THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Understanding the structure and variation within the genome of the pathogenic ascomycete fungus *Leptosphaeria maculans*

Manuel Matias Zander

Bachelor of Biotechnology (Honours)

Graduate Certificate in Research Commercialisation

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2015

School of Agriculture and Food Science

Abstract

The ascomycete fungus *Leptosphaeria maculans* is a major pathogen of *Brassica* species, particularly canola (*Brassica napus*; rapeseed; oil-seed rape), and the primary cause of crop losses of canola in Australia, causing blackleg disease. *L. maculans*, a filamentous ascomycete, is the causal agent of phoma stem canker, commonly referred to as blackleg. In late stages of infection, it spreads through the stem vasculature causing lesions, leading to poor growth, lodging and eventually plant death. This fungus is found in canola-growing regions worldwide such as Australia, Canada and Europe. Increased production of canola in these regions has led to a rise in the severity of the disease. In Australia alone, *L. maculans* infection is responsible for an estimated Australian $100 million in crop losses each year, with average losses ranging from 15–48 % and significant efforts are underway to improve resistance to this disease.

Understanding the characteristics of *L. maculans* is vital for developing an effective and sustainable approach to the management of blackleg disease on *Brassica* species. The completion of the *L. maculans* genome sequence was a significant development in the study of this fungal pathogen and provides a reference genome to which molecular markers can be physically mapped. This has been highly useful in other plant pathogens with sequenced genomes, such as the wheat pathogen *Parastagonospora nodorum* and the cereal pathogen *Fusarium graminearum*. Importantly, a reference genome also allows mapping of whole genome re-sequencing data, which is becoming a high-throughput, cost-effective method to study genome-wide diversity, particularly for the relatively small, lower complexity genomes of many fungal species. By re-sequencing the genome of different *L. maculans* isolates, variations in genome sequence and structure can be elucidated.

Advances in genome sequencing technologies have revolutionised plant and fungal genomics. They have made genome sequencing, re-sequencing and Single Nucleotide Polymorphism (SNP) discovery highly accessible, high-throughput and cost-effective. The process of whole genome re-sequencing involves aligning millions of short sequence reads to a reference genome sequence. Once this has been achieved, it is possible to identify genetic variation between individuals, which can be linked to variation in phenotype to provide molecular genetic markers and insights into gene function. Sequence variation can have a major impact on how an organism develops and responds to the environment.

This thesis describes the implementation of several approaches to elucidating the genome structure and variation of a number of *L. maculans* isolates, including SNPs and presence/absence variations (PAVs).

Initially, the re-sequencing of two *L. maculans* isolates for the identification of 21,814 SNPs was performed. I demonstrated the application of a novel SNP calling method, SGSautoSNP and its robustness and sensitivity in identifying polymorphisms in *L. maculans*. I described the use of these SNPs for phylogenetic analysis, genome analysis, including SNP properties and density in relation to genomic position and predicted function. This method correctly predicted polymorphisms in *AvrLm* genes, which are important in the pathogen's interaction with its host plant. Whole-genome polymorphic trends such as genome-wide SNP density and transition/transversion ratios were also determined with this approach. The SNPs from this study were subsequently applied for the genotyping of 59 *L. maculans* isolates from around Australia in a separate study conducted within our group (Patel et al., 2015).

Subsequently a larger scale SNP prediction was performed using ten *L. maculans* isolates with known avirulence (*AvrLm*) gene content, based on infection studies toward *Brassica* species. The genome re-sequencing of these isolates was performed and yielded high genome coverage, ranging from 26 times to 266 times coverage. This resulted in the identification of 47,097 SNPs with an average of 1 SNP every 953 bp. This provides a greater resource for further study of *L. maculans* variation across individuals and populations than previous work. Genome analysis was performed and analyses of SNP properties and positions were undertaken. Importantly, the SGSautoSNP prediction correctly predicted the mutations within *AvrLm* genes in these isolates, indicating that the infection assays and the computational approaches can complement each other and indeed can be used to determine novel infection-related genes.

Furthermore, prediction and analysis of presence/absence variations (PAVs) was undertaken in order to understand genome structure and variation within *L. maculans*. The PAVs allow a better understanding of larger polymorphisms within this genome that can be several hundred base pairs (bp) long, as compared to the singular nature of SNPs. I analysed the positions and the occurrence of these variations and their effect on the genome in both coding and non-coding regions of the ten *L. maculans* isolates used in the SNP analysis. Results of these analyses indicate that a number of highly variable regions

exist within the genome of *L. maculans.* This was particularly evident on SuperContig 13, where a number of secondary metabolite genes, of which some are involved in plant infection processes, are located. Other genes of interest, such as genes involved in fatty acid metabolism, antibiotic and antifungal resistance, were also shown to be affected by PAVs. This approach was also effective in identifying the presence/absence of avirulence genes, known to be present or lost from previous studies, such as *AvrLm1* and *AvrLm6*.

The aim of this project was to contribute to a greater understanding of *Leptosphaeria maculans* and its genomic qualities, and that in turn this can help improve efforts to reduce the occurrence of this disease in Australia and abroad. An improved understanding of this pathogen will aid in developing more resistant plant varieties and thus improve the yields but also the sustainability of the canola industry for the long term.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Journal Publications:

Zander, M., Alamery, S., Lorberg, S.K., Patel, D.A., Tollenaere, R., Edwards, D. and Batley, J. (2015) Understanding Effector-Triggered Immunity of Plants to Pathogens: A Genomics Approach. *Plant Biotechnology Journal* (Submitted)

Patel, D.A., **Zander, M**., Van de Wouw, A.P., Mason, A.S., Edwards, D. and Batley, J. (2015) Population diversity of *Leptosphaeria maculans* in Australia. *International Journal of Biology* **7**, 18-36

Golicz, A.A., Martinez, P.A., **Zander, M**., Patel, D.A., Van De Wouw, A.P., Visendi, P., Fitzgerald, T.L., Edwards, D. and Batley, J. (2015) Gene loss in the fungal canola pathogen *Leptosphaeria maculans*. *Functional & Integrative Genomics* **15**, 189-196

Lai, K., Lorenc, M.T., Lee, H.C., Berkman, P.J., Bayer, P.E., Visendi, P., Ruperