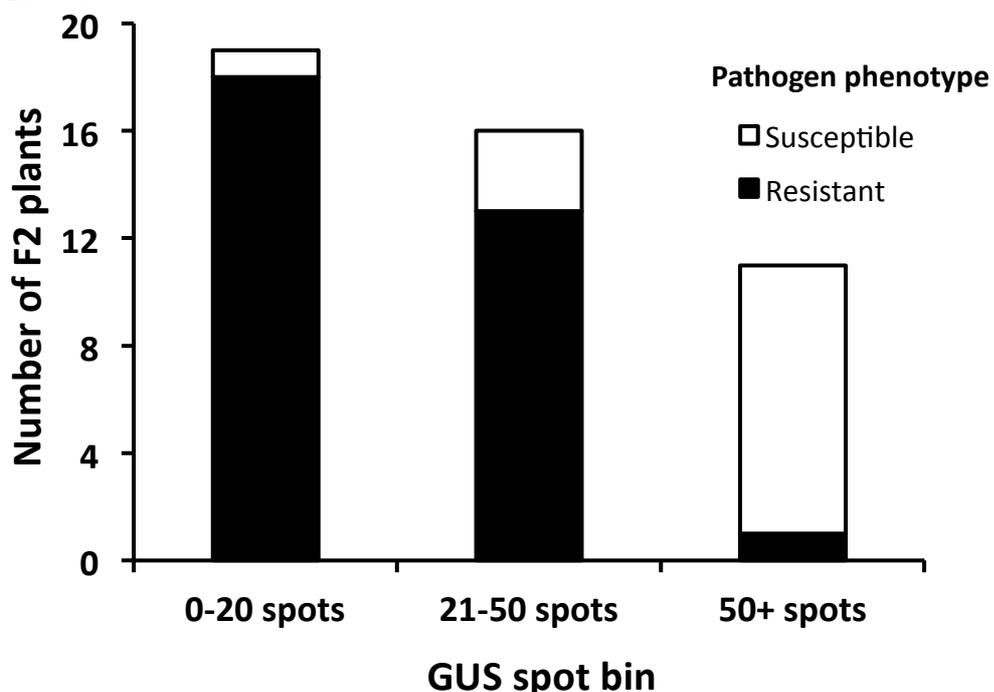
I set out to test the hypothesis that *SSP16*<sup>Nc14</sup> and *SSP18*<sup>Nc14</sup> are recognised by R-proteins, probably belonging to one of the NB-LRR classes. In this chapter I report on i) classical genetic mapping, ii) comparative genomics and iii) *R* gene enrichment and next-generation mapping to identify the gene conferring recognition of *SSP16*<sup>Nc14</sup>, and subsequent experiments to test candidate *R* genes.

### 6.2 RESULTS

#### 6.2.1 CO-SEGREGATION OF *SSP16*<sup>Nc14</sup> RECOGNITION AND *NC14* RESISTANCE IN *HR-5*

A cross between accessions *HR-5* (resistant) and *Ws-2* (susceptible) was made by Dr. Alexandre Robert-Seilaniantz. Two *F*<sub>1</sub> plants were selfed to produce two populations of *F*<sub>2</sub> seeds. I tested 48 plants of each population and found segregations indistinguishable from 3:1 (R:S) in both (Chi<sup>2</sup> test, p=0.50). This suggests that a single dominant gene confers resistance in *HR-5*. To determine whether the same gene that confers resistance also confers the recognition of

SSP16<sup>Nc14</sup>, I devised a strategy whereby the same plants could be bombarded and tested with Nc14. Instead of bombarding two leaves, I arranged the leaves so that two control leaves and four F<sub>2</sub> leaves (from two F<sub>2</sub> plants) could be bombarded simultaneously, and that each F<sub>2</sub> plant could be designated as recognising, or not recognising, SSP16<sup>Nc14</sup>. I screened 46 F<sub>2</sub> plants using this method, and revealed that 18/19 plants showing a strong GUS eclipse in this assay were resistant to the pathogen. Of an additional 16 plants that showed less than 50 GUS spots, 13 were resistant. Finally, of 11 plants that showed no GUS eclipse (ie more than 50 spots), only 1 was resistant. These data are shown in figure 6.1. Overall the segregation with the pathogen was 32:14 (indistinguishable from 3:1, Chi<sup>2</sup> test, p=0.46). The leaves from the two HR-5 control plants showed 0-20 spots. Assuming that the SSP16<sup>Nc14</sup> recognition is unrelated to pathogen resistance, but is conferred by a single locus, then one quarter of each GUS spot bin should be susceptible (a ratio of 4.75:4:2.75- these numbers are 25% of the total number in each bin). However the observed ratio is 1:3:10, significantly different by Chi<sup>2</sup> test (p=5.4x10<sup>-5</sup>). Similar results were found in a second experiment. Considering the inherent variability of the GUS-eclipse method, these data suggest that the resistance to Nc14 is conferred by the same or a closely linked locus as that which confers the recognition of SSP16<sup>Nc14</sup>.

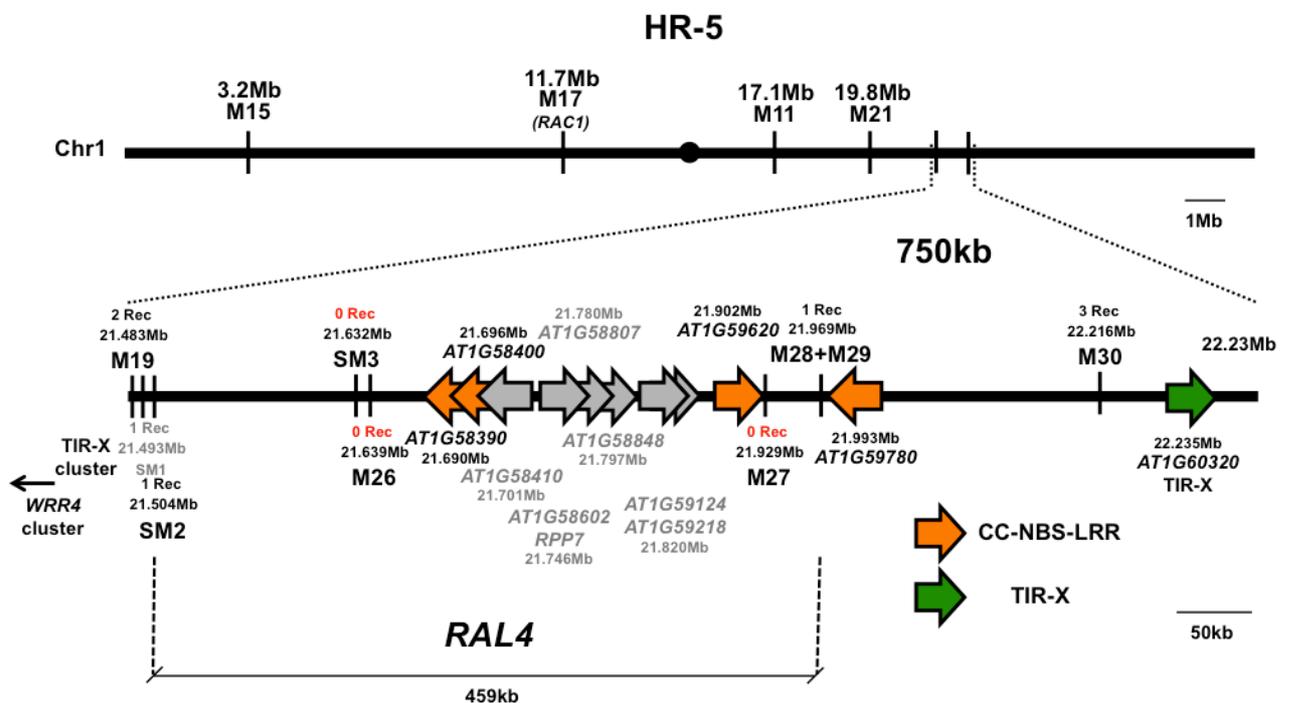


**Figure 6.1. Evidence for the co-segregation of SSP16<sup>Nc14</sup> recognition and Nc14 resistance.** Two leaves from each of 46 F<sub>2</sub> plants were bombarded with 35S:SSP16<sup>Nc14</sup> and 35S:GUS and assessed for their level of GUS spots, either strong GUS eclipse (0-20 spots), intermediate (20-50 spots) or no GUS eclipse (>50 spots). Each plant was tracked and scored with Nc14 pathogen 14 DPI.

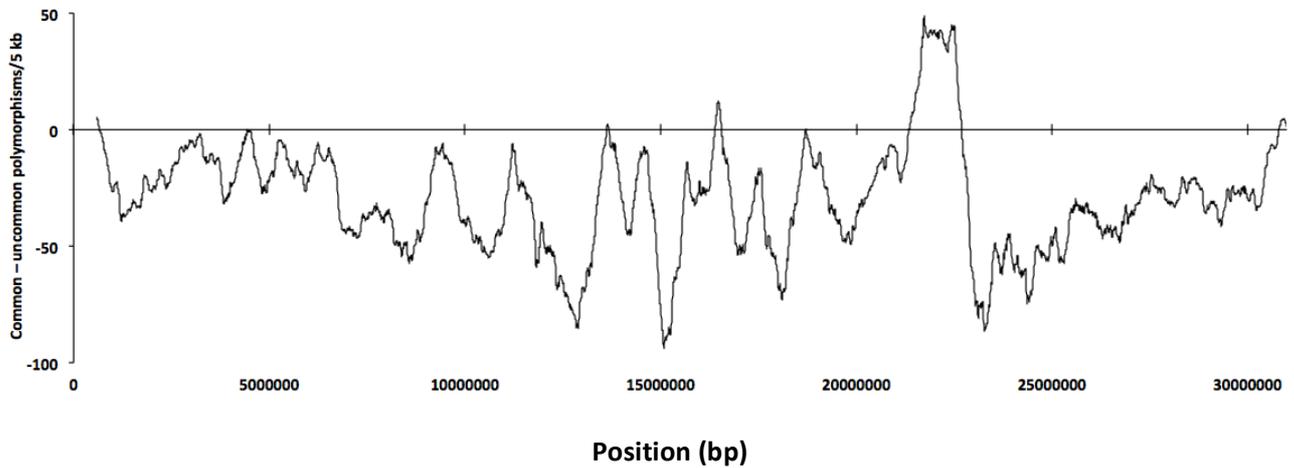
### 6.2.2 MAPPING OF THE RESISTANCE TO *ALBUGO LAIBACHII* 4 (*RAL4*) LOCUS

To identify the *Resistance to Albugo laibachii* 4 (*RAL4*) locus, I used positional cloning. I designed and verified molecular markers based on the alignments of 100 bp and 76 bp HR-5 and Ws-2 Illumina reads to the Col-0 genome. The HR-5 reads were provided by the Salk institute as part of the 1001 genomes project (Weigel and Mott, 2009) and the Ws-2 reads were generated in-house for a previous project by Dr Alexandre Robert-Seilaniantz. Using these read alignments, I scanned the genome and identified potential Simple Sequence Length Polymorphism (SSLP) markers, where one accession appeared to have a deletion of between 50 and 200 bp. I then designed primers to amplify across these deletions such that gel electrophoresis could detect the difference in the size of the product. Such markers are easier to use than dCAPs markers, which often require optimisation. Initially I identified 14 markers that covered the whole *At* genome (at least one marker per chromosome arm). I screened at least 30 susceptible individuals with each of these markers and found around 50% recombination (ie a mix of both parental genotypes and heterozygous polymorphism) at all but two positions. At marker 17 (M17), around 11.7 Mb on chromosome 1, I found a ratio of 3:9:70 (HR-5:Het:Ws-2), indicating a skew towards the susceptible genotype. Further up on chromosome 1 with M11 (17.1 Mb), I observed a ratio of 0:7:50, indicating that this is probably closer to the causal locus. Finally, with an additional marker and additional F<sub>2</sub> susceptible samples at M18 (21.2 Mb), I found a ratio of 0:1:181. In the region around this marker, there are multiple *R* genes. At 21.17 Mb lies *WRR4* and several related TIR-NB-LRR encoding genes (Borhan et al., 2008), and at 21.74 Mb lies *RPP7* and several related CC-NB-LRR encoding genes (Eulgem et al., 2007). To fine map the region, I designed further markers and increased the size of the genotyping pool of F<sub>2</sub> susceptible plants to 500. The results of the fine-mapping are shown in figure 6.2. When no further SSLP markers could be found, several sequencing markers (SM) were generated. Select recombinants were amplified at these sequences that contained SNPs between HR-5 and Ws-2, and the products were Sanger sequenced. A region in HR-5 corresponding to 459 kb containing 9 CC-NB-LRRs in Col-0 was defined as carrying the *Resistance to Albugo laibachii* 4 (*RAL4*) locus.

As an alternative method to identify the region harbouring *RAL4*, I used a direct comparison of the genomes of HR-5 and Ren-11. Ren-11 has the same phenotype as HR-5; it also can only resist Nc14 (chapter 3). I have evidence from a bombardment experiment, that is not presented in this thesis because the experiment was only performed once, that Ren-11 can also recognise SSP16<sup>Nc14</sup>. Therefore I hypothesised that it might contain the same *R* gene. By extracting all of the polymorphisms predicted for each accession from the Salk Illumina data, I carried out a 5 kb sliding window analysis of genome-genome similarity for these two accessions, subtracting the number of uncommon polymorphisms from the number of common ones within each window. This analysis revealed only a 1.3 Mb region of substantial similarity; the region between 21.4 and 22.7 Mb on chromosome 1, overlapping with the *RAL4* locus identified through classical mapping. Most regions of the HR-5 and Ren-11 genomes were more different than they are similar (figure 6.3).



**Figure 6.2. Fine mapping at the *RAL4* locus.** The region containing *RAL4* was fine mapped using a population of 500 susceptible F<sub>2</sub> plants from a cross between HR-5 and Ws-2. M = Marker and SM = Sequencing Marker. Size bars indicated as a guide only; not precisely to scale.

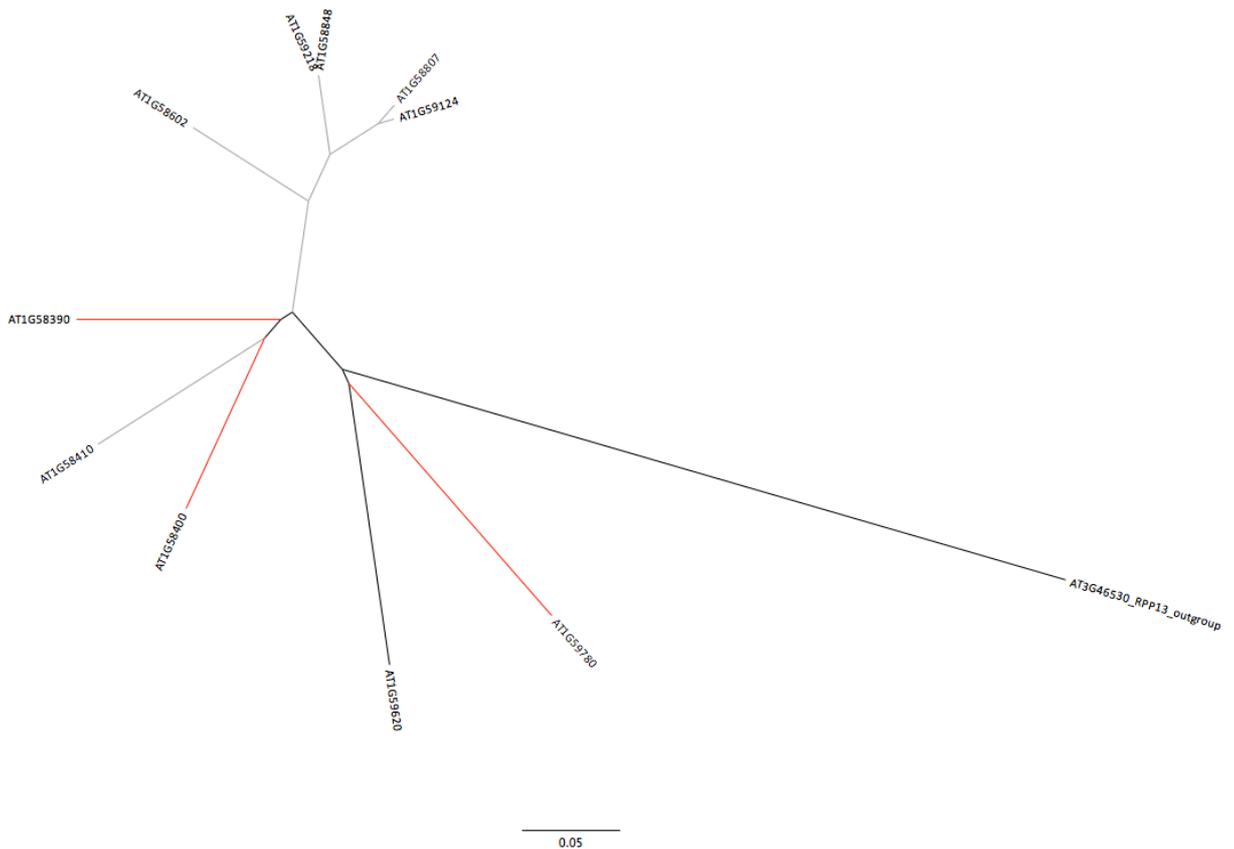


**Figure 6.3. A polymorphism identity plot between HR-5 and Ren-11 also reveals the *RAL4* region.** The number of common and uncommon polymorphisms against the Col-0 genome in sliding 5 kb intervals were subtracted from one another. These data were plotted along chromosome 1.

### 6.2.3 *RAL4* CANDIDATE GENES

An examination of the *R* gene candidates within the *RAL4* interval in Col-0 revealed 9 putative CC-NB-LRR encoding genes: *AT1G58390*, *AT1G58400*, *AT1G58410*, *AT1G58602*, *AT1G58807*, *AT1G58848*, *AT1G59124*, *AT1G59218* and *AT1G59620*. These genes are closely related and, on a phylogenetic tree of *At* CC-NB-LRRs, as constructed by Meyers et al (2003), cluster together in a monophyletic group. Figure 6.4 shows a radial phylogeny of the cluster, with *RPP13* as an out-group. The alignment of HR-5 reads to the Col-0 genome revealed that 6 of these genes are probably not present in this accession. Table 6.1 shows the read depth and breadth of these genes in the alignment. The genes that are present appear to be *AT1G58390*, *AT1G58400* and *AT1G59620*, however they are very polymorphic in HR-5 compared to Col-0. In order to a) resolve the sequence of these genes and their promoters and b) check for divergent relatives of the *RPP7* cluster, I assembled the 100 bp paired-end Salk Illumina data for HR-5 using Velvet (Zerbino et al., 2008). After several iterations, a useful assembly of 121 Mb and an N50 of 22.7 kb was generated. The Col-0 sequences of each of the *RPP7/RAL4* cluster genes were blasted (blastn) against the assembly and the HR-5 allele of each gene was identified. I identified a single contig of 27.5 kb that contained the HR-5 alleles of *AT1G58390* and *AT1G58400*. This contig also revealed that *AT1G58410* is absent in the HR-5 genome: I identified where flanking sequences from both sides of the gene in Col-0 are fused in HR-5. I also identified the full-length HR-5 alleles of *AT1G59620* and *AT1G59780* from HR-5 on contigs of 46 and 15 kb respectively. The top hits

from the blastn of all the other *RPP7/RAL4* cluster genes hit the same three contigs, suggesting that these genes are not present in HR-5. Analysis of the alignments of the Col-0 and HR-5 alleles of the three candidate genes suggest that there are 58 non-synonymous polymorphisms in *AT1G58390*, 76 in *AT1G58400* and 3 in *AT1G9620*. *AT1G59780*, excluded by a single recombinant, has 29 non-synonymous polymorphisms between HR-5 and Col-0 alleles. In addition to the 3 CC-NB-LRR encoding genes within the *RAL4* interval there are 82 predicted protein-encoding genes in Col-0, according to The Arabidopsis Information Resource version 10 annotation. See appendix table 6A1 for the list of non NB-LRR encoding genes in the *RAL4* interval. Of note, there is one receptor-like protein (RLP), annotated in *At* as *RPL9*. RPL9 is predicted to have a secretion signal, 7 leucine-rich repeat domains and a transmembrane domain.



**Figure 6.4. Radial phylogeny of the genes in the *RPP7/RAL4* cluster in Col-0.** A clustal omega nucleotide alignment was generated, and used as the basis for a neighbor-joining tree. The radial alignment helps to visually distinguish the various groups of more closely related genes. Those with red branches are present in HR-5 and within the *RAL4* mapping interval. Those with grey are apparently not present in the HR-5 genome. The scale bar indicates substitutions/site.

Gene	Length	Mean Coverage	Percentage Covered
<i>AT1G58390</i>	2724	39.68	100
<i>AT1G58400</i>	2703	36.89	100
<i>AT1G58410</i>	2700	0	0
<i>AT1G58602</i>	3417	6.86	68
<i>AT1G58807</i>	2568	4.16	57
<i>AT1G58848</i>	3150	3.08	47
<i>AT1G59124</i>	2568	2.76	51
<i>AT1G59218</i>	3150	2.67	40
<i>AT1G59620</i>	2529	38.28	100
<i>AT1G59780</i>	2721	42.25	100

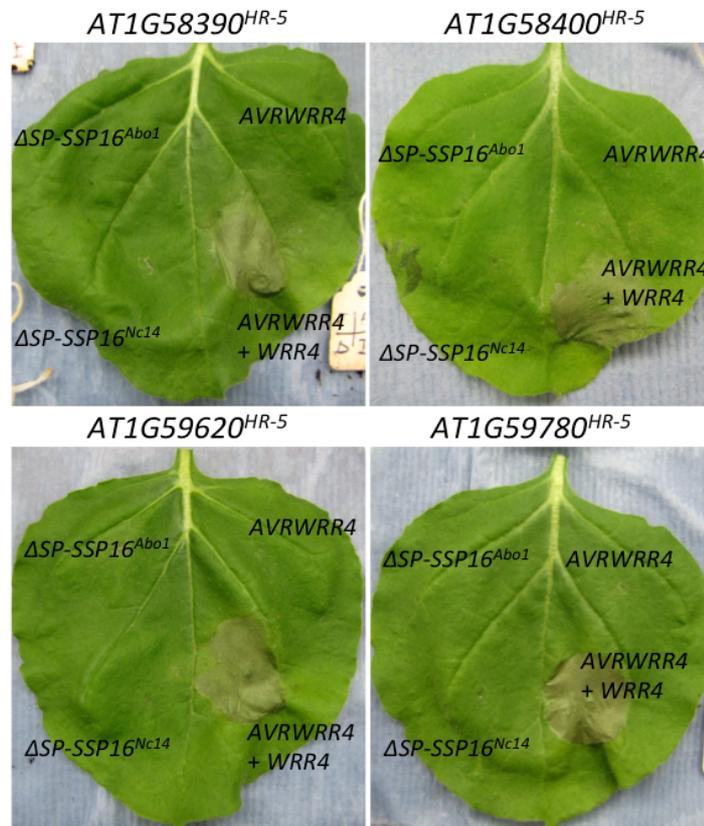
**Table 6.1. Statistics for the breadth and depth of genes in the *RPP7/RAL4* cluster in an alignment of HR-5 reads against the Col-0 genome.** These statistics were collected using a custom Perl script. Note, *AT1G59780* was excluded as *RAL4* through a single recombinant F2 susceptible but is nominally included in the cluster.

#### 6.2.4 CLONING AND TESTING OF *RAL4* CANDIDATE GENES

Each of the three *RAL4* candidates and *AT1G59780* were cloned full-length from HR-5 gDNA into a *p35S* promoter binary vector. I transformed these into *Agrobacterium tumefaciens*. In addition, *p35S::ΔSP-SSP16<sup>Nc14</sup>* and *ΔSP-SSP16<sup>Abo1</sup>* were also transformed into this strain. In order to assess if these genes could trigger a hypersensitive response when expressed in any combination, the agrobacterium co-infiltration assay (Van der Hoorn et al., 2000) in *Nicotiana benthamiana* was used. These experiments revealed that the co-expression of *ΔSP-SSP16<sup>Nc14</sup>* and any of the *RAL4* candidates in *N. benthamiana* (figure 6.5) did not result in a HR. To test if any of the *RAL4* candidates confer resistance to *AI* in *At*, full length genomic sequences including ~2 kb of upstream native promoter sequence from HR-5 were cloned into a binary vector and transformed into *A. tumefaciens* and the susceptible accession Col-0 was transformed with this construct. These *At* experiments are on-going and the results will be reported in future work.

NDR1 and EDS1 are proteins each required for some R proteins to function in *At*. HR-5 was crossed to the Col-0 *ndr1-1* mutant (Century et al., 1997) and *Ws-2 eds1-1* mutant (Parker et al., 1996) to test if either of these proteins are required for *RAL4* function. In two separate HR-5 x Col-0 *ndr1-1* F<sub>2</sub> populations, resistance to Nc14

segregated 3:1 (R:S) (540:188 and 652:201, both indistinguishable from 3:1 by  $\chi^2$  test,  $p=0.61$  and  $p=0.33$ ) suggesting that NDR1 is not required for RAL4 function. The HR-5 x Ws-2 *eds1-1* F<sub>2</sub> populations will be tested in a future work.

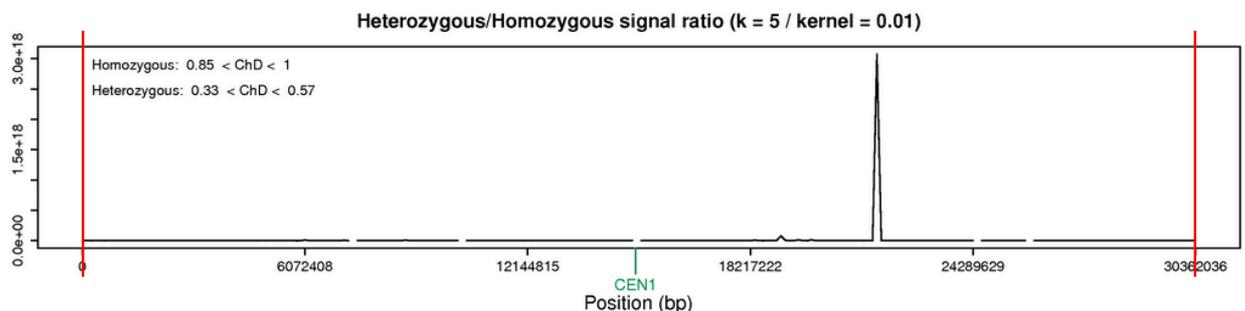


**Figure 6.5. Agroinfiltration of *N. benthamiana* with various *p35S RAL4* candidate constructs and *p35S:SSP16* alleles.** The positive control used on every leaf is the combination of a gene from *Ac* that triggers *WRR4* mediated HR in *N. benth*. These positive control constructs were provided by Dr Volkan Cevik. Photos taken 5 dpi.

### 6.2.5 R GENE ENRICHMENT SEQUENCING AND ILLUMINA BASED MAPPING

As a further measure to identify the genetic basis of resistance to Nc14, I implemented a strategy based around the enrichment of NB-LRR encoding genes (informally, *R* genes) from DNA samples using synthetic biotinylated RNA baits, as used by Jupe et al (2012). The development of the system for *Brassicaceae* RenSeq is discussed in detail in chapter 2b. Using the method I obtained 300 bp paired-end Illumina MiSeq data for the NB-LRRome of Col-0, HR-5, Ws-2 and a bulked susceptible F<sub>2</sub> population from the HR-5 x Ws-2 cross, as well as several other accessions, species and crosses that will not be discussed here. The objective of

obtaining this data is twofold: to confirm the *RAL4* locus in HR-5 and to identify any additional NB-LRRs that may be within the *RAL4* interval. First, the reads from Ws-2, HR-5, and the susceptible F<sub>2</sub> bulk were aligned to the Col-0 genome. Manual inspection of the *RAL4* genes revealed that 99% of the reads in the F<sub>2</sub> bulk at the *AT1G58390* and *AT1G58400* loci were derived from the Ws-2 alleles (in fact *AT1G58400* is deleted in Ws-2, so read coverage was very low). In order to check the location of the *RAL4* locus in a systematic way, the alignment data was uploaded to the “Next-gen mapping” server produced by Austin et al, (2011) (<http://bar.utoronto.ca/NGM>). This algorithm analyses Illumina alignment data to find linked loci in bulked sequencing experiments of either EMS mutants or wild accessions. Analysing the Ws-2 x HR-5 bulk F<sub>2</sub> susceptible, and considering polymorphic sites of “chastity” (a term referring to the degree of homozygosity at a polymorphic site) of between 85 and 100% homozygous susceptible resulted in a single strong peak at the *RAL4* locus (figure 6.6).



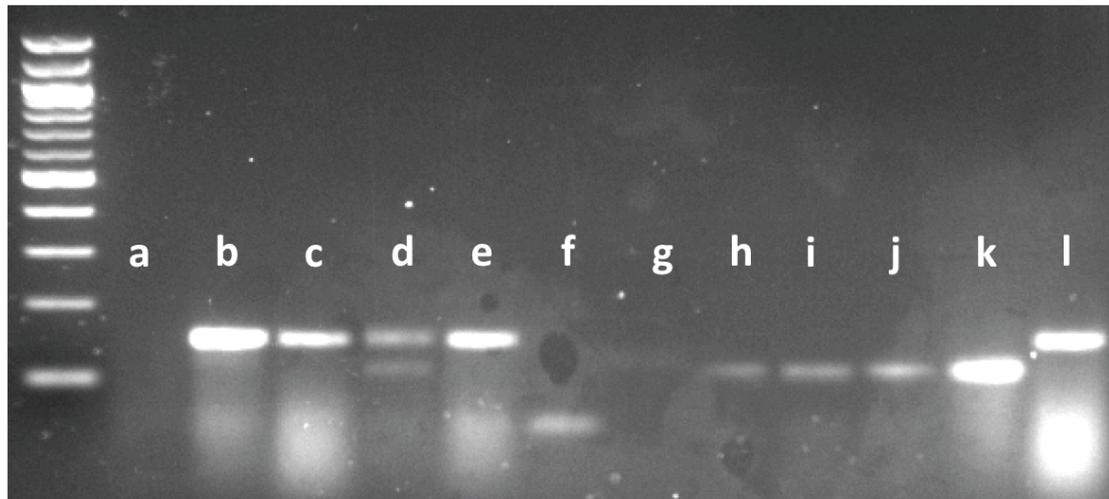
**Figure 6.6. “Next generation mapping” analysis of the Illumina data from the Ws-2 x HR-5 bulk susceptible RenSeq.** Alignment data was uploaded to <http://bar.utoronto.ca/NGM>, and analyzed. This plot shows the peak produced on chromosome 1 when a homozygous SNP “chastity” of between 85 and 100% is specified.

In order to establish if there are any NB-LRR encoding genes in HR-5 but absent from Ws-0 or Col-0 that co-segregate with *RAL4* in HR-5, I developed a simple informatics pipeline. Briefly, the HR-5 RenSeq reads were assembled using SPAdes (Bankevich et al., 2012). I then compiled a nucleotide blast database of 199 known NB-LRR or NB-LRR like- encoding genes from Col-0. I identified the highest identity and longest alignment length hits for each contig, thus identifying the most

likely allelic pairs of genes from HR-5 and Col-0. Contigs of length >1500 bp and covered at least 15x and that did not have a good hit (at least 80% identity) to a Col-0 NB-LRR encoding gene were investigated further. These contigs were blasted (megablast and discontinuous megablast) against the NCBI nucleotide collection, and were subjected to a GENESCAN (Burge and Karlin, 1997) ORF prediction and subsequently searched against the PFAM database (Finn et al., 2014). From this analysis I identified several novel NB-LRR encoding genes. These include 2 HR-5 specific TIR-NB-LRRs, 1 TIR-NB-LRR shared in Ws-2 and HR-5 and not Col-0, several putative partial genes (encoding 1 TIR, NB or LRR domain). The reads from the bulked-susceptible sample were aligned against the HR-5 'NB-LRRome' assembly, and each novel NB-LRR was checked for co-segregation with *RAL4*. Each was covered with many HR-5 derived reads, and the HR-5 NB-LRRs absent from Ws-2 were all heterozygous, indicating that they are not linked to *RAL4*.

#### 6.2.6 *RAL4* IMMUNITY CANNOT BE SUPPRESSED BY VIRULENT *ALBUGO LAIBACHII*

In previous chapters I have described the remarkable ability of *Al* to suppress resistance to various other pathogens in the context of *Al* pre-infection, including multiple *Hpa* resistances (Cooper et al., 2008). In order to test if virulent *Al* isolates can suppress *RAL4* mediated immunity, I developed a simple assay. I screened the genomes of the 5 *Al* isolates virulent on HR-5 for micro-deletions relative to Nc14. I discovered one such micro-deletion in Ash4, around which I designed PCR markers. The marker reveals a size difference in the PCR product between Nc14 and Ash4, and was therefore named AISSLP1. I then infected HR-5, along with the requisite controls, with Ash4 and after 10 days sprayed with Nc14. After a further 14 days DNA was carefully extracted (to avoid contamination) and AISSLP1 applied. I found that whilst PCR products for both Nc14 and Ash4 are found on pre-infected Col-0 plants (susceptible to both isolates), only Ash4 DNA could be amplified from the pre-infected HR-5 plants (figure 6.7). HR-5 plants sprayed with a mix of Nc14 and Ash4 spores at day 0 also showed only the Ash4 band. The experiment was repeated twice with the same result. I also tried a 4 day pre-infection which had the same outcome, suggesting that *RAL4* immunity cannot be suppressed by pre-infection with a virulent *Al* isolate.



**Figure 6.7. *RAL4* resistance cannot be suppressed by a virulent isolate, SSLP marker data.** HR-5 and Col-0 were inoculated with the virulent isolate Ash4 and after 10 days inoculated with Nc14 (avirulent on HR-5). To test if pre-infection with Ash4 could suppress HR-5/*RAL4* immunity and allow Nc14 growth, a PCR SSLP marker was applied after a further 14 days. The larger band is specific to Nc14, the smaller specific to Ash4. The lanes in the above agarose gel picture were loaded with the following: (a) Col-0: d0 H<sub>2</sub>O (b) Col-0: d0 H<sub>2</sub>O, d10 Nc14 (c) Col-0: d0 Nc14 (d) Col-0: d0 Ash4, d10 Nc14 (note double band indicating co-infection) (e) Col-0: d0 Nc14+Ash4 (f) HR-5: d0 H<sub>2</sub>O (g) HR-5: d0 Nc14 (h) HR-5: d0 Ash4 (i) HR-5: d0 Ash4, d10 Nc14 (j) HR-5: d0 Nc14+Ash4 (k) Ash4 DNA control (l) Nc14 DNA control. Ladder: 100 bp NEB DNA ladder.

### 6.3 DISCUSSION

In this chapter I described the mapping of a locus that confers resistance to HR-5 and probably recognition of *SSP16*<sup>Nc14</sup>. I found a 3:1 segregation of resistance to susceptibility in a cross between *Ws-2* and HR-5. This suggested a single dominant locus confers resistance. I then used the innovative technique of bombarding an F<sub>2</sub> population and testing the same plants for pathogen resistance. I observed a correlation between plants with abundant GUS sectors and susceptibility (and the reciprocal), although there were a small number of exceptions. This suggests that the same genetic locus confers both resistance to *A/Nc14* and recognition of *SSP16*<sup>Nc14</sup>.

Initially I used classical genetic mapping to define the *RAL4* locus in HR-5. The resistance mapped to a cluster of CC-NB-LRRs on chromosome 1, containing *RPP7* in Col-0 (Eulgem et al., 2007). This cluster of CC-NB-LRRs, known colloquially as the

*RPP7* cluster, is a complex locus that has apparently arisen through tandem duplication (Guo et al., 2011). *RPP7* itself is exceptional in terms of its regulation: a COPIA-R7 transposable element was recruited to *RPP7*'s first intron and the level of expression of the active splice variant of gene is regulated by the level of methylation at specific sites in the intron (Tsuchiya and Eulgem, 2013). This appears to be a unique feature of *RPP7*. *RPP7* and 5 other CC-NB-LRR encoding genes appear to be missing in HR-5, leaving 3 *RAL4* candidate genes. *RAC1*, the only cloned *R* gene against *Al*, encodes a TIR-NB-LRR. However various CC-NB-LRRs have been identified as sources of resistance to *Hpa* in *At* (Eulgem et al., 2007; McDowell et al., 1998; Bittner-Eddy et al., 2000) and other pathogens as reviewed in chapter 1. Interestingly, *RAL4* resistance functions independently of NDR1. The closest CC-NB-LRR relatives of the *RAL4* candidates (*RPP7*, *RPP8* and *RPP13*) also do not require NDR1 (McDowell et al., 2000; Bittner-Eddy et al., 2001). *RPP8* however, was shown to be ineffective in an *eds1/sid2* double mutant (Venugopal et al., 2009). *RAL4* function may therefore also be dependent on the products of these genes.

In *Agrobacterium*-mediated transient expression assays in *N. benthamiana* and *N. tabacum*, none of the *RAL4* candidates triggered a SSP16-dependent HR. The mechanism of recognition of pathogen effectors by CC-NB-LRRs is often via "indirect" recognition. For example, two CC-NB-LRR encoding genes *RPM1* and *RPS2* guard the RIN4 protein against modification by several bacterial effectors (Day et al., 2006). Another CC-NB-LRR, *RPS5*, is activated by the cleavage of its guardee, PBS1, by the effector AvrPphB at the plasma membrane (Qi et al., 2014). The effector putatively recognised by *RAL4*,  $\Delta$ SP-SSP16<sup>Nc14</sup> localised to the plasma membrane with an N-terminal tag (chapter 5). It is therefore possible that the reason that none of the *RAL4* candidates function in *Nicotiana spp* is that the guardee protein is either not present or substantially divergent in these species which are distantly related to *At*. It is also possible that due to selection pressure from *Al*, the signalling pathway for *RAL4* has become divergent compared to *Nicotiana spp*. *RPS5* requires its guardee PBS1 to trigger AvrPphB dependent HR in *Nicotiana* transient assays (Qi et al., 2014). As of yet, no signalling components required for *RPP7*, *RPP8* or *RPP13* have been identified, although there are

quantitative reductions in function in *eds1/ndr1* double mutants (McDowell et al., 2000).

It is also possible that there is a problem of protein accumulation, and in a further study epitope-tagged versions of both SSP16 and the RAL4 candidates will be tested in *Nicotiana spp.* and protein level checked by Western blotting.

In order to check if there are undiscovered novel candidates within the *RAL4* interval in HR-5, I developed and employed a *Brassicaceae* RenSeq method (Jupe et al., 2012; chapter 2b). Using this method in combination with next-generation mapping confirmed the *RAL4* mapping location. It also proved that the method could be used to quickly map multiple further *R* genes in future studies, since with the enrichment it was possible to multiplex 23 different samples and still achieve the high sequencing depth required to map to a high resolution. It also proved to be an effective method to identify accession specific novel NB-LRR encoding genes not in the Col-0 genome, but no further *RAL4* candidates were identified. Using these data also allowed me to prove that the putatively missing *RPP7/RAL4* cluster genes are truly absent in HR-5.

We must also consider that *RAL4* may not be a gene encoding a NB-LRR. Within the 83 genes also within the interval, there is one RLP (RLP9). The predicted protein product of this gene is structurally similar to several *R* genes cloned from tomato against *Cladosporium fulvum* and *Verticillium dahliae* consisting of a secretion signal, multiple leucine-rich repeats and a C-terminal transmembrane domain (Jones et al., 1994; Kawchuk et al., 2001). However since no genes of this class have been identified as encoding *R* genes in *At*, it is considered unlikely to be *RAL4*.

It has been previously established that *Al* is capable of suppressing resistance conferred by various R-proteins, and allow co-infection with avirulent races of *Hpa* (Cooper et al., 2008). Although Cooper et al (2008) showed that *Al* can suppress *RPP7* immunity against *Hpa*, the recognition and resistance of *Al* by a CC-NB-LRR suggests that *Al* is not as proficient at suppressing this resistance pathway. To test if this is indeed the case I set up a co-infection experiment. Nc14 (incompatible) was inoculated on HR-5 4 days after inoculation of Ash4 (compatible). Ten days

following the second infection, it was not possible to detect an Nc14 specific PCR product on HR-5. This suggests that Ash4 does not have the capability to suppress *RAL4* immunity to allow the growth of avirulent Nc14. In future work it may be enlightening to discover how *RAL4* signalling leads to resistance and how this differs from the currently described mechanisms of R protein mediated defence activation.

## CHAPTER 7: GENERAL DISCUSSION AND OUTLOOK

In the last twenty-five years, major progress has been made in the understanding of the genetic and molecular components that define the outcome of interactions between plants and pathogens. Numerous *R* genes and recognized effector (*Avr*) genes were genetically defined, allowing investigations of the physical mechanisms of recognition, activation of defense and co-evolutionary relationships between key host and pathogen components (Jones and Dangl, 2006; Dodds and Rathjen, 2010).

*Albugo* species infect an extensive range of hosts including many Brassicaceae crop species (Kamoun et al., 2014). Despite this, beyond detailed descriptions of the infection structures and life-cycle, until recently very little was known about the mechanisms by which *Albugo* sp. parasitize their hosts and how this is prevented by resistant plants. The species *Albugo candida* was re-organized into two distinct species, *Ac* an apparent generalist, and *Al* the Arabidopsis specialist (Thines et al., 2009).

The genome sequence of *Al* (Kemen et al., 2011) raised numerous questions. It encodes a new class of secreted effector candidates, the CHXCs. However, the low number (35) of these in the genome led to the questions, are these real, and are they the only *Al* effectors? Will the *Al* or *Ac* *Avr* genes encode CHXC effectors, or will they belong to another class of secreted proteins? The two reported *Al* genomes also showed a relatively low level of nucleotide diversity, but was this an artefact of the sequencing of only two isolates?

Previous studies of *AVR* genes in similar pathogens such as *Hpa* suggested that high allelic diversity should be expected at *Avr* gene loci, with these genes undergoing strong adaptive and balancing selection (Hall et al., 2009; Stukenbrock and McDonald, 2009). Advances in genome sequencing technology have rendered feasible the concept of identifying causal loci from population genome sequencing through association with polymorphisms. Would this be an effective method to identify pathogen *Avr* genes?

Although one TIR-NB-LRR encoding gene has been defined as an *Al* *R* gene (Borhan et al., 2004), important questions remain about *Al* resistance. For example would all *Al* *R* genes encode TIR-NB-LRRs? Can the remarkable defense suppression capabilities of *Al* (Cooper et al., 2008) extend to *R* genes against *Al*?

During my PhD, I focused on addressing these questions, and further questions that arose as the work progressed. In this discussion I will address the progress made in answering these questions and the new questions that my results have raised in the context of recent literature.

### 7.1 POPULATION VARIATION IN *ALBUGO LAIBACHII*

With the splitting of *Ac* into *Ac* and *Al*, both of which can parasitize *At* (Thines et al., 2009), the ecological question of which species is predominantly found in association with *At* in the field emerges. In chapter 3, I describe the collection of field isolates from various locations around the UK and one area in Germany. My results suggest that *Al* is the predominant *Albugo* species growing on *At* in these areas. It is difficult to be certain, as in order not to waste material, infection from collected samples was made in some cases directly onto *At* Col-0 which may have selected against *Ac* isolates growing on the plants in the field as the *Arabidopsis* accession Col-0 has a broad-spectrum *Ac* resistance gene *WRR4* (Borhan et al., 2008). Only the collection of additional field isolates from *At* can reveal the balance of *Al* and *Ac* infections in nature. I didn't further address the phenomenon of symptomless *Albugo* infection, as observed by Ploch and Thines (2011).

I developed a procedure to genetically identify the species of *Albugo* in a field isolate and to distinguish different *Al* isolates. Colleagues in the lab are already using these methods. For example Agathe Jouet identified the further field *At* isolates Wild 1, Wild 19A and Wild 43 which were mentioned in chapter 5. Her results support my hypothesis that *Al* is the dominant species on *At*, and that the isolates she collected are unique *Al* isolates.

Kemen et al, (2011) reported the genome sequences of two *Al* isolates. These isolates showed a low level of nucleotide diversity. This raised the question of the

degree of nucleotide diversity in the wider *Al* population. In chapter 4, I present the genome sequencing data for 4 further purified isolates. My results suggest that the level of diversity between Nc14 and Em1 is at the typical level between any two given isolates. How is such a low level of diversity maintained? These data could suggest that the *Al* population has recently experienced a population bottleneck effect, but the general heterogeneity of the *At* accessions for resistance, and the fact that most are susceptible suggests that this is probably not the case. On the other hand, very high spore mobility (one interpretation of the similarity of the German isolate to the UK ones) may mean that more competitive race-types are able to quickly dominate the population. Sequencing further samples could reveal deeper underlying diversity in the population.

My analysis of the genetic diversity between the 6 isolates did reveal signatures of recombination throughout the genomes of each isolate. This suggests that in nature, sexual recombination occurs between isolates, and that a future avenue for *Al* research might be to attempt to generate crosses between isolates, for example to help identify *Avr* genes.

## 7.2 ASSOCIATION GENOMICS TO IDENTIFY *AVR* GENE CANDIDATES

A major theme of my work has been developing and testing the utility of association genomics to identify candidate *Avr* genes. Although numerous papers have reported the identification of *AVR* genes using similar methods (Armstrong et al., 2005; Yoshida et al., 2009; de Jonge et al., 2012), none of these have attempted it with a high-throughput genome-wide method. I developed a bioinformatics pipeline to correlate non-synonymous polymorphisms predicted using Illumina data, described in chapter 2b. I showed that using this method on Illumina data for 5 *Hpa* races and race virulence data, it was possible to predict that *ATR1*, a well characterized *Hpa* *Avr* gene (Rehmany et al., 2005), would encode an effector recognized in the *At* accession Nd-1. *ATR5* (Bailey et al., 2011) was also one of 3 strong candidates for the *Avr* gene conditioning avirulence on the accession Ler-1 (where it is recognized by RPP5). More complicated genetic scenarios, involving the

recognition of multiple effectors by single accessions, however exposed the weakness of using relatively few isolates. For example *ATR13* (Allen et al., 2004) could not be predicted. It has been reported that oomycete pathogens are able to avoid recognition through the silencing of *Avr* genes (Qutob, Chapman and Gijzen, 2013). Such an event would be beyond the scope of my association genomics pipeline to detect. Nevertheless, I persevered with the method, and applied it to my own sequencing and virulence data for the 6 *Al* isolates.

Using the disease phenotype information and polymorphisms from 6 isolates, I predicted *Avr* gene candidates. However, in order to reduce the number of candidates, I used additional statistical tests. It has been widely reported that effector, and in particular *Avr*, encoding genes are generally under strong adaptive and balancing selection due to the co-evolutionary pressures of plant-pathogen interactions (Dawkins and Krebs, 1979; Stahl et al., 1999; Anderson et al., 2010). Using the pN/pS ratio (Yang, 2007) as a measure of adaptive selection and the linkage disequilibrium-based test called Fu's  $F_s$  (Fu, 1997) as a measure of balancing selection, I refined my list of candidate *Avr* genes. My results suggest that a subset of the secretome is far more diverse and evolving at a much more rapid rate than the majority of *Al* genes. As noted in chapter 4, 30% of the non-synonymous changes observed in the secretome were found in 31 out of 929 genes. Several CHXC class effectors were found in this group of rapidly evolving secreted protein encoding genes, strengthening the case that at least some of them are effectors. Genome density plotting was consistent with some of these rapidly evolving genes occupying more gene-sparse regions as reported in *Phytophthora* spp. by Raffaele et al, (2010b). Plotting the coordinates of non-synonymous changes along the length of the rapidly evolving effectors suggested that they are undergoing stronger adaptive selection towards the 3' end, similar to results found in *Pi* effectors by Win et al, (2007). Considered together, these data are consistent with a limited number of effectors being under strong evolutionary pressure to either evade recognition or to co-evolve with their host targets.

### 7.3 DISCOVERY OF RECOGNIZED EFFECTORS FROM *ALBUGO LAIBACHII*

Testing of several candidate *Avr* genes in chapter 5 revealed that two of them, *SSP16* and *SSP18* seem to be recognized in the accessions HR-5 and Ksk-1 respectively. Both of the predicted secreted proteins do not belong to the CHXC or RXLR classes of effectors, but do show extreme levels of amino-acid level diversity relative to most *Al* genes. I showed that the predicted virulent alleles of *SSP16* are able to quantitatively evade recognition in HR-5 in GUS eclipse experiments and do not compromise *PstDC3000* growth in strains that carry *SSP16*-delivering EDV constructs. The virulent allele of *SSP18* showed full evasion of recognition by Ksk-1 in both GUS eclipse and *PstDC3000* growth experiments. Additionally, I showed that the highly polymorphic *SSP16* C-terminus alone was sufficient to cause a GUS eclipse in HR-5. Following the discoveries of ATR13, ATR1 and *Avr3a*, Rehmany et al, (2005) quickly noticed that they shared a common RXLR motif in their N-terminal post-signal peptide cleavage region, and were part of a large family of related genes. However, I was able to find no evidence of such a large gene family related to *SSP16* or *SSP18* in *Al*, or any other species whose genome sequence was available to me. A major question directly arising from the apparent cytoplasmic recognition of these effectors is: how do they enter plant cells in a natural infection? There has been much speculation about the mechanism of uptake of RXLR, CRN and CHXC type effectors, and whether these N-terminal motifs are indeed required for uptake or may be involved in host cell targeting rather than uptake (Petre and Kamoun, 2014). The fact remains however that there are numerous fungal effectors for which there is evidence of host cell uptake, such as *Melampsora lini* *AvrM* (Rafiqi et al., 2010), *Uromyces fabae* RTP1 (Kemen et al., 2005), *Ustilago maydis* Cmu1 (Djamei et al., 2011) and various *Magnaporthe oryzae* effectors (Giraldo et al, 2013) where the nature of a host-cell translocation motif is unclear. Therefore it is plausible that *SSP16* and *SSP18* naturally enter infected host cells, like various fungal effectors, via an unknown mechanism.

At this stage I have no data about the possible virulence effects of these two proteins, beyond their sub-cellular localizations. To determine if they have a virulence effect, *p35S::At* transgenic lines have been generated and will be tested for enhanced susceptibility or resistance to various pathogens.

### 7.3 *RAL4*; A GENETIC LINK BETWEEN RECOGNITION AND RESISTANCE

After the discovery of two apparently recognized effectors, I needed a way to confirm if they were *Avr* genes. Since the resistance to *AI* in HR-5 segregated 3:1, I hypothesized that a single dominant *R* gene conditioned resistance. I reasoned that if *SSP16* was indeed the Nc14 *Avr* gene then  $F_2$  individuals that recognized it should also be resistant to Nc14. In chapter 6, to test this hypothesis, I conducted GUS-eclipse experiments on  $F_2$  plants followed by pathogen infection. I found that almost all the plants that didn't recognize *SSP16* were also susceptible to Nc14. Considering the overall variability of the GUS eclipse method, my results suggest that *SSP16* can be tentatively considered as the *Avr* gene of Nc14 on HR-5. Unfortunately, there is no established protocol to transform or cross *AI* isolates. Therefore it is difficult to imagine a direct method to test this hypothesis.

I then mapped the *RAL4* locus to the previously identified *RPP7* (Eulgem et al., 2007) cluster on *At* chromosome 1. In the *RAL4*-containing accession, HR-5, 6 of the 9 putative NB-LRRs in the mapping interval are absent. This leaves 3 clear candidate genes for *RAL4*. All of these genes encode CC-NB-LRR type genes. In *A. tumefaciens* transient assays in *N. benthamiana*, none of these candidates triggered cell death upon co-expression with *SSP16*. There are various hypotheses discussed in chapter 6 as to why this might be the case, but the ultimate test will come when stable *At* transgenic complementation lines for each candidate can be tested with Nc14 and *SSP16*.

In order to verify the lack of other NB-LRR candidates in the mapping interval, I undertook an *R* gene enrichment sequencing (RenSeq) experiment (Jupe et al., 2012). I sequenced the Col-0, Ws-2, HR-5 and HR-5 x Ws-2 bulked Nc14 susceptible NB-LRRomes with Illumina MiSeq technology. These results confirmed the location of the *RAL4* locus through next-generation mapping (Austin et al., 2011). The assembly of HR-5 reads, and checks of linkage using the bulked susceptible reads revealed that HR-5 harbors several non- Col-0 NB-LRR encoding genes, but none of these were linked to *RAL4*. My results suggest that the RenSeq technique could be

useful for mapping and identifying novel NB-LRRs from other *At* accessions. Nevertheless, I cannot exclude that *RAL4* is one of the numerous non-NB-LRR encoding genes within the mapping interval.

Although the identity of *RAL4* is still uncertain, it is notable that the CC-NB-LRRs RPM1, RPS2 (Day, Dahlbeck and Staskawicz, 2006) and RPS5 (Qi et al., 2014) guard proteins that are localized at the plasma membrane. According to biochemical fractionation RPP13, closely related to the *RAL4* candidates, also associates with the plasma membrane (Leonelli, 2011). As shown in chapter 5, GFP-SSP16 localizes to the plasma membrane. Therefore it is tempting to speculate that SSP16 could target an unknown host target at the plasma membrane, which is guarded by *RAL4* (figure 7.1).

*Al* is capable of suppressing resistance conferred by TIR-NB-LRRs and CC-NB-LRRs to allow co-infection with avirulent races of *Hpa* (Cooper et al., 2008). I tested if this applied to *RAL4*. My results suggest that *RAL4* immunity could not be suppressed by the pre-infection of a virulent *Al* isolate. This indicates that *RAL4* may have evolved to be specifically resilient against perturbation by *Al* effectors, which in the light of the co-evolutionary battle between these organisms is logical. Although *RAL4* putatively encodes a CC-NB-LRR paralog of *RPP7*, Cooper et al, (2008) report that *RPP7* can be suppressed by *Al* infection, and resolution of this inconsistency requires the final cloning of *RAL4*. Similar to *RPP7*, *RPP8* and *RPP13* (McDowell et al., 2000; Bittner-Eddy et al., 2001), *RAL4* is NDR-1 independent. Elucidation of the components involved in *RAL4* immunity might lead to understanding novel pathways in plant immunity that could be usefully applied in the future.

The basis of the recognition of SSP18 in Ksk-1 was not addressed in this thesis. Given that it is recognized in Ksk-1, I speculate that it is recognized by the R protein encoded by either *RAC1* or *RAC3* (Borhan et al., 2001; Borhan et al., 2004).

#### 7.4 CONCLUSIONS AND OUTLOOK

I present data on the sequencing of 4 and comparison of 6 *Al* isolates, and the use of these data to identify two novel recognised effectors from *Al*. These effectors are the first of their kind discovered in oomycete plant pathogens. I have mapped the

recognition of one of these effectors to an *AI* “suppression-proof” locus that contains 3 CC-NB-LRRs. If a CC-NB-LRR is encoded by *RAL4*, this will be the first of its type identified for *Albugo* resistance. Figure 7.1 incorporates my findings into the existing literature.

My work demonstrates the utility of combined population genomic analyses to predict candidate genes for encoding recognised effectors under strong selection pressures, and enhances our knowledge of the population biology of *AI* and its effectors and the mechanism of resistance to *AI* in *At*. The technological developments of the last few years, and some good fortune, have enabled rapid progress within the scope of a single PhD that would have been unimaginable 5 years previously.

Future work will need to confirm the basis of the recognition of these effectors, and elucidate whether they have virulence effects and targets. While it remains to be seen exactly how my findings could be translated directly into benefits in economically and agronomically important systems, it has provided a substantial methodological contribution, showing that association genomics combined with sophisticated tools to measure adaptive and balancing selection, enable strong effector candidates to be discerned. In view of the potential importance of effector-guided breeding (Vleeshouwers and Oliver, 2014), these methods are likely to be of widespread utility in analysis of the genomes of crop pathogens, and this information could be used to underpin crop breeding.

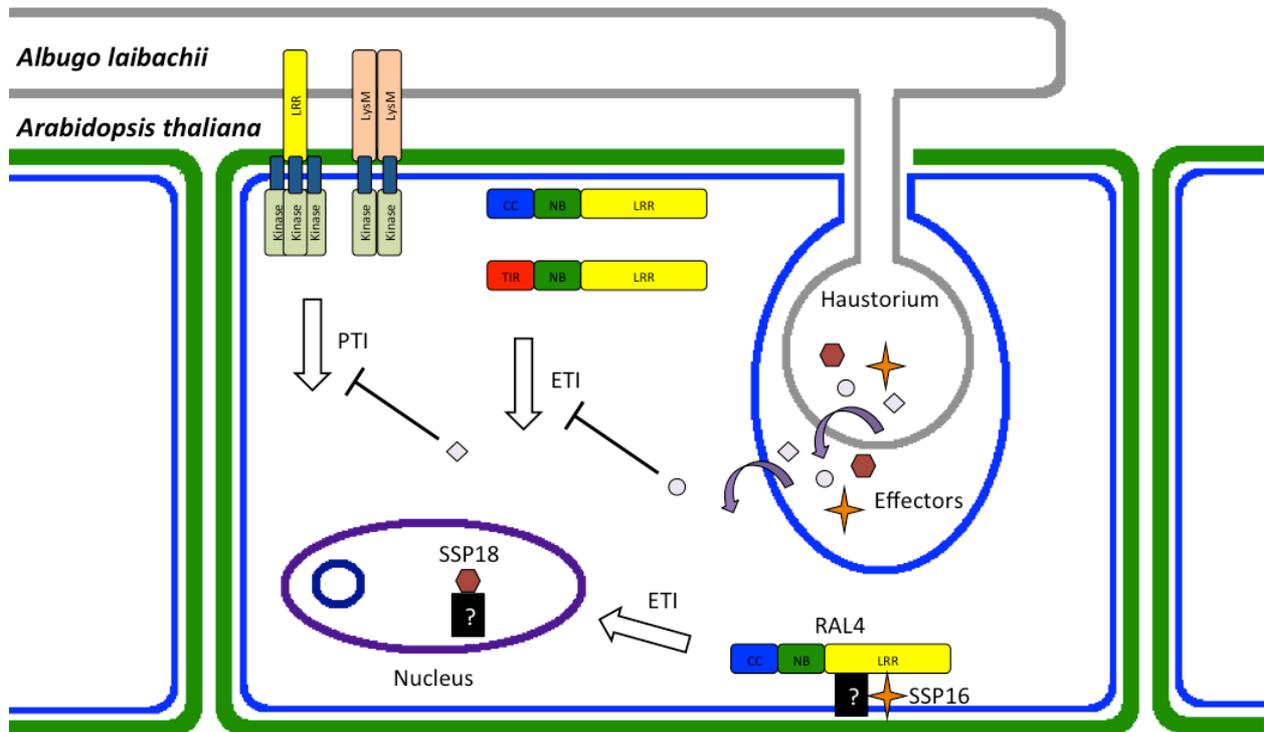
This work could be argued to sit squarely within the current conceptual model of the field, as developed by Jones and Dangl, (2006). This essay established the ‘zig-zag-zig’ model of plant immunity where: 1. Pathogens are detected and plants activate PTI. 2. Secreted pathogen effectors suppress this immunity. 3. Plant R proteins recognise these effectors or their activities and activate ETI. 4. Effectors either evade detection through polymorphism (leading to elevated diversity of R proteins and effectors) or other effectors suppress the activation of defence triggered by this recognition (ETS). 5. A perpetual co-evolutionary struggle between R proteins, effectors and effector targets ensues. This leads to all kinds of

extraordinary evolutionary inventions, including decoys (van der Hoorn and Kamoun, 2008) and guard/guardee relationships.

Although my findings are undoubtedly important in the development of our understanding of *Albugo* spp. and do have implications for the future discovery of recognised effectors from population genomic data, these data do not enable one to derive a significant conceptual advance. A pessimistic way to sum up my work could be, “more recognised effectors and more R proteins”. The optimistic future developments in the project that I propose will probably only lead to more effector targets/guardees/decoys that will confirm the current conceptual model. Is this a problem rooted in the approach that I have taken from the beginning? My strategy was based around the very conceptual model I have described; I hoped to find what I have found.

In the defence of my work, it could be argued that every effector or R protein study conducted since Jones and Dangl (2006) has confirmed concepts therein proposed (or within the decoy hypothesis). Though we continue to accumulate knowledge of specific interactions, amassing myriad effectors, R proteins, targets, guardees, decoys and regulators, we do not advance our conceptual understanding to a higher plane. In spite of our best intentions, for example examining an understudied genus like *Albugo* in a search for novelty, we return to the same kinds of answers. From where will the next conceptual advance emerge? Perhaps the answer is to ask different questions.

Nevertheless, we shouldn't lose sight of our other goal (after making conceptual advances), which is rooted in the fact that plants are the source of our food. The better understanding of the relationship between effectors, their targets and R proteins in both the mechanistic and co-evolutionary sense are key to sustainable mitigation of the damage done to crop plants by pathogens.



**Figure 7.1. Tentative general model of *Albugo laibachii* interaction with infected *Arabidopsis* cells.** *Al* invaginates a haustorium into the host cell and secretes effectors. Some of these effectors translocate into the host cell and suppress immunity. SSP18, localizes to the nucleus, presumably to carry out its virulence activity, but in Ksk-1 it is recognized by an R protein that triggers ETI. SSP16 localizes to the cell membrane presumably to carry out its virulence activity, but in HR-5 it is recognized by a RAL4 encoded CC-NB-LRR, triggering an ETI that *Al* cannot suppress.

## 9. LIST OF REFERENCES

- Adhikari, T.B., Liu, J.Q., Mathur, S., Wu, C.R.X., and Rimmer, S.R.** (2003). Genetic and molecular analyses in crosses of race 2 and race 7 of *Albugo candida*. *Phytopathology* **93**, 959-965.
- Adl, S.M., Simpson, A.G., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser, S.S., Brugerolle, G., Fensome, R.A., Fredericq, S., James, T.Y., Karpov, S., Kugrens, P., Krug, J., Lane, C.E., Lewis, L.A., Lodge, J., Lynn, D.H., Mann, D.G., McCourt, R.M., Mendoza, L., Moestrup, O., Mozley-Standridge, S.E., Nerad, T.A., Shearer, C.A., Smirnov, A.V., Spiegel, F.W., and Taylor, M.F.** (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* **52**, 399-451.
- Alfano, J.R., and Collmer, A.** (2004). Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annu Rev Phytopathol* **42**, 385-414.
- Allen, R.L., Bittner-Eddy, P.D., Grenville-Briggs, L.J., Meitz, J.C., Rehmany, A.P., Rose, L.E., and Beynon, J.L.** (2004). Host-parasite coevolutionary conflict between *Arabidopsis* and downy mildew. *Science* **306**, 1957-1960.
- Anderson, J.P., Gleason, C.A., Foley, R.C., Thrall, P.H., Burdon, J.B., and Singh, K.B.** (2010). Plants versus pathogens: an evolutionary arms race. *Funct Plant Biol* **37**, 499-512.
- Armstrong, M.R., Whisson, S.C., Pritchard, L., Bos, J.I., Venter, E., Avrova, A.O., Rehmany, A.P., Böhme, U., Brooks, K., Cherevach, I., Hamlin, N., White, B., Fraser, A., Lord, A., Quail, M.A., Churcher, C., Hall, N., Berriman, M., Huang, S., Kamoun, S., Beynon, J.L., and Birch, P.R.** (2005). An ancestral oomycete locus contains late blight avirulence gene *Avr3a*, encoding a protein that is recognized in the host cytoplasm. *Proc Natl Acad Sci U S A* **102**, 7766-7771.
- Asai, S., Ohta, K., and Yoshioka, H.** (2008). MAPK signaling regulates nitric oxide and NADPH oxidase-dependent oxidative bursts in *Nicotiana benthamiana*. *Plant Cell* **20**, 1390-1406.
- Asai, T., Tena, G., Plotnikova, J., Willmann, M., Chiu, W., Gomez-Gomez, L., Boller, T., Ausubel, F., and Sheen, J.** (2002). MAP kinase signalling cascade in *Arabidopsis* innate immunity. *Nature* **415**, 977-983.
- Austin, R.S., Vidaurre, D., Stamatiou, G., Breit, R., Provart, N.J., Bonetta, D., Zhang, J., Fung, P., Gong, Y., Wang, P.W., McCourt, P., and Guttman, D.S.** (2011). Next-generation mapping of *Arabidopsis* genes. *Plant J* **67**, 715-725.
- Awasthi, R.P., Nashaat, N.I., Kolte, S.J., Tewari, A.K., Meena, P.D., and Renu, B.** (2012). Screening of putative resistant sources against Indian and exotic isolates of *Albugo candida* inciting white rust in rapeseed-mustard. *Journal of Oilseed Brassica* **3**, 27-37.
- Bailey, K., Cevik, V., Holton, N.J., Byrne-Richardson, J., Sohn, K.H., Coates, M., Woods-Tör, A., Aksoy, H.M., Hughes, L., Baxter, L., Jones, J.D., Beynon, J., Holub, E.B., and Tör, M.** (2011). Molecular cloning of *ATR5Emoy2* from *Hyaloperonospora arabidopsidis*, an avirulence determinant that triggers RPP5-mediated defense in *Arabidopsis*. *Mol Plant Microbe Interact*.
- Bains, S.S., and Jhooty, J.S.** (1985). Association of *Peronospora parasitica* with *Albugo candida* on brassica-juncea leaves. *Phytopathologische Zeitschrift-Journal of Phytopathology* **112**, 28-31.
- Bakker, E.G., Toomajian, C., Kreitman, M., and Bergelson, J.** (2006). A genome-wide survey of R gene polymorphisms in *Arabidopsis*. *Plant Cell* **18**, 1803-1818.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., and Pevzner, P.A.** (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**,

455-477.

- Baxter, L., Tripathy, S., Ishaque, N., Boot, N., Cabral, A., Kemen, E., Thines, M., Ah-Fong, A., Anderson, R., Badejoko, W., Bittner-Eddy, P., Boore, J.L., Chibucos, M.C., Coates, M., Dehal, P., Delehaunty, K., Dong, S., Downton, P., Dumas, B., Fabro, G., Fronick, C., Fuerstenberg, S.I., Fulton, L., Gaulin, E., Govers, F., Hughes, L., Humphray, S., Jiang, R.H., Judelson, H., Kamoun, S., Kyung, K., Meijer, H., Minx, P., Morris, P., Nelson, J., Phuntumart, V., Qutob, D., Rehmany, A., Rougon-Cardoso, A., Ryden, P., Torto-Alalibo, T., Studholme, D., Wang, Y., Win, J., Wood, J., Clifton, S.W., Rogers, J., Van den Ackerveken, G., Jones, J.D., McDowell, J.M., Beynon, J., and Tyler, B.M.** (2010). Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* **330**, 1549-1551.
- Bhattacharjee, S., Hiller, N.L., Liolios, K., Win, J., Kanneganti, T.D., Young, C., Kamoun, S., and Haldar, K.** (2006). The malarial host-targeting signal is conserved in the Irish potato famine pathogen. *PLoS Pathog* **2**, e50.
- Bittner-Eddy, P.D., and Beynon, J.L.** (2001). The Arabidopsis downy mildew resistance gene, RPP13-Nd, functions independently of NDR1 and EDS1 and does not require the accumulation of salicylic acid. *Mol Plant Microbe Interact* **14**, 416-421.
- Boch, J., and Bonas, U.** (2010). *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* **48**, 419-436.
- Bonants, P., deWeerd, M.H., van Gent-Pelzer, M., Lacourt, I., Cooke, D., and Duncan, J.** (1997). Detection and identification of *Phytophthora fragariae* Hickman by the polymerase chain reaction. *European Journal of Plant Pathology* **103**, 345-355.
- Borhan, M.H., Brose, E., Beynon, J.L., and Holub, E.B.** (2001). White rust (*Albugo candida*) resistance loci on three Arabidopsis chromosomes are closely linked to downy mildew (*Peronospora parasitica*) resistance loci. *Mol Plant Pathol* **2**, 87-95.
- Borhan, M.H., Holub, E.B., Beynon, J.L., Rozwadowski, K., and Rimmer, S.R.** (2004). The Arabidopsis TIR-NB-LRR gene RAC1 confers resistance to *Albugo candida* (white Rust) and is dependant on EDS1 but not PAD4. *Molecular Plant-Microbe Interactions* **17**, 711-719.
- Borhan, M.H., Holub, E.B., Kindrachuk, C., Omid, M., Bozorgmanesh-Frad, G., and Rimmer, S.R.** (2010). WRR4, a broad-spectrum TIR-NB-LRR gene from *Arabidopsis thaliana* that confers white rust resistance in transgenic oilseed brassica crops. *Molecular Plant Pathology* **11**, 283-291.
- Borhan, M.H., Gunn, N., Cooper, A., Gulden, S., Tör, M., Rimmer, S.R., and Holub, E.B.** (2008). WRR4 encodes a TIR-NB-LRR protein that confers broad-spectrum white rust resistance in Arabidopsis thaliana to four physiological races of *Albugo candida*. *Mol Plant Microbe Interact* **21**, 757-768.
- Borson, N.D., Salo, W.L., and Drewes, L.R.** (1992). A lock-docking oligo(dT) primer for 5' and 3' RACE PCR. *PCR Methods Appl* **2**, 144-148.
- Botella, M.A., Parker, J.E., Frost, L.N., Bittner-Eddy, P.D., Beynon, J.L., Daniels, M.J., Holub, E.B., and Jones, J.D.** (1998). Three genes of the Arabidopsis RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. *Plant Cell* **10**, 1847-1860.
- Botton, S.A., Pereira, D.I., Costa, M.M., Azevedo, M.I., Argenta, J.S., Jesus, F.P., Alves, S.H., and Santurio, J.M.** (2011). Identification of *Pythium insidiosum* by Nested PCR in Cutaneous Lesions of Brazilian Horses and Rabbits. *Curr Microbiol* **62**, 1225-1229.
- Boudsocq, M., Willmann, M.R., McCormack, M., Lee, H., Shan, L., He, P., Bush, J., Cheng, S.H., and Sheen, J.** (2010). Differential innate immune signalling via Ca(2+) sensor protein kinases. *Nature* **464**, 418-422.
- Brown, J.K., and Tellier, A.** (2011). Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annu Rev Phytopathol* **49**, 345-367.
- Burch-Smith, T.M., Schiff, M., Caplan, J.L., Tsao, J., Czymmek, K., and Dinesh-Kumar, S.P.**

- (2007). A novel role for the TIR domain in association with pathogen-derived elicitors. *PLoS Biol* **5**, e68.
- Burge, C., and Karlin, S.** (1997). Prediction of complete gene structures in human genomic DNA. *J Mol Biol* **268**, 78-94.
- Caillaud, M.C., Asai, S., Rallapalli, G., Piquerez, S., Fabro, G., and Jones, J.D.** (2013). A downy mildew effector attenuates salicylic acid-triggered immunity in *Arabidopsis* by interacting with the host mediator complex. *PLoS Biol* **11**, e1001732.
- Caillaud, M.C., Piquerez, S.J., Fabro, G., Steinbrenner, J., Ishaque, N., Beynon, J., and Jones, J.D.** (2012). Subcellular localization of the Hpa RxLR effector repertoire identifies a tonoplast-associated protein HaRxL17 that confers enhanced plant susceptibility. *Plant J* **69**, 252-265.
- Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., Wang, X., Ott, F., Müller, J., Alonso-Blanco, C., Borgwardt, K., Schmid, K.J., and Weigel, D.** (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* **43**, 956-963.
- Century, K.S., Shapiro, A.D., Repetti, P.P., Dahlbeck, D., Holub, E., and Staskawicz, B.J.** (1997). NDR1, a pathogen-induced component required for *Arabidopsis* disease resistance. *Science* **278**, 1963-1965.
- Chen, Y., Liu, Z., and Halterman, D.A.** (2012). Molecular determinants of resistance activation and suppression by *Phytophthora infestans* effector IPI-O. *PLoS Pathog* **8**, e1002595.
- Cheng, Y.T., Germain, H., Wiermer, M., Bi, D., Xu, F., García, A.V., Wirthmueller, L., Després, C., Parker, J.E., Zhang, Y., and Li, X.** (2009). Nuclear pore complex component MOS7/Nup88 is required for innate immunity and nuclear accumulation of defense regulators in *Arabidopsis*. *Plant Cell* **21**, 2503-2516.
- Choi, Y.J., Shin, H.D., and Thines, M.** (2009). The host range of *Albugo candida* extends from Brassicaceae through Cleomaceae to Capparaceae. *Mycological Progress* **8**, 329-335.
- Cingolani, P., Platts, A., Wang, I.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M.** (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92.
- Clarke, C.R., Chinchilla, D., Hind, S.R., Taguchi, F., Miki, R., Ichinose, Y., Martin, G.B., Leman, S., Felix, G., and Vinatzer, B.A.** (2013). Allelic variation in two distinct *Pseudomonas syringae* flagellin epitopes modulates the strength of plant immune responses but not bacterial motility. *New Phytol* **200**, 847-860.
- Clough, S.J., and Bent, A.F.** (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**, 735-743.
- Coates, M., Beynon, J., VanAlfen, N., Bruening, G., and Leach, J.** (2010). *Hyaloperonospora arabidopsidis* as a Pathogen Model. *Annual Review of Phytopathology*, Vol 48 **48**, 329-345.
- Coll, N.S., Epple, P., and Dangl, J.L.** (2011). Programmed cell death in the plant immune system. *Cell Death Differ* **18**, 1247-1256.
- Cooper, A., Latunde-Dada, A., Woods-Tör, A., Lynn, J., Lucas, J., Crute, I., and Holub, E.** (2008). Basic compatibility of *Albugo candida* in *Arabidopsis thaliana* and *Brassica juncea* causes broad-spectrum suppression of innate immunity. *Mol Plant Microbe Interact* **21**, 745-756.
- Crute, I., Beynon, J., Dangl, J., Holub, E., Mauch-Mani, B., Slusarenko, A., Staskawicz, B., and Ausubel, F.** (1994). Microbial pathogenesis of *Arabidopsis*. In *Arabidopsis* (Cold Spring Harbor Laboratory Press), pp. 705- 747.
- Dangl, J.L., and Jones, J.D.** (2001). Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826-833.

- Dawkins, R., and Krebs, J.R.** (1979). Arms races between and within species. *Proc R Soc Lond B Biol Sci* **205**, 489-511.
- Day, B., Dahlbeck, D., and Staskawicz, B.J.** (2006). NDR1 interaction with RIN4 mediates the differential activation of multiple disease resistance pathways in Arabidopsis. *Plant Cell* **18**, 2782-2791.
- de Jonge, R., van Esse, H.P., Maruthachalam, K., Bolton, M.D., Santhanam, P., Saber, M.K., Zhang, Z., Usami, T., Lievens, B., Subbarao, K.V., and Thomma, B.P.** (2012). Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc Natl Acad Sci U S A* **109**, 5110-5115.
- Deslandes, L., Olivier, J., Peeters, N., Feng, D.X., Khounlotham, M., Boucher, C., Somssich, I., Genin, S., and Marco, Y.** (2003). Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proc Natl Acad Sci U S A* **100**, 8024-8029.
- Djamei, A., Schipper, K., Rabe, F., Ghosh, A., Vincon, V., Kahnt, J., Osorio, S., Tohge, T., Fernie, A.R., Feussner, I., Feussner, K., Meinicke, P., Stierhof, Y.D., Schwarz, H., Macek, B., Mann, M., and Kahmann, R.** (2011). Metabolic priming by a secreted fungal effector. *Nature* **478**, 395-398.
- Dodds, P.N., and Rathjen, J.P.** (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet* **11**, 539-548.
- Dodds, P.N., Lawrence, G.J., Catanzariti, A.M., Teh, T., Wang, C.I., Ayliffe, M.A., Kobe, B., and Ellis, J.G.** (2006). Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc Natl Acad Sci U S A* **103**, 8888-8893.
- Dong, S., Stam, R., Cano, L.M., Song, J., Sklenar, J., Yoshida, K., Bozkurt, T.O., Oliva, R., Liu, Z., Tian, M., Win, J., Banfield, M.J., Jones, A.M., van der Hoorn, R.A., and Kamoun, S.** (2014). Effector specialization in a lineage of the Irish potato famine pathogen. *Science* **343**, 552-555.
- Dubiella, U., Seybold, H., Durian, G., Komander, E., Lassig, R., Witte, C.P., Schulze, W.X., and Romeis, T.** (2013). Calcium-dependent protein kinase/NADPH oxidase activation circuit is required for rapid defense signal propagation. *Proc Natl Acad Sci U S A* **110**, 8744-8749.
- Eitas, T.K., and Dangl, J.L.** (2010). NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Curr Opin Plant Biol* **13**, 472-477.
- Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S.** (2009). Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS One* **4**, e5553.
- Eulgem, T., and Somssich, I.E.** (2007). Networks of WRKY transcription factors in defense signaling. *Curr Opin Plant Biol* **10**, 366-371.
- Eulgem, T., Tsuchiya, T., Wang, X.J., Beasley, B., Cuzick, A., Tör, M., Zhu, T., McDowell, J.M., Holub, E., and Dangl, J.L.** (2007). EDM2 is required for RPP7-dependent disease resistance in Arabidopsis and affects RPP7 transcript levels. *Plant J* **49**, 829-839.
- Fabro, G., Steinbrenner, J., Coates, M., Ishaque, N., Baxter, L., Studholme, D.J., Körner, E., Allen, R.L., Piquerez, S.J., Rougon-Cardoso, A., Greenshields, D., Lei, R., Badel, J.L., Caillaud, M.C., Sohn, K.H., Van den Ackerveken, G., Parker, J.E., Beynon, J., and Jones, J.D.** (2011). Multiple candidate effectors from the oomycete pathogen *Hyaloperonospora arabidopsidis* suppress host plant immunity. *PLoS Pathog* **7**, e1002348.
- Fan, J., Crooks, C., and Lamb, C.** (2008). High-throughput quantitative luminescence assay of the growth in planta of *Pseudomonas syringae* chromosomally tagged with *Photobacterium luminescens* luxCDABE. *Plant J* **53**, 393-399.

- Felix, G., Duran, J.D., Volko, S., and Boller, T.** (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J* **18**, 265-276.
- Ferreira, M.A., and Purcell, S.M.** (2009). A multivariate test of association. *Bioinformatics* **25**, 132-133.
- Feys, B.J., Wiermer, M., Bhat, R.A., Moisan, L.J., Medina-Escobar, N., Neu, C., Cabral, A., and Parker, J.E.** (2005). Arabidopsis SENESCENCE-ASSOCIATED GENE101 stabilizes and signals within an ENHANCED DISEASE SUSCEPTIBILITY1 complex in plant innate immunity. *Plant Cell* **17**, 2601-2613.
- Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L., Tate, J., and Punta, M.** (2014). Pfam: the protein families database. *Nucleic Acids Res* **42**, D222-230.
- Flor, H.H.** (1971). Current Status of the Gene-For-Gene Concept. *Annual Review of Phytopathology* **9**, 275-296.
- Franchi, L., Warner, N., Viani, K., and Nuñez, G.** (2009). Function of Nod-like receptors in microbial recognition and host defense. *Immunol Rev* **227**, 106-128.
- Frank, S.A.** (1992). Models of plant-pathogen coevolution. *Trends Genet* **8**, 213-219.
- Fu, Y.X.** (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915-925.
- Galesloot, T.E., van Steen, K., Kiemeny, L.A., Janss, L.L., and Vermeulen, S.H.** (2014). A comparison of multivariate genome-wide association methods. *PLoS One* **9**, e95923.
- Galán, J.E., and Collmer, A.** (1999a). Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284**, 1322-1328.
- Gardiner A, McMullan, M., Bailey, K., Kemen, E., Ward, B., Cevik, V., Robert-Seilaniantz, A., Schultz-Larsen, T., Balmuth, AL., Holub, E., van Osterhout, C., Jones, J.D.G.** (2015) Immunosuppression enables expanded host ranges and can explain mosaic genome structures in *Albugo candida* races. *eLife*.
- Gan, X., Stegle, O., Behr, J., Steffen, J.G., Drewe, P., Hildebrand, K.L., Lyngsoe, R., Schultheiss, S.J., Osborne, E.J., Sreedharan, V.T., Kahles, A., Bohnert, R., Jean, G., Derwent, P., Kersey, P., Belfield, E.J., Harberd, N.P., Kemen, E., Toomajian, C., Kover, P.X., Clark, R.M., Rättsch, G., and Mott, R.** (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* **477**, 419-423.
- Gaulin, E., Madoui, M.A., Bottin, A., Jacquet, C., Mathé, C., Couloux, A., Wincker, P., and Dumas, B.** (2008). Transcriptome of *Aphanomyces euteiches*: new oomycete putative pathogenicity factors and metabolic pathways. *PLoS One* **3**, e1723.
- Geng, X., Jin, L., Shimada, M., Kim, M.G., and Mackey, D.** (2014). The phytotoxin coronatine is a multifunctional component of the virulence armament of *Pseudomonas syringae*. *Planta*.
- Gilroy, E.M., Breen, S., Whisson, S.C., Squires, J., Hein, I., Kaczmarek, M., Turnbull, D., Boevink, P.C., Lokossou, A., Cano, L.M., Morales, J., Avrova, A.O., Pritchard, L., Randall, E., Lees, A., Govers, F., van West, P., Kamoun, S., Vleeshouwers, V.G., Cooke, D.E., and Birch, P.R.** (2011). Presence/absence, differential expression and sequence polymorphisms between PiAVR2 and PiAVR2-like in *Phytophthora infestans* determine virulence on R2 plants. *New Phytol* **191**, 763-776.
- Giraldo, M.C., Dagdas, Y.F., Gupta, Y.K., Mentlak, T.A., Yi, M., Martinez-Rocha, A.L., Saitoh, H., Terauchi, R., Talbot, N.J., and Valent, B.** (2013). Two distinct secretion systems facilitate tissue invasion by the rice blast fungus *Magnaporthe oryzae*. *Nat Commun* **4**, 1996.
- Glazebrook, J.** (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu Rev Phytopathol* **43**, 205-227.
- Glez-Peña, D., Gómez-López, G., Reboiro-Jato, M., Fdez-Riverola, F., and Pisano, D.G.** (2011). PileLine: a toolbox to handle genome position information in next-

generation sequencing studies. *BMC Bioinformatics* **12**, 31.

- Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D.B., Lander, E.S., and Nusbaum, C.** (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* **27**, 182-189.
- Goritschnig, S., Krasileva, K.V., Dahlbeck, D., and Staskawicz, B.J.** (2012). Computational prediction and molecular characterization of an oomycete effector and the cognate Arabidopsis resistance gene. *PLoS Genet* **8**, e1002502.
- Grant, M.R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Innes, R.W., and Dangl, J.L.** (1995). Structure of the Arabidopsis RPM1 gene enabling dual specificity disease resistance. *Science* **269**, 843-846.
- Grooters, A.M.** (2003). Pythiosis, lagenidiosis, and zygomycosis in small animals. *Vet Clin North Am Small Anim Pract* **33**, 695-720, v.
- Guo, Y.L., Fitz, J., Schneeberger, K., Ossowski, S., Cao, J., and Weigel, D.** (2011). Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in Arabidopsis. *Plant Physiol* **157**, 757-769.
- Gómez-Gómez, L., and Boller, T.** (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. *Mol Cell* **5**, 1003-1011.
- Haas, B.J., Kamoun, S., Zody, M.C., Jiang, R.H., Handsaker, R.E., Cano, L.M., Grabherr, M., Kodira, C.D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T.O., Ah-Fong, A.M., Alvarado, L., Anderson, V.L., Armstrong, M.R., Avrova, A., Baxter, L., Beynon, J., Boevink, P.C., Bollmann, S.R., Bos, J.I., Bulone, V., Cai, G., Cakir, C., Carrington, J.C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M.A., Fugelstad, J., Gilroy, E.M., Gnerre, S., Green, P.J., Grenville-Briggs, L.J., Griffith, J., Grünwald, N.J., Horn, K., Horner, N.R., Hu, C.H., Huitema, E., Jeong, D.H., Jones, A.M., Jones, J.D., Jones, R.W., Karlsson, E.K., Kunjeti, S.G., Lamour, K., Liu, Z., Ma, L., Maclean, D., Chibucos, M.C., McDonald, H., McWalters, J., Meijer, H.J., Morgan, W., Morris, P.F., Munro, C.A., O'Neill, K., Ospina-Giraldo, M., Pinzón, A., Pritchard, L., Ramsahoye, B., Ren, Q., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D.C., Schumann, U.D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D.J., Sykes, S., Thines, M., van de Vondervoort, P.J., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S., Fry, W., Meyers, B.C., van West, P., Ristaino, J., Govers, F., Birch, P.R., Whisson, S.C., Judelson, H.S., and Nusbaum, C.** (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**, 393-398.
- Hall, S.A., Allen, R.L., Baumber, R.E., Baxter, L.A., Fisher, K., Bittner-Eddy, P.D., Rose, L.E., Holub, E.B., and Beynon, J.L.** (2009). Maintenance of genetic variation in plants and pathogens involves complex networks of gene-for-gene interactions. *Mol Plant Pathol* **10**, 449-457.
- Halter, T., Imkamp, J., Blaum, B.S., Stehle, T., and Kemmerling, B.** (2014). BIR2 affects complex formation of BAK1 with ligand binding receptors in plant defense. *Plant Signal Behav* **9**.
- Hayafune, M., Berisio, R., Marchetti, R., Silipo, A., Kayama, M., Desaki, Y., Arima, S., Squeglia, F., Ruggiero, A., Tokuyasu, K., Molinaro, A., Kaku, H., and Shibuya, N.** (2014). Chitin-induced activation of immune signaling by the rice receptor CEBiP relies on a unique sandwich-type dimerization. *Proc Natl Acad Sci U S A* **111**, E404-413.
- Heller, A., and Thines, M.** (2009). Evidence for the importance of enzymatic digestion of epidermal walls during subepidermal sporulation and pustule opening in white blister rusts (Albuginaceae). *Mycol Res* **113**, 657-667.
- Hogenhout, S.A., Van der Hoorn, R.A., Terauchi, R., and Kamoun, S.** (2009). Emerging

- concepts in effector biology of plant-associated organisms. *Mol Plant Microbe Interact* **22**, 115-122.
- Holub, E.B.** (2001). The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat Rev Genet* **2**, 516-527.
- Holub, E.B.** (2008). Natural history of *Arabidopsis thaliana* and oomycete symbioses. *European Journal of Plant Pathology* **122**, 91-109.
- Holub, E.B., Beynon, L.J., and Crute, I.R.** (1994). Phenotypic and genotypic characterization of interactions between isolates of *peronospora-parasitica* and accessions of *arabidopsis-thaliana*. *Molecular Plant-Microbe Interactions* **7**, 223-239.
- Holub, E.B., Brose, E., Tor, M., Clay, C., Crute, I.R., and Beynon, J.L.** (1995). Phenotypic and genotypic variation in the interaction between *Arabidopsis thaliana* and *Albugo candida*. *Molecular Plant-Microbe Interactions* **8**, 916-928.
- Horton, M.W., Hancock, A.M., Huang, Y.S., Toomajian, C., Atwell, S., Auton, A., Mulyati, N.W., Platt, A., Sperone, F.G., Vilhjálmsson, B.J., Nordborg, M., Borevitz, J.O., and Bergelson, J.** (2012). Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat Genet* **44**, 212-216.
- Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K.** (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res* **35**, W585-587.
- Ishaque, N., Furzer, O.J., E Kemen, G Fabro, D Maclean, D Studholme, JDG Jones.**  
Comparative analysis of 8 *Arabidopsis* downy-mildew genomes reveals strong selective pressures on an elevated proportion of the secretome and novel effector candidates (*in preparation*).
- Jiang, S., Yao, J., Ma, K.W., Zhou, H., Song, J., He, S.Y., and Ma, W.** (2013). Bacterial effector activates jasmonate signaling by directly targeting JAZ transcriptional repressors. *PLoS Pathog* **9**, e1003715.
- Jones, D.A., Thomas, C.M., Hammond-Kosack, K.E., Balint-Kurti, P.J. and Jones, J.D.** (1994) Isolation of the tomato Cf-9 gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science*, **266**, 789-793.
- Jones, J., and Dangl, J.** (2006). The plant immune system. *Nature* **444**, 323-329.
- Jupe, F., Witek, K., Verweij, W., Sliwka, J., Pritchard, L., Etherington, G.J., Maclean, D., Cock, P.J., Leggett, R.M., Bryan, G.J., Cardle, L., Hein, I., and Jones, J.D.** (2013). 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. *Plant J* **76**, 530-544.
- Kadota, Y., Sklenar, J., Derbyshire, P., Stransfeld, L., Asai, S., Ntoukakis, V., Jones, J.D., Shirasu, K., Menke, F., Jones, A., and Zipfel, C.** (2014). Direct regulation of the NADPH oxidase RBOHD by the PRR-associated kinase BIK1 during plant immunity. *Mol Cell* **54**, 43-55.
- Kamoun, S., Furzer, O., Jones, J.D., Judelson, H.S., Ali, G.S., Dalio, R.J., Roy, S.G., Schena, L., Zambounis, A., Panabières, F., Cahill, D., Ruocco, M., Figueiredo, A., Chen, X.R., Hulvey, J., Stam, R., Lamour, K., Gijzen, M., Tyler, B.M., Grünwald, N.J., Mukhtar, M.S., Tomé, D.F., Tör, M., Van den Ackerveken, G., McDowell, J., Daayf, F., Fry, W.E., Lindqvist-Kreuzer, H., Meijer, H.J., Petre, B., Ristaino, J., Yoshida, K., Birch, P.R., and Govers, F.** (2014). The Top 10 oomycete pathogens in molecular plant pathology. *Mol Plant Pathol*.
- Kawchuk, L.M., Hachey, J., Lynch, D.R., Kulcsar, F., van Rooijen, G., Waterer, D.R., Robertson, A., Kokko, E., Byers, R., Howard, R.J., Fischer, R. and Pruffer, D.** (2001) Tomato Ve disease resistance genes encode cell surface-like receptors. *Proc Natl Acad Sci U S A*, **98**, 6511-6515.
- Kaur, P., Sivasithamparam, K., and Barbetti, M.J.** (2008). Pathogenic behaviour of strains of *Albugo candida* from *Brassica juncea* (Indian mustard) and *Raphanus*

- raphanistrum* (wild radish) in Western Australia. Australasian Plant Pathology **37**, 353-356.
- Ke, X.L., Wang, J.G., Gu, Z.M., Li, M., and Gong, X.N.** (2009). Morphological and molecular phylogenetic analysis of two *Saprolegnia* sp. (Oomycetes) isolated from silver crucian carp and zebra fish. Mycol Res **113**, 637-644.
- Kemen, E., Kemen, A.C., Rafiqi, M., Hempel, U., Mendgen, K., Hahn, M., and Voegelé, R.T.** (2005). Identification of a protein from rust fungi transferred from haustoria into infected plant cells. Mol Plant Microbe Interact **18**, 1130-1139.
- Kemen, E., Gardiner, A., Schultz-Larsen, T., Kemen, A.C., Balmuth, A.L., Robert-Seilaniantz, A., Bailey, K., Holub, E., Studholme, D.J., Maclean, D., and Jones, J.D.** (2011). Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. PLoS Biol **9**, e1001094.
- Kim, H.S., Desveaux, D., Singer, A.U., Patel, P., Sondek, J., and Dangl, J.L.** (2005). The *Pseudomonas syringae* effector AvrRpt2 cleaves its C-terminally acylated target, RIN4, from *Arabidopsis* membranes to block RPM1 activation. Proc Natl Acad Sci U S A **102**, 6496-6501.
- Kimura, M.** (1968). Evolutionary rate at the molecular level. Nature **217**, 624-626.
- King, S.R., McLellan, H., Boevink, P.C., Armstrong, M.R., Bukharova, T., Sukarta, O., Win, J., Kamoun, S., Birch, P.R., and Banfield, M.J.** (2014). Phytophthora infestans RXLR effector PexRD2 interacts with host MAPKKK  $\epsilon$  to suppress plant immune signaling. Plant Cell **26**, 1345-1359.
- Knepper, C., Savory, E.A., and Day, B.** (2011). Arabidopsis NDR1 is an integrin-like protein with a role in fluid loss and plasma membrane-cell wall adhesion. Plant Physiol **156**, 286-300.
- Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., and Yanagawa, H.** (2009). Six classes of nuclear localization signals specific to different binding grooves of importin alpha. J Biol Chem **284**, 478-485.
- Krasileva, K.V., Dahlbeck, D., and Staskawicz, B.J.** (2010). Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. Plant Cell **22**, 2444-2458.
- Krasileva, K.V., Zheng, C., Leonelli, L., Goritschnig, S., Dahlbeck, D., and Staskawicz, B.J.** (2011). Global analysis of Arabidopsis/downy mildew interactions reveals prevalence of incomplete resistance and rapid evolution of pathogen recognition. PLoS One **6**, e28765.
- Lacombe, S., Rougon-Cardoso, A., Sherwood, E., Peeters, N., Dahlbeck, D., van Esse, H.P., Smoker, M., Rallapalli, G., Thomma, B.P., Staskawicz, B., Jones, J.D., and Zipfel, C.** (2010). Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. Nat Biotechnol **28**, 365-369.
- Le Floch, G., Rey, P., Benizri, E., Benhamou, N., and Tirilly, Y.** (2003). Impact of auxin-compounds produced by the antagonistic fungus *Pythium oligandrum* or the minor pathogen *Pythium* group F on plant growth. Plant and Soil **257**, 459-470.
- Leonard, K.J.** (1994). Stability of equilibria in a gene-for-gene coevolution model of host-parasite interactions. Phytopathology **84**, 70-77.
- Leonelli, L.** (2011) Genetic, Biochemical, and Structural Studies of the *Hyaloperonospora arabidopsidis* Effector ATR13 and its Cognate *Arabidopsis thaliana* R-gene, RPP13. UC Berkeley Electronic Theses and Dissertations. <http://escholarship.org/uc/item/649803bw>.
- Li, H., and Durbin, R.** (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26**, 589-595.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, G.P.D.P.** (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**, 2078-2079.

- Librado, P., and Rozas, J.** (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.
- Liebrand, T.W., van den Berg, G.C., Zhang, Z., Smit, P., Cordewener, J.H., America, A.H., Sklenar, J., Jones, A.M., Tameling, W.I., Robatzek, S., Thomma, B.P., and Joosten, M.H.** (2013). Receptor-like kinase SOBIR1/EVR interacts with receptor-like proteins in plant immunity against fungal infection. *Proc Natl Acad Sci U S A* **110**, 10010-10015.
- Links, M.G., Holub, E., Jiang, R.H., Sharpe, A.G., Hegedus, D., Beynon, E., Sillito, D., Clarke, W.E., Uzuhashi, S., and Borhan, M.H.** (2011). De novo sequence assembly of *Albugo candida* reveals a small genome relative to other biotrophic oomycetes. *BMC Genomics* **12**, 503.
- Liu, J., Elmore, J.M., Lin, Z.J., and Coaker, G.** (2011). A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. *Cell Host Microbe* **9**, 137-146.
- Liu, J., Elmore, J.M., Fuglsang, A.T., Palmgren, M.G., Staskawicz, B.J., and Coaker, G.** (2009). RIN4 functions with plasma membrane H<sup>+</sup>-ATPases to regulate stomatal apertures during pathogen attack. *PLoS Biol* **7**, e1000139.
- Liu, T., Liu, Z., Song, C., Hu, Y., Han, Z., She, J., Fan, F., Wang, J., Jin, C., Chang, J., Zhou, J.M., and Chai, J.** (2012). Chitin-induced dimerization activates a plant immune receptor. *Science* **336**, 1160-1164.
- Lu, D., Wu, S., Gao, X., Zhang, Y., Shan, L., and He, P.** (2010). A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin receptor complex to initiate plant innate immunity. *Proc Natl Acad Sci U S A* **107**, 496-501.
- Macho, A.P., and Zipfel, C.** (2014). Plant PRRs and the activation of innate immune signaling. *Mol Cell* **54**, 263-272.
- Macho, A.P., Schwessinger, B., Ntoukakis, V., Brutus, A., Segonzac, C., Roy, S., Kadota, Y., Oh, M.H., Sklenar, J., Derbyshire, P., Lozano-Durán, R., Malinovsky, F.G., Monaghan, J., Menke, F.L., Huber, S.C., He, S.Y., and Zipfel, C.** (2014). A bacterial tyrosine phosphatase inhibits plant pattern recognition receptor activation. *Science* **343**, 1509-1512.
- Mackey, D., Holt, B.F., Wiig, A., and Dangl, J.L.** (2002). RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell* **108**, 743-754.
- Maekawa, T., Cheng, W., Spiridon, L.N., Töller, A., Lukasik, E., Saijo, Y., Liu, P., Shen, Q.H., Micluta, M.A., Somssich, I.E., Takken, F.L., Petrescu, A.J., Chai, J., and Schulze-Lefert, P.** (2011). Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. *Cell Host Microbe* **9**, 187-199.
- Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., and Turner, D.J.** (2010). Target-enrichment strategies for next-generation sequencing. *Nat Methods* **7**, 111-118.
- Mann, H.B., Whitney, D.R.,** (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics* **18** **1**, 50-60.
- Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P.** (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-913.
- Martin, D., and Rybicki, E.** (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562-563.
- Mascheretti, S., Croucher, P.J., Kozanitas, M., Baker, L., and Garbelotto, M.** (2009). Genetic epidemiology of the sudden oak death pathogen *Phytophthora ramorum* in California. *Mol Ecol* **18**, 4577-4590.
- Mathieson, I., and McVean, G.** (2012). Differential confounding of rare and common

- variants in spatially structured populations. *Nat Genet* **44**, 243-246.
- McCann, H.C., Nahal, H., Thakur, S., and Guttman, D.S.** (2012). Identification of innate immunity elicitors using molecular signatures of natural selection. *Proc Natl Acad Sci U S A* **109**, 4215-4220.
- McDonald, J.H., and Kreitman, M.** (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652-654.
- McDonald, M.C., Oliver, R.P., Friesen, T.L., Brunner, P.C., and McDonald, B.A.** (2013). Global diversity and distribution of three necrotrophic effectors in *Phaeosphaeria nodorum* and related species. *New Phytol* **199**, 241-251.
- McDowell, J.M., Cuzick, A., Can, C., Beynon, J., Dangl, J.L., and Holub, E.B.** (2000). Downy mildew (*Peronospora parasitica*) resistance genes in *Arabidopsis* vary in functional requirements for NDR1, EDS1, NPR1 and salicylic acid accumulation. *Plant J* **22**, 523-529.
- McDowell, J.M., Dhandaydham, M., Long, T.A., Aarts, M.G., Goff, S., Holub, E.B., and Dangl, J.L.** (1998). Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of *Arabidopsis*. *Plant Cell* **10**, 1861-1874.
- McGaugh, S.E., Heil, C.S., Manzano-Winkler, B., Loewe, L., Goldstein, S., Himmel, T.L., and Noor, M.A.** (2012). Recombination modulates how selection affects linked sites in *Drosophila*. *PLoS Biol* **10**, e1001422.
- Mentlak, T.A., Kombrink, A., Shinya, T., Ryder, L.S., Otomo, I., Saitoh, H., Terauchi, R., Nishizawa, Y., Shibuya, N., Thomma, B.P., and Talbot, N.J.** (2012). Effector-mediated suppression of chitin-triggered immunity by *magnaporthe oryzae* is necessary for rice blast disease. *Plant Cell* **24**, 322-335.
- Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W.** (2003). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**, 809-834.
- Mindrinis, M., Katagiri, F., Yu, G.L., and Ausubel, F.M.** (1994). The *A. thaliana* disease resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* **78**, 1089-1099.
- Moncalvo, J.M., Wang, H.H., and Hseu, R.S.** (1995). Phylogenetic-relationships in *ganoderma* inferred from the internal transcribed spacers and 25s ribosomal dna-sequences. *Mycologia* **87**, 223-238.
- Mooney, M.A., Nigg, J.T., McWeeney, S.K., and Wilmot, B.** (2014). Functional and genomic context in pathway analysis of GWAS data. *Trends Genet* **30**, 390-400.
- Nandety, R.S., Caplan, J.L., Cavanaugh, K., Perroud, B., Wroblewski, T., Michelmore, R.W., and Meyers, B.C.** (2013). The role of TIR-NBS and TIR-X proteins in plant basal defense responses. *Plant Physiol* **162**, 1459-1472.
- Narusaka, M., Shirasu, K., Noutoshi, Y., Kubo, Y., Shiraishi, T., Iwabuchi, M., and Narusaka, Y.** (2009). RRS1 and RPS4 provide a dual Resistance-gene system against fungal and bacterial pathogens. *Plant J* **60**, 218-226.
- Nemri, A., Atwell, S., Tarone, A.M., Huang, Y.S., Zhao, K., Studholme, D.J., Nordborg, M., and Jones, J.D.** (2010). Genome-wide survey of *Arabidopsis* natural variation in downy mildew resistance using combined association and linkage mapping. *Proc Natl Acad Sci U S A* **107**, 10302-10307.
- Nürnberger, T., Brunner, F., Kemmerling, B., and Piater, L.** (2004a). Innate immunity in plants and animals: striking similarities and obvious differences. *Immunol Rev* **198**, 249-266.
- O'Neill, L.A., Golenbock, D., and Bowie, A.G.** (2013). The history of Toll-like receptors - redefining innate immunity. *Nat Rev Immunol* **13**, 453-460.
- Ogasawara, Y., Kaya, H., Hiraoka, G., Yumoto, F., Kimura, S., Kadota, Y., Hishinuma, H., Senzaki, E., Yamagoe, S., Nagata, K., Nara, M., Suzuki, K., Tanokura, M., and Kuchitsu, K.** (2008). Synergistic activation of the *Arabidopsis* NADPH oxidase

- AtrbohD by Ca<sup>2+</sup> and phosphorylation. *J Biol Chem* **283**, 8885-8892.
- Oh, S.K., Young, C., Lee, M., Oliva, R., Bozkurt, T.O., Cano, L.M., Win, J., Bos, J.I., Liu, H.Y., van Damme, M., Morgan, W., Choi, D., Van der Vossen, E.A., Vleeshouwers, V.G., and Kamoun, S.** (2009). In planta expression screens of *Phytophthora infestans* RXLR effectors reveal diverse phenotypes, including activation of the *Solanum bulbocastanum* disease resistance protein Rpi-blb2. *Plant Cell* **21**, 2928-2947.
- Ong, F.S., Lin, J.C., Das, K., Grosu, D.S., and Fan, J.B.** (2013). Translational utility of next-generation sequencing. *Genomics* **102**, 137-139.
- Parker, J.E., Holub, E.B., Frost, L.N., Falk, A., Gunn, N.D., and Daniels, M.J.** (1996). Characterization of eds1, a mutation in *Arabidopsis* suppressing resistance to *Peronospora parasitica* specified by several different RPP genes. *Plant Cell* **8**, 2033-2046.
- Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A.J., Thomson, N.R., Quail, M., Smith, F., Walker, D., Libberton, B., Fenton, A., Hall, N., and Brockhurst, M.A.** (2010). Antagonistic coevolution accelerates molecular evolution. *Nature* **464**, 275-278.
- Petre, B., and Kamoun, S.** (2014). How do filamentous pathogens deliver effector proteins into plant cells? *PLoS Biol* **12**, e1001801.
- Petrie, G.A.** (1975). Prevalence Of Oospores Of *Albugo cruciferarum* In Brassica Seed Samples From Western Canada 1967-1973. *Canadian Plant Disease Survey* **55**, 19-24.
- Ploch, S., Choi, Y.J., Rost, C., Shin, H.D., Schilling, E., and Thines, M.** (2010). Evolution of diversity in *Albugo* is driven by high host specificity and multiple speciation events on closely related Brassicaceae. *Molecular Phylogenetics and Evolution* **57**, 812-820.
- Qi, D., Dubiella, U., Kim, S.H., Sloss, D.I., Downen, R.H., Dixon, J.E., and Innes, R.W.** (2014). Recognition of the protein kinase AVRPPHB SUSCEPTIBLE1 by the disease resistance protein RESISTANCE TO PSEUDOMONAS SYRINGAE5 is dependent on s-acylation and an exposed loop in AVRPPHB SUSCEPTIBLE1. *Plant Physiol* **164**, 340-351.
- Quinlan, A.R., and Hall, I.M.** (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842.
- Qutob, D., Chapman, B.P., and Gijzen, M.** (2013). Transgenerational gene silencing causes gain of virulence in a plant pathogen. *Nat Commun* **4**, 1349.
- Raffaele, S., Win, J., Cano, L.M., and Kamoun, S.** (2010a). Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. *BMC Genomics* **11**, 637.
- Raffaele, S., Farrer, R.A., Cano, L.M., Studholme, D.J., MacLean, D., Thines, M., Jiang, R.H., Zody, M.C., Kunjeti, S.G., Donofrio, N.M., Meyers, B.C., Nusbaum, C., and Kamoun, S.** (2010b). Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* **330**, 1540-1543.
- Rech, G.E., Sanz-Martín, J.M., Anisimova, M., Sukno, S.A., and Thon, M.R.** (2014). Natural selection on coding and non-coding DNA sequences is associated with virulence genes in a plant pathogenic fungus. *Genome Biol Evol*.
- Rehmany, A.P., Gordon, A., Rose, L.E., Allen, R.L., Armstrong, M.R., Whisson, S.C., Kamoun, S., Tyler, B.M., Birch, P.R., and Beynon, J.L.** (2005). Differential recognition of highly divergent downy mildew avirulence gene alleles by RPP1 resistance genes from two *Arabidopsis* lines. *Plant Cell* **17**, 1839-1850.
- Rentel, M.C., Leonelli, L., Dahlbeck, D., Zhao, B., and Staskawicz, B.J.** (2008). Recognition of the *Hyaloperonospora parasitica* effector ATR13 triggers resistance against oomycete, bacterial, and viral pathogens. *Proc Natl Acad Sci U S A* **105**, 1091-1096.
- Rimmer, S.R., Mathur, S., and Wu, C.R.** (2000). Virulence of isolates of *Albugo candida* from western Canada to Brassica species. *Canadian Journal of Plant Pathology* **22**,

- Robert-Seilaniantz, A., Grant, M., and Jones, J.D.** (2011). Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism. *Annu Rev Phytopathol* **49**, 317-343.
- Rodewald, J., and Trognitz, B.** (2013). Solanum resistance genes against *Phytophthora infestans* and their corresponding avirulence genes. *Mol Plant Pathol* **14**, 740-757.
- Roux, M., Schwessinger, B., Albrecht, C., Chinchilla, D., Jones, A., Holton, N., Malinovsky, F.G., Tör, M., de Vries, S., and Zipfel, C.** (2011). The Arabidopsis leucine-rich repeat receptor-like kinases BAK1/SERK3 and BKK1/SERK4 are required for innate immunity to hemibiotrophic and biotrophic pathogens. *Plant Cell* **23**, 2440-2455.
- Römer, P., Hahn, S., Jordan, T., Strauss, T., Bonas, U., and Lahaye, T.** (2007). Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science* **318**, 645-648.
- Saunders, D.G., Win, J., Kamoun, S., and Raffaele, S.** (2014). Two-dimensional data binning for the analysis of genome architecture in filamentous plant pathogens and other eukaryotes. *Methods Mol Biol* **1127**, 29-51.
- Schorneck, S., van Damme, M., Bozkurt, T.O., Cano, L.M., Smoker, M., Thines, M., Gaulin, E., Kamoun, S., and Huitema, E.** (2010). Ancient class of translocated oomycete effectors targets the host nucleus. *Proc Natl Acad Sci U S A* **107**, 17421-17426.
- Schorneck, S., Huitema, E., Cano, L., Bozkurt, T., Oliva, R., Van Damme, M., Schwizer, S., Raffaele, S., Chaparro-Garcia, A., Farrer, R., Segretin, M., Bos, J., Haas, B., Zody, M., Nusbaum, C., Win, J., Thines, M., and Kamoun, S.** (2009). Ten things to know about oomycete effectors. *Mol Plant Pathol* **10**, 795-803.
- Schwessinger, B., Roux, M., Kadota, Y., Ntoukakis, V., Sklenar, J., Jones, A., and Zipfel, C.** (2011). Phosphorylation-dependent differential regulation of plant growth, cell death, and innate immunity by the regulatory receptor-like kinase BAK1. *PLoS Genet* **7**, e1002046.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., and Higgins, D.G.** (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.
- Sigrist, C.J., Cerutti, L., de Castro, E., Langendijk-Genevaux, P.S., Bulliard, V., Bairoch, A., and Hulo, N.** (2010). PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* **38**, D161-166.
- Sinapidou, E., Williams, K., Nott, L., Bahkt, S., Tör, M., Crute, I., Bittner-Eddy, P., and Beynon, J.** (2004). Two TIR:NB:LRR genes are required to specify resistance to *Peronospora parasitica* isolate Cala2 in Arabidopsis. *Plant J* **38**, 898-909.
- Sohn, K., Lei, R., Nemri, A., and Jones, J.** (2007). The downy mildew effector proteins ATR1 and ATR13 promote disease susceptibility in *Arabidopsis thaliana*. *Plant Cell* **19**, 4077-4090.
- Sohn K.H., Zhang Y., Jones J.D.** (2009) The Pseudomonas syringae effector protein, AvrRPS4, requires in planta processing and the KRVY domain to function. *Plant J* **57**, 1079–1091.
- Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X., Zhu, L.H., Fauquet, C., and Ronald, P.** (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* **270**, 1804-1806.
- Soylu, S., Keshavarzi, M., Brown, I., and Mansfield, J.W.** (2003). Ultrastructural characterisation of interactions between *Arabidopsis thaliana* and *Albugo candida*. *Physiological and Molecular Plant Pathology* **63**, 201-211.
- Stahl, E.A., Dwyer, G., Mauricio, R., Kreitman, M., and Bergelson, J.** (1999). Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. *Nature* **400**, 667-671.

- Stephens, M., and Scheet, P.** (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am J Hum Genet* **76**, 449-462.
- Stern, A., Doron-Faigenboim, A., Erez, E., Martz, E., Bacharach, E., and Pupko, T.** (2007). Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Res* **35**, W506-511.
- Stranger, B.E., Stahl, E.A., and Raj, T.** (2011). Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* **187**, 367-383.
- Stukenbrock, E.H., and McDonald, B.A.** (2009). Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Mol Plant Microbe Interact* **22**, 371-380.
- Su, Z., Cardin, N., Donnelly, P., Marchini, J., and Wellcome Trust Case, C.** (2009). A Bayesian Method for Detecting and Characterizing Allelic Heterogeneity and Boosting Signals in Genome-Wide Association Studies. *Statistical Science* **24**, 430-450.
- Sun, Y., Li, L., Macho, A.P., Han, Z., Hu, Z., Zipfel, C., Zhou, J.M., and Chai, J.** (2013). Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. *Science* **342**, 624-628.
- Sánchez-Vallet, A., Saleem-Batcha, R., Kombrink, A., Hansen, G., Valkenburg, D.J., Thomma, B.P., and Mesters, J.R.** (2013). Fungal effector Ecp6 outcompetes host immune receptor for chitin binding through intrachain LysM dimerization. *Elife* **2**, e00790.
- Tajima, F.** (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.
- Takken, F.L., and Goverse, A.** (2012). How to build a pathogen detector: structural basis of NB-LRR function. *Curr Opin Plant Biol* **15**, 375-384.
- Tellier, A., and Brown, J.K.** (2011). Spatial heterogeneity, frequency-dependent selection and polymorphism in host-parasite interactions. *BMC Evol Biol* **11**, 319.
- Terauchi, R., and Yoshida, K.** (2010). Towards population genomics of effector-effector target interactions. *New Phytol* **187**, 929-939.
- Thines, M.** (2014). Phylogeny and evolution of plant pathogenic oomycetes—a global overview (*Eur J Plant Pathol*).
- Thines, M., and Spring, O.** (2005). A revision of *Albugo* (chromista, peronosporomycetes). *Mycotaxon* **92**, 443-458.
- Thines, M., and Kamoun, S.** (2010). Oomycete-plant coevolution: recent advances and future prospects. *Curr Opin Plant Biol* **13**, 427-433.
- Thines, M., Choi, Y.J., Kemen, E., Ploch, S., Holub, E.B., Shin, H.D., and Jones, J.D.** (2009). A new species of *Albugo* parasitic to *Arabidopsis thaliana* reveals new evolutionary patterns in white blister rusts (*Albuginaceae*). *Persoonia* **22**, 123-128.
- Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P.** (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192.
- Thudi, M., Li, Y., Jackson, S.A., May, G.D., and Varshney, R.K.** (2012). Current state-of-art of sequencing technologies for plant genomics research. *Brief Funct Genomics* **11**, 3-11.
- Tian, M., Win, J., Savory, E., Burkhardt, A., Held, M., Brandizzi, F., and Day, B.** (2011). 454 Genome sequencing of *Pseudoperonospora cubensis* reveals effector proteins with a QXLR translocation motif. *Mol Plant Microbe Interact* **24**, 543-553.
- Torto, T.A., Li, S., Styer, A., Huitema, E., Testa, A., Gow, N.A., van West, P., and Kamoun, S.** (2003). EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. *Genome Res* **13**, 1675-1685.
- Tsuchiya, T., and Eulgem, T.** (2013). An alternative polyadenylation mechanism coopted to

the Arabidopsis RPP7 gene through intronic retrotransposon domestication. Proc Natl Acad Sci U S A **110**, E3535-3543.

- Tyler, B.M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R.H., Aerts, A., Arredondo, F.D., Baxter, L., Bensasson, D., Beynon, J.L., Chapman, J., Damasceno, C.M., Dorrance, A.E., Dou, D., Dickerman, A.W., Dubchak, I.L., Garbelotto, M., Gijzen, M., Gordon, S.G., Govers, F., Grunwald, N.J., Huang, W., Ivors, K.L., Jones, R.W., Kamoun, S., Krampis, K., Lamour, K.H., Lee, M.K., McDonald, W.H., Medina, M., Meijer, H.J., Nordberg, E.K., Maclean, D.J., Ospina-Giraldo, M.D., Morris, P.F., Phuntumart, V., Putnam, N.H., Rash, S., Rose, J.K., Sakihama, Y., Salamov, A.A., Savidor, A., Scheuring, C.F., Smith, B.M., Sobral, B.W., Terry, A., Torto-Alalibo, T.A., Win, J., Xu, Z., Zhang, H., Grigoriev, I.V., Rokhsar, D.S., and Boore, J.L.** (2006). *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* **313**, 1261-1266.
- van der Biezen, E.A., and Jones, J.D.** (1998). Plant disease-resistance proteins and the gene-for-gene concept. *Trends Biochem Sci* **23**, 454-456.
- van der Hoorn, R.A., and Kamoun, S.** (2008). From Guard to Decoy: a new model for perception of plant pathogen effectors. *Plant Cell* **20**, 2009-2017.
- van der Hoorn, R.A., Laurent, F., Roth, R., and De Wit, P.J.** (2000). Agroinfiltration is a versatile tool that facilitates comparative analyses of Avr9/Cf-9-induced and Avr4/Cf-4-induced necrosis. *Mol Plant Microbe Interact* **13**, 439-446.
- van der Sijde, M.R., Ng, A., and Fu, J.** (2014). Systems genetics: From GWAS to disease pathways. *Biochim Biophys Acta* **1842**, 1903-1909.
- van West, P., de Bruijn, I., Minor, K.L., Phillips, A.J., Robertson, E.J., Wawra, S., Bain, J., Anderson, V.L., and Secombes, C.J.** (2010). The putative RxLR effector protein SpHtp1 from the fish pathogenic oomycete *Saprolegnia parasitica* is translocated into fish cells. *FEMS Microbiol Lett* **310**, 127-137.
- Venugopal, S.C., Jeong, R.D., Mandal, M.K., Zhu, S., Chandra-Shekara, A.C., Xia, Y., Hersh, M., Stromberg, A.J., Navarre, D., Kachroo, A., and Kachroo, P.** (2009). Enhanced disease susceptibility 1 and salicylic acid act redundantly to regulate resistance gene-mediated signaling. *PLoS Genet* **5**, e1000545.
- Vleeshouwers, V.G., and Oliver, R.P.** (2014). Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. *Mol Plant Microbe Interact* **27**, 196-206.
- Wang, G., Fiers, M., Ellendorff, U., Wang, Z., de Wit, P.J.G.M., Angenent, G.C., and Thomma, B.P.H.J.** (2010). The Diverse Roles of Extracellular Leucine-rich Repeat-containing Receptor-like Proteins in Plants. *Critical Reviews in Plant Sciences* **29**, 285-299.
- Weigel, D., and Mott, R.** (2009). The 1001 genomes project for Arabidopsis thaliana. *Genome Biol* **10**, 107.
- Weßling, R., Epple, P., Altmann, S., He, Y., Yang, L., Henz, S.R., McDonald, N., Wiley, K., Bader, K.C., Gläßer, C., Mukhtar, M.S., Haigis, S., Ghamsari, L., Stephens, A.E., Ecker, J.R., Vidal, M., Jones, J.D., Mayer, K.F., Ver Loren van Themaat, E., Weigel, D., Schulze-Lefert, P., Dangl, J.L., Panstruga, R., and Braun, P.** (2014). Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe* **16**, 364-375.
- Whisson, S.C., Boevink, P.C., Moleleki, L., Avrova, A.O., Morales, J.G., Gilroy, E.M., Armstrong, M.R., Grouffaud, S., van West, P., Chapman, S., Hein, I., Toth, I.K., Pritchard, L., and Birch, P.R.** (2007). A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* **450**, 115-118.
- Williams, S.J., Sohn, K.H., Wan, L., Bernoux, M., Sarris, P.F., Segonzac, C., Ve, T., Ma, Y., Saucet, S.B., Ericsson, D.J., Casey, L.W., Lonhienne, T., Winzor, D.J., Zhang, X., Coerdts, A., Parker, J.E., Dodds, P.N., Kobe, B., and Jones, J.D.** (2014). Structural

- basis for assembly and function of a heterodimeric plant immune receptor. *Science* **344**, 299-303.
- Wilton, M., Subramaniam, R., Elmore, J., Felsensteiner, C., Coaker, G., and Desveaux, D.** (2010). The type III effector HopF2Pto targets Arabidopsis RIN4 protein to promote *Pseudomonas syringae* virulence. *Proc Natl Acad Sci U S A* **107**, 2349-2354.
- Win, J., Morgan, W., Bos, J., Krasileva, K.V., Cano, L.M., Chaparro-Garcia, A., Ammar, R., Staskawicz, B.J., and Kamoun, S.** (2007). Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell* **19**, 2349-2369.
- Xiang, T., Zong, N., Zhang, J., Chen, J., Chen, M., and Zhou, J.M.** (2011). BAK1 is not a target of the *Pseudomonas syringae* effector AvrPto. *Mol Plant Microbe Interact* **24**, 100-107.
- Yang, Z.** (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591.
- Yang, Z., and Nielsen, R.** (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* **17**, 32-43.
- Yoshida, K., Saitoh, H., Fujisawa, S., Kanzaki, H., Matsumura, H., Tosa, Y., Chuma, I., Takano, Y., Win, J., Kamoun, S., and Terauchi, R.** (2009). Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* **21**, 1573-1591.
- Zerbino, D.R., and Birney, E.** (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821-829.
- Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J.D., Boller, T., and Felix, G.** (2006). Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. *Cell* **125**, 749-760.
- Zong, N., Xiang, T., Zou, Y., Chai, J., and Zhou, J.M.** (2008). Blocking and triggering of plant immunity by *Pseudomonas syringae* effector AvrPto. *Plant Signal Behav* **3**, 583-585.

## 10. APPENDICES

Differential group number	<i>Nc14</i>	<i>Em1</i>	<i>Abo1</i>	<i>Sua1</i>	<i>Went1</i>	<i>Ash4</i>	Accessions
1	S	R	S	R	R	R	CIBC-5, Ge-0
2	R	R	S	S	S	R	HR-10
3	R	S	S	S	S	S	HR-5, Ren-11, As-77, BAT1, Fri2, EKN 3, GrA-5
4	R	R	S	S	R	S	Knox-18, San-2, T1010, UIIA-2
5	R	R	S	R	S	R	NFA-10
6	R	R	S	S	S	S	Sq-1, T450, TDr9, S294BeL4
7	R	S	S	S	R	S	Fly2-1, Fly2-2, Hov1-7, Kni-1, Rev-3, T1160, T800, T860, UIIA-1
8	R	S	S	S	R	R	Udul 1-34
9	R	R	R	R	R	R	Ts-1, Sf-2
10	R	R	S	R	R	R	Ksk-1
11	S	R	R	R	S	S	Ei-2, RRS-7
12	S	R	S	S	S	S	Kin-0, Pna-17
13	S	R	R	R	R	R	UK-1
14	S	R	R	S	S	S	Ts-5

**Appendix table 4A1. Re-organised table structured around the fourteen observed differential *A1* virulence groups.** Accessions phenotyped with the various *A1* isolates were grouped together according to the differential resistance/susceptibility pattern that they displayed.

HR-10			
Gene	Corellated NS polymorphisms	Annotation (automatic)	Notes
AINc14C103G6107	26	unknown	
AINc14C204G8769	26	unknown	
AINc14C65G4610	24	unknown	SSP24
AINc14C303G10412	3	unknown	RxL14
AINc14C43G3600	3	SSP6	

HR-5, Ren-11, As-77, BAT1, Fri2, FkN3, GrA-5			
Gene	Corellated NS polymorphisms	Annotation (automatic)	Notes
AINc14C46G3742	17	inositol-3 putative	
AINc14C260G9794	6	unknown	SSP
AINc14C28G2718	6	unknown SSP16	SSP16
AINc14C260G9793	4	unknown	
AINc14C365G11050	4	unknown	
AINc14C142G7291	3	unknown	SSP27

Knox-18, San-2, T1010, UJIA-2			
Gene	Corellated NS polymorphisms	Annotation (automatic)	Notes
AINc14C28G2718	21	unknown SSP16	SSP16
AINc14C18G1839	12	unknown	possible aspartate protease
AINc14C361G10994	7	unknown	CxHC2
AINc14C236G9385	5	DEAD/DEAH box RNA helicase putative	
AINc14C316G10530	5	unknown	SSP15
AINc14C34G3068	5	DEAD/DEAH box RNA helicase putative	

NFA-10			
Gene	Corellated NS polymorphisms	Annotation (automatic)	Notes
AINc14C236G9399	3	conserved hypothetical protein	
AINc14C34G3052	3	conserved hypothetical protein	

Sq-1, T450, TDr9, S294Bel4			
Gene	Corellated NS polymorphisms	Annotation (automatic)	Notes
AINc14C28G2718	7	unknown SSP16	SSP16
AINc14C325G10635	4	unknown	SSP19
AINc14C143G7305	3	ethanolamine-phosphate cytidyltransferase putative	

Fly2-1, Fly2-2, Hov1-7, Kni-1, Rev-3, T1160, T800, T860, UJIA-1			
Gene	Corellated NS polymorphisms	Annotation (automatic)	Notes
AINc14C260G9794	13	unknown	SSP
AINc14C28G2718	8	unknown SSP16	SSP16

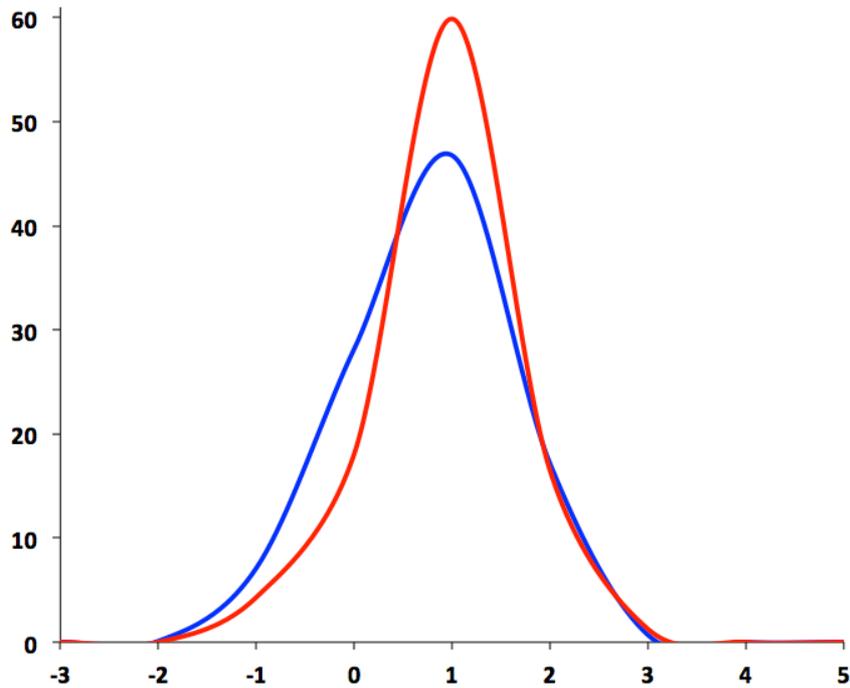
  

Ksk-1			
Gene	Corellated NS polymorphisms	Annotation (automatic)	Notes
AINc14C169G7963	33	unknown	SSP17
AINc14C56G4264	22	unknown	SSP18
AINc14C35G3154	14	unknown	SSP21
AINc14C163G7833	5	unknown	
AINc14C56G4265	5	unknown	
AINc14C64G4566	4	unknown	elicitin-like

Uk-1			
Gene	Corellated NS polymorphisms	Annotation (automatic)	Notes
AINc14C46G3742	17	inositol-3 putative	
AINc14C260G9794	6	unknown	SSP
AINc14C28G2718	6	unknown SSP16	SSP16
AINc14C260G9793	4	unknown	
AINc14C365G11050	4	unknown	

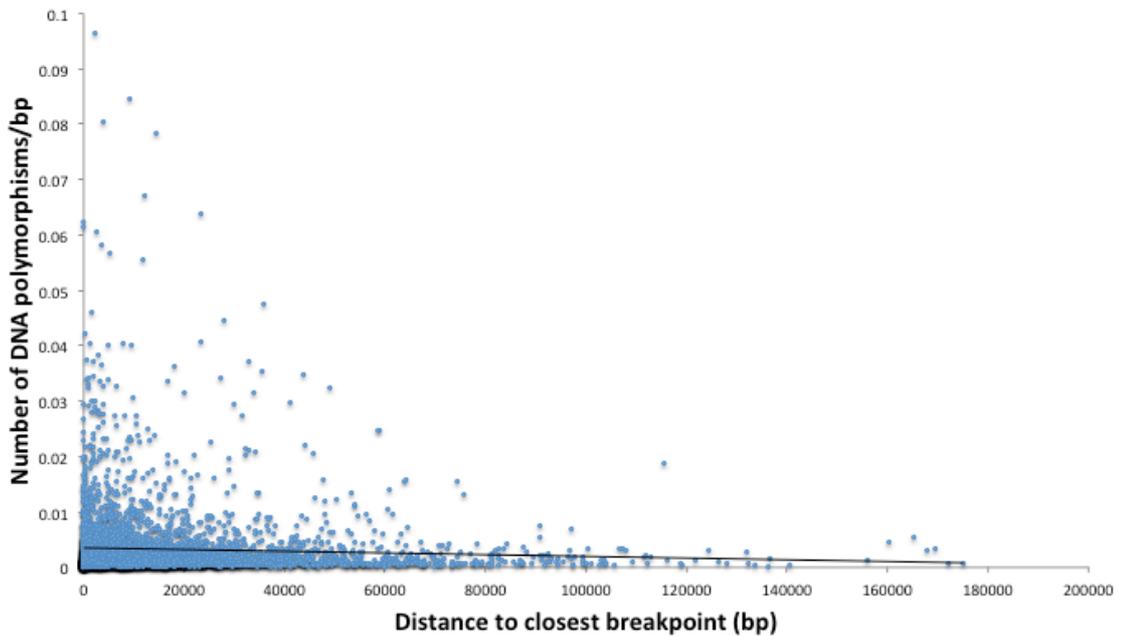
**Appendix table 4A2. Full table of AI association genomic predictions.**



**Appendix figure 4A1. A distribution of Tajima's D for all AI genes, generated from the analysis of the 6 isolates.** Tajima's D was calculated for each gene using the VariTale method as described in Chapter 2. Genes encoding proteins with a predicted signal peptide are in red, and those without in blue.

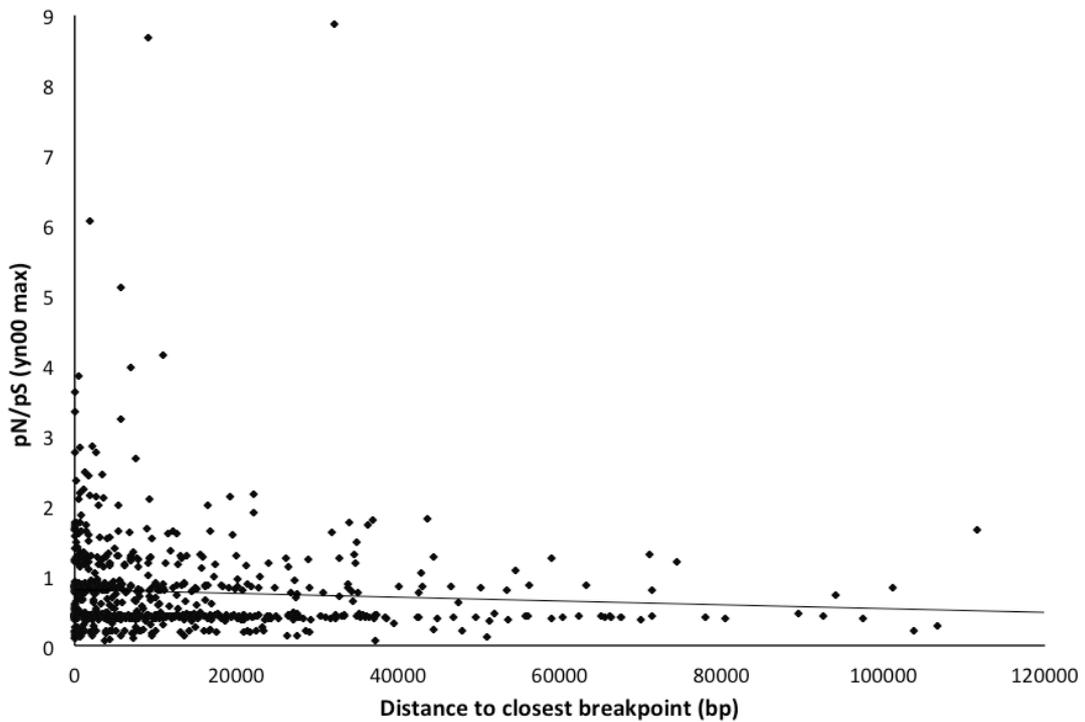
Contig	Gene	Tajima's D	Annotation
18	1839	2.7189	unknown
65	4610	2.6353	SSP24
177	8157	2.5327	kazal protease inhibitor putative
205	8798	2.5327	conserved hypothetical protein
361	10994	2.5327	unknown
176	8132	2.3858	unknown
313	13412	2.3858	CHXC29
205	8792	2.2762	hypothetical protein UM06115.1
453	11745	2.2762	nd
376	11176	2.2509	dolichyl-diphosphooligosaccharide-protein glycosyltransferas
155	7617	2.123	unknown
178	8172	2.1159	carbohydrate-binding protein putative
168	7937	1.9729	conserved hypothetical protein
212	8942	1.97	unknown
211	8913	1.9544	unknown
205	8793	1.9229	hypothetical protein UM06115.1
295	10287	1.9229	conserved hypothetical protein
16	1784	1.8912	unknown
16	1785	1.8912	unknown
294	10281	1.8912	conserved hypothetical protein

**Appendix table 4A3. Top ranked predicted secreted protein encoding genes by Tajima's D.** Tajima's D was calculated for each gene using the VariTale method as described in Chapter 2.



**Appendix figure 4A3. Correlation of polymorphism and distance to closest recombination breakpoint.**

The number of nucleotide polymorphisms in each gene was extracted and plotted with the distance of each gene to its closest predicted recombination breakpoint. ~8000 genes were analyzed, and a non-significant negative correlation was observed (Pearson test: -0.054).



**Appendix figure 4A4. Correlation of pN/pS and distance to closest recombination breakpoint.**

The pN/pS of each gene was determined using VariTale and plotted with the distance of each gene to its closest predicted recombination breakpoint and ~8000 genes analyzed.

Locus Identifier	Gene Model Description
AT1G59077	BEST Arabidopsis thaliana protein match is: B-block binding subunit of TFIIC (TAIR:AT1G59453.1)
AT1G58170	Disease resistance-responsive (dirigent-like protein) family protein
AT1G58235	unknown protein
AT1G58220	Homeodomain-like superfamily protein
AT1G59725	DNAJ heat shock family protein
AT1G58684	Ribosomal protein S5 family protein
AT1G59030	GDSL-like Lipase/Acylhydrolase superfamily protein
AT1G58320	PLAC8 family protein
AT1G58725	GDSL-like Lipase/Acylhydrolase superfamily protein
AT1G58460	unknown protein
AT1G58265	Cytochrome P450 superfamily protein
AT1G59359	Ribosomal protein S5 family protein
AT1G58280	Phosphoglycerate mutase family protein
AT1G59710	FUNCTIONS IN: molecular_function unknown
AT1G59550	UBX domain-containing protein
AT1G58120	BEST Arabidopsis thaliana protein match is: methyltransferases (TAIR:AT5G01710.1)
AT1G59630	F-box associated ubiquitination effector family protein
AT1G58420	Uncharacterised conserved protein UCP031279
AT1G59535	unknown protein
AT1G58225	unknown protein
AT1G58936	Inositol-pentakisphosphate 2-kinase family protein
AT1G58110	Basic-leucine zipper (bZIP) transcription factor family protein
AT1G58248	Encodes a Plant thionin family protein
AT1G59312	Inositol-pentakisphosphate 2-kinase family protein
AT1G58245	Encodes a Plant thionin family protein
AT1G59406	GDSL-like Lipase/Acylhydrolase superfamily protein
AT1G59171	Inositol-pentakisphosphate 2-kinase family protein
AT1G59660	Nucleoporin autopeptidase
AT1G59690	F-box associated ubiquitination effector family protein
AT1G58090	F-box and associated interaction domains-containing protein
AT1G58230	BEACH-DOMAIN HOMOLOG B
AT1G58983	Ribosomal protein S5 family protein
AT1G58766	BEST Arabidopsis thaliana protein match is: B-block binding subunit of TFIIC (TAIR:AT1G59453.1)
AT1G59453	B-block binding subunit of TFIIC
AT1G58242	Plant thionin family protein
AT1G59675	F-box family protein
AT1G58310	F-box/RNI-like superfamily protein
AT1G59722	unknown protein
AT1G58150	unknown protein
AT1G58643	Inositol-pentakisphosphate 2-kinase family protein
AT1G59510	Encodes CF9.
AT1G59650	Encodes CW14.
AT1G59520	Encodes CW7.
AT1G59500	encodes an IAA-amido synthase that conjugates Asp and other amino acids to auxin in vitro.
AT1G58290	Encodes a protein with glutamyl-tRNA reductase (GluTR) activity
AT1G58370	Encodes a protein with xylanase activity.
AT1G58430	Encodes an anther-specific proline-rich protein.
AT1G58520	RXW8
AT1G58440	Encodes a putative protein that has been speculated, based on sequence similarities, to have squalene monooxygenase activity.
AT1G58380	XW6
AT1G59540	Encodes a kinesin-like protein.
AT1G59590	ZCF37 mRNA, complete cds
AT1G59560	ZCF61
AT1G59600	ZCW7
AT1G58340	Encodes a plant MATE (multidrug and toxic compound extrusion) transporter
AT1G58350	ZW18
AT1G58330	ZW2
AT1G58270	ZW9 mRNA, complete cds
AT1G58360	Encodes AAP1 (amino acid permease 1), a neutral amino acid transporter expressed in seeds.
AT1G58080	ATP phosphoribosyl transferase, catalyses first step of histidine biosynthesis
AT1G59530	basic leucine-zipper 4 (bZIP4)
AT1G58180	beta carbonic anhydrase 6 (BCA6)
AT1G59640	A basic helix-loop-helix encoding gene (BIGPETAL, BPE) involved in the control of petal size.
AT1G59720	Pentatricopeptide Repeat Protein containing the DYW motif.
AT1G58122	Upstream open reading frames (uORFs) are small open reading frames found in the 5' UTR of a mature mRNA.
AT1G58260	member of CYP79C subfamily of cytochrome p450s. Encodes a putative xylan endohydrolase.
AT1G59610	A high molecular weight GTPase whose GTP-binding domain shows a low homology to those of other plant dynamin-like proteins.
AT1G58210	Encodes a member of the NET superfamily of proteins that potentially couples different membranes to the actin cytoskeleton in plant cells.
AT1G59680	embryo sac development arrest 1 (EDA1)
AT1G59670	Encodes glutathione transferase belonging to the tau class of GSTs. Naming convention according to Wagner et al. (2002).
AT1G59700	Encodes glutathione transferase belonging to the tau class of GSTs. Naming convention according to Wagner et al. (2002).
AT1G58300	Encodes a member (HO4) of the heme oxygenase family.
AT1G58160	Mannose-binding lectin superfamily protein
AT1G59580	encodes a mitogen-activated kinase involved in innate immunity
AT1G58200	A member of MscS-like gene family, structurally very similar to MSL2.
AT1G58190	receptor like protein 9 (RLP9)
AT1G58470	encodes an RNA-binding protein
AT1G58250	SABRE, putative gene of unknown function, homologous to maize apt1 gene.
AT1G58100	Encodes TCP8, belongs to the TCP transcription factor family known to bind site II elements in promoter regions.
AT1G58450	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins.
AT1G59730	thioredoxin H-type 7 (TH7)

**Appendix table 6A1. List of non NB-LRR encoding genes in the *RAL4* locus corresponding region in Col-0. Gene model predictions and annotations between Arabidopsis Chr1 positions 21504314 and 21953456 were extracted from TAIR (Arabidopsis.org).**

## List of primers used in this study

Use, Name, Sequence, Design credits (if not OJF)

### **Albugo laibachii genotyping primers**

RXLR10 Fw OF13 CGTCAACGTAGACCTGTGC  
RXLR10 Rv OF14 GCCTAACTTGTCACACCTGC  
rDNA ITS Fw DC6 GAGGGACTTTTGGGTAATCA  
rDNA ITS Rv LR0 GCTTAAGTTCAGCGGGT Moncalvo et al, 1995

### **SSP16 sequencing and cloning**

#### **Genomic Sequencing**

Fw OF55 CCAGAACGAATTCTACTGCG  
Rv OF56 CATGATGAAACCATCTTAACAATGC

#### **RACE-PCR**

fw OF76\_3'\_nested\_close GCGTCAGCCAACATTCTTAAAGAAATTC  
Rv OF\_77\_5'\_nested\_close GTTCTGGGTACTTGCCATCTTAC  
fw OF84\_3' CAGTGAAGTGATCAATCAACACACGATTGTTCTAC  
rv OF85\_5' GCCATTTGGACTGAGCCGACCATATTGC  
fw OF86\_3'\_nested CTGACTCTGAGGATTGGGTATTGGAGATTGGC  
rv OF87\_5'\_nested TCTCTGCTGTGTGCAGCCCTAACATGTACAC

#### **Gateway cloning**

SSP16\_TOPO\_start Fw OF88\_SSP16 CACCATGGGTTTCAAAAAAGTCAATC  
SSP16\_TOPO\_stop Rv OF89\_SSP16 TTAGTCGAGTTCAAGGTGAATCTTGCTTTG  
SSP16\_TOPO\_start\_no\_sp Fw OF95\_SSP16 CACCATGTGCTTCATCAAGAAGGAAGCGACGTGC  
SSP16\_abo1\_var\_allele\_spec Rv OF110\_SSP16 TTAGTCGAGTTAAGGTGAATCTGGCCTTTG

#### **GoldenGate cloning**

SSP16\_GG\_16 Fw OF\_115\_SSP16 tatggtctcaaATGTGCTTCATCAAGAAGGAAGCGACG  
SSP16\_GG\_16 Rv OF\_121\_SSP16 tataGGTCTCtaagcTTAGTCGAGTTCAAGGTGAATCTTGCTTTGG

#### **Truncations (GoldenGate)**

SSP16-truncation1 Fw OF127-16T2-GG tataGGTCTCtaATGGGTGCATCTTGAACAATTTCCACTACGAC  
SSP16-truncation2 Fw OF136-16T6-GG tatggtctcaaATGGAAGCAAAACCGACATGGGAGGTGGTGTGTTTTTAAG  
SSP16-truncation3 Fw OF130-16T5-GG tataGGTCTCtaATGAGTAAGTATCGTGTGCTGTTTGGAAAAATGAC

#### **GFP-tagging (GoldenGate)**

GFP Fw OF292-GFP-N-ter88-F tataGGTCTCaaATGATGGTGAGCAAGGGCGAGG  
GFP Rv (SSP16 fusion) OF293-GFP-N-ter88-R-fusion\_16\_spec tataGGTCTCaAGCAGATCTAATAGCCGCGTTTTGTACAGCTC  
SSP16 Fw (GFP fusion) OF294-SSP16-F-GFP-N-ter-fusion tataGGTCTCaTGCTTCATCAAGAAGGAAGCGACG

### **SSP18 sequencing and cloning**

#### **Genomic Sequencing**

Fw OF65 CATTTCGAATTAGAACGCCAATGC  
Rv OF66 GCTCACTGCCTTCTTACGATCA

#### **Gateway cloning**

SSP18\_TOPO\_start Fw OF\_98\_SSP18 CACCATGCTTTCCGCTCCAGTGC  
SSP18\_TOPO\_stop Rv OF99\_SSP18 CACCATGCGAAATGCTCTGCG  
SSP18\_TOPO\_start\_no\_sp Fw OF100\_SSP18 TTATTTTTGGACCGCTTTTTACCGGAG

#### **GoldenGate cloning**

SSP18\_GG Fw OF\_120\_GG\_18-F tatggtctcaaATGCGAAATGCTCTGCGAATCGAGTCAGAAACGG  
SSP18\_GG domestication1 Rv OF117-SSP18-dom1-GG CTGgatgacAATGTAGTCATGAAACCGtattatGTCTTctata  
SSP18\_GG domestication2 Fw OF118-SSP18-dom2-GG tataGAAGCttaaataAAAATCAAAATGCACACCACACTGAACATTCC  
SSP18\_GG Rv OF119-SSP18-dom3-GG tataGGTCTCtaagcTTATTTTTGGACCGCTTTTTACCGGAggtcGc

#### **GFP-tagging GoldenGate**

GFP Rv (SSP18 fusion) OF295-GFP-N-ter88-R-fusion\_18\_spec tataGGTCTCaGTGAGATCTAATAGCCGCGTTTTGTACAGCTC  
SSP18 Fw (GFP fusion) OF296-SSP18-F-GFP-N-ter-fusion tataGGTCTCaTACGAAATGCTCTGCGAATCGAG

#### **SSP15 sequencing and cloning**

#### **Genomic Sequencing**

Fw OF32 GCTTTCATCAATTCACGTTTTG  
Rv OF33 TCAGGAGAGCTGGAGGATCC

#### **Gateway cloning**

SSP15\_TOPO\_start Fw OF34 CACCATGGTACAACACAAGCG  
SSP15\_TOPO\_stop Rv OF36\_stop TTAGTCTCTATTAATAAAATTCATAAAACCACG  
SSP15\_TOPO\_start\_no\_sp Fw OF96\_TOPO\_start\_no\_sp CACCATGCTTACTACCTCGCATGA

### **SSP17 sequencing and cloning**

#### **Genomic Sequencing**

Fw OF63 CTAGTCCGGATGAAAAACTCATGG  
Rv OF64 CCAATTATTCAGTATGAATCACAATCC

#### **Gateway cloning**

SSP17\_TOPO\_start Fw OF82\_RxL5\_TOPO\_start CACCATGAATCACAATCCTTCC  
SSP17\_TOPO\_stop Rv OF83\_RxL5\_TOPO\_stop CTATAATTTATTAGATCGGCTCCCTAGTCC  
SSP17\_TOPO\_start\_no\_sp Fw OF97\_TOPO\_start\_no\_sp CACCATGGAGTCAAATCGTGCCTTA  
Gateway clone verification

M13 Fw M13F TGTAACGACGGCCAGTG Unknown

M13 Rv M13R CAGGAAACAGCTATGACCATG

**GoldenGate clone verification**

piCH86988 clones vc\_197 GTAACATCGCTGCAATCCACCATg Volkan Cevik  
vc\_198 cgaaccgcggttaaggatctg Volkan Cevik  
piCH86966 clones VC\_199 gccggtcttgcatgattatc Volkan Cevik  
VC\_200 ggttctgtggttgccacatac Volkan Cevik

**Ws-2xHR-5 SSLP markers**

chr1\_M15\_3.2mb Fw OF173 GAACAGATGTAAGAAACATTGGGTTCTCCTTTTAC  
chr1\_M15\_3.2mb Rv OF174 CTGAGACCGTCAATCGTGTGCAAG  
chr1\_M17\_11.7mb Fw OF177 GAAAGAAAGAATGAATTGCTACTGTGTTAAAAATGATAAG  
chr1\_M17\_11.7mb Rv OF178 CTTATCATTTTTAACACAGTAGCAATTCATTCTTTCTTTC  
chr1\_M11\_17.14mb Fw OF165 CATCCTAGATGCGAAAGATAACATAGAAAGCC  
chr1\_M11\_17.14mb Rv OF166 GTTCTCATCATGCAAATTAATTTTATTGCAATTATGATTAGTCC  
chr1\_M21\_19.8mb Fw OF185 CGTTGAGAGTAGATTATAGTGAAGCAAACC  
chr1\_M21\_19.8mb Rv OF186 CGTACTACTTGAATCGGCGCATTATGTTTC  
chr1\_M19\_21.4mb Fw OF181 GGAAGTGCATTTTATTTGCTTAGCAAATGATGC  
chr1\_M19\_21.4mb Rv OF182 GAAGTAAAAGGACCAACGAGAATCAACCC  
chr1\_M26\_21.63Mb Fw OF217 GACAAGCTCGGTGAAGTTAGAGGTG  
chr1\_M26\_21.63Mb Rv OF218 TAAGCTGTGGCAAAGATAAACGGAC  
chr1\_M27\_21.9Mb Fw OF219 CCTCATTGGCTCAACCTGGAAAAATCAATATTC  
chr1\_M27\_21.9Mb Rv OF220 GCTCTTATCACGACCGACTGTACC  
chr1\_M28\_21.9Mb Fw OF221 GCAATAGAATTTAGAATATCACTTCATGTTACGTCGTAC  
chr1\_M28\_21.9Mb Rv OF222 GATGTTTGAATCATATCGTATAGCTCTTCTCAAATGG  
chr1\_M29\_21.9Mb Fw OF223 GGTACCATCTTATCTTTGCCCTTCTCG  
chr1\_M29\_21.9Mb Rv OF224 GTGCTTGTACTGATAGTAGTATGAAACCCCTGAG  
chr1\_M30\_22.2Mb Fw OF229 CTTGGGAAAGAAGTACTCTATATATAGAGAGGC  
chr1\_M30\_22.2Mb Rv OF230 GTTAGAAGAAAATAAAATGTCAAACCTCAACTCAATTTGTTTC  
chr1\_M31\_22.6Mb Fw OF231 CGGTGATTCACTCAACTACCGTGTTCAC  
chr1\_M31\_22.6Mb Rv OF232 CGATGATAGTGTCTCTTTGGTAGCAGTATGA  
Chr2:7531219 Fw WA110 TTCCGTGGGAGTTGGAGGAAGAC Wiebke Apel  
Chr2:7531219 Rv WA111 ctttcgcttattctcagctatg  
Chr2:10176739 Fw WA90 gattgaattcttatcgagagatg Wiebke Apel  
Chr2:10176739 Rv WA91 catcaattacagtataacactaacc  
Chr2:16184272 Fw Map8 ATGTCCAATTGACCAACCG Torsten Schultz-Larsen  
Chr2:16184272 Rv Map8 CAAAATAACACCCCAACT  
Chr3:786320 Fw Map10 CATCCGAATGCCATTGTTC Torsten Schultz-Larsen  
Chr3:786320 Rv Map10 AGCTGCTTCTTATAGCGTCC  
chr3\_M4\_10mb Fw OF151 GGAAGATAACGTCCATTGATCGCACTAGA  
chr3\_M4\_10mb Rv OF152 GGTGAGAGTGTATGTTACAAGACTAGATTTATCTGAA  
Chr3\_19.133Mb-F VC190 ggctcgtgtcgtgttggtcgcgtc Volkan Cevik  
Chr3\_19.133Mb-R VC191 gtggttctttggagagaaatccact  
chr4\_M9\_0.13mb Fw OF161 GACCCCGATCGTTCCTGATT  
chr4\_M9\_0.13mb Rv OF162 GTGTTATTTAATCAGTGAAACTGCCAC  
chr4\_M10\_3.3mb Fw OF163 CTTTCCACTTTGATACCTCTTGAGAGATTG  
chr4\_M10\_3.3mb Rv OF164 GCAAGAGCTTGAGTGAAGCTCACAG  
chr5\_M12\_3.5mb Fw OF167 CAAGCTGATGAGGAGAGTGGTCCG  
chr5\_M12\_3.5mb Rv OF168 CTTGGTCATCAACATTGCCAAATAGTTAGTGG  
chr5\_M7\_18mb Fw OF157 TGCGGAAGAAGCGCAAGG  
chr5\_M7\_18mb Rv OF158 GTTCCTTATTCTACAGAATAAAGCCTG

**Ws-2xHR-5 Sequencing markers**

SM1 Fw OF243 CAAAACACACAACCTAGAGCAGCACC  
SM1 Rv OF244 CCACATTGATGACACTTGTGAGGAAGG  
SM2 Fw OF245 CACTGGGCTTATTGTCGGTCCCT  
SM2 Rv OF246 CTGTTTAGAATCTGTTGGCGGCCATC  
SM3 Fw OF247 GAAGTAATGCAAGATGAGAATCCAATAAGTGTGTC  
SM3 Rv OF248 CAAGAGTTCCAGATTTGGAGAATCTTGAAG

**RAL4 candidate cloning**

**GoldenGate**

**For piCH86988**

58390-GG Fw OF253 tataGGTCTCaaATGGCTGGAGAACTTGTGTCGTTTG  
58390-dGG Rv OF254 tataGGTCTCttgtgATAATCTTCGGCGGTGGATATTCC  
58390-dGG Fw 2 OF335 tataGGTCTCtacaATGcAgagacTATTCAAGATGTTGG  
58390-dGG Rv 2 OF256 tataGGTCTCtactAGATTTCTCAAACCTAACCCTGTCTTCTGTG  
58390-dGG Fw 3 OF336 tataGGTCTCtagtaAAGTTGgagacTTTGGTTTATTTCTGcAC  
58390-GG Rv 3 OF290 tataGGTCTCtaagcTCAATTGAAGTAGCCTCCTATGAATTCACCG  
58400-GG Fw OF225 tataggtctcaaATGGTTCGAGGCAATTGTTTCATTTGG  
58400-dGG Rv OF337 tataGGTCTCtACGAGATTACACAacCCCAACTTTATCTC  
58400-dGG Fw OF227 tataGGTCTCtTCGTAAACTTGgaAaccTTAGAGAATTTCTCAAC  
58400-GG Rv OF226 tataGGTCTCtaagcTTATTTGTAGTCTTTTCGAATTTAACAGAAGG  
59620-GG Fw OF285 tataGGTCTCaaATGGCTGAGACACTTTTGTCAATTTGGAG  
59620-GG Rv OF286 tataGGTCTCtaagcTAAAGAAATCGAACAAGAGGAATGTGTTGGAC

59780-GG Fw OF251 tataGGTCTCaaATGCAGGACTTATATATGGTTGATTCAATTGTATCG  
59780-GG Rv OF252 tataGGTCTCtaagcTCAGATGATTGGACTAGGGAAAGAATATATCACC

**USER**

58390HR-5\_pro Fw OF343 GGCTTAAUAGGTAATGCTATTGTATATCATCCCTATACAAATTGA  
58390HR-5 Rv OF349 GGTTAAUTCATTTGAAGTAGCCTCTATGAATTCAACG  
58400HR-5\_pro Fw OF350 GGCTTAAUAGCTTGCATGTGCCTTCATTGTGTT  
58400HR-5 Rv OF351 GGCTTAAUTCATTTGTAGTCCTTTTCAAAATTAACAGAAGG  
59620HR-5\_pro Fw OF347 GGCTTAAUAGAGATATGGATCTGGCGGCC  
59620HR-5 Rv OF348 GGTTTAAUUTAAAGAAATCGAACAAAGAGGAATGTGTTGGAC

**RAL4 candidate cloning sequencing**

OF272-58390-seqF1 GTTGAAGCATCGAAATCATTAAATGTCTTTG  
OF273-58390-seqF2 GATTTGGACAGAGCTACCTAGAGG  
OF274-58390-seqF3 CCTTGAATATCTTTATATTGTGGGTACTCACTCT  
OF234-At1G58400 tcgtaaacctggagaccttagagaatttc  
OF235-At1G58400 ccttctgttaaatcgaaaaggactacaataa  
OF236-At1G584005pr600rev ACACCCATGCGAGTCTATCAAAT  
OF237-At1G58400mid-rev CGTTTCCAATCATGAAAAGTATTTTTGCGAG  
OF238-At1G5840053pr800rev CTTCTTCAAACCTGGCTTTATAGAGATCTAATAC  
OF239-At1G5840053pr800fwd GAGTTGGAAGCTATTAGGTTCAAGCT  
OF240-At1G5840053pr400fwd CTGATCAACACACTTCCCTTCTCAC  
OF241-At1G58400-promfwd GTATCTTAACGCGAAATGAAACATTCAACACT  
OF242-At1G58400-promrv CATAATGGCTCCTGCTAAGTAGTTGAAAAATAATAATCAA  
BC\_C3f\_59620 AGTGTCTTGTGGGGTTGGA Baptiste Castel  
BC\_C3r\_59620 ATATATTCTGGCCCAACTTTCC  
OF275-59780-seqF1 CAGAAGAAACTCTTTCAGTTGTTGAAAC  
OF276-59780-seqF2 GAGCCAACACTACGCTCATTTAATGAGA  
OF277-59780-seqF3 CGGTGATTGATTTTTGACTGCATAATGTTG  
OF278-59780-seqF4 GTATTTGAGTTTATATCAGGCATCTGTAACCTA  
OF279-59780-seqF5 CCTATGTTGGGAGGAGAATGGTTTGC

**Albugo laibachii SSLP1**

AISSLP1 Fw OF283 CTTCACTTTGTCATCACCACACAG  
AISSLP1 Rv OF284 CAGTGACCACAGAGTACTTTTATGC

**NDR-1 genotyping**

NDR-1\_spec Fw OF352 GTGTGTCTACTGAGTC  
NDR-1\_spec Rv OF353 TCACAGCTGGTCTCACCT

**Illumina library amplification**

P5 AATGATACGGCGACCACCGA  
P7 CAAGCAGAAGACGGCATAACGA

**Barcoding of RenSeq libraries**

index\_8nt\_1 AJI\_1 CAAGCAGAAGACGGCATAACGAGATcgttggttGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT Agathe Jouet (all)  
index\_8nt\_2 AJI\_2 CAAGCAGAAGACGGCATAACGAGATtctggttGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_5 AJI\_3 CAAGCAGAAGACGGCATAACGAGATggcggttGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_11 AJI\_4 CAAGCAGAAGACGGCATAACGAGATtagtctgtGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_26 AJI\_5 CAAGCAGAAGACGGCATAACGAGATtgggtcttGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_130 AJI\_6 CAAGCAGAAGACGGCATAACGAGATtagagttctGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_294 AJI\_7 CAAGCAGAAGACGGCATAACGAGATccattggGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_298 AJI\_8 CAAGCAGAAGACGGCATAACGAGATccagctggGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_320 AJI\_9 CAAGCAGAAGACGGCATAACGAGATgcagacggGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_325 AJI\_10 CAAGCAGAAGACGGCATAACGAGATaccggaggGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_331 AJI\_11 CAAGCAGAAGACGGCATAACGAGATcctgtaggGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_398 AJI\_12 CAAGCAGAAGACGGCATAACGAGATatgaatagGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_468 AJI\_13 CAAGCAGAAGACGGCATAACGAGATtataactcGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_520 AJI\_14 CAAGCAGAAGACGGCATAACGAGATtacgtagcGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_534 AJI\_15 CAAGCAGAAGACGGCATAACGAGATcatactccGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_546 AJI\_16 CAAGCAGAAGACGGCATAACGAGATatacccgtGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_554 AJI\_17 CAAGCAGAAGACGGCATAACGAGATcaactaccGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_561 AJI\_18 CAAGCAGAAGACGGCATAACGAGATgctgaaccGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_655 AJI\_19 CAAGCAGAAGACGGCATAACGAGATatataagaGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_682 AJI\_20 CAAGCAGAAGACGGCATAACGAGATcgtcgccaGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_687 AJI\_21 CAAGCAGAAGACGGCATAACGAGATgtcaaccaGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_699 AJI\_22 CAAGCAGAAGACGGCATAACGAGATgaccgaaGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
index\_8nt\_703 AJI\_23 CAAGCAGAAGACGGCATAACGAGATgttcagaaGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT  
AJ\_universal\_primer AJI\_U AATGATACGGCGACCACCGAGATCTACTCTTTCCCTACACGACGCTCTCCGATC\*T

**RRS1 qPCR**

Fw OF\_SS\_29 GGTAAGAAATCCTCCATGGACAA  
Rv OF\_SS\_36 AGATGAGGCAGAGGTAGTTATGG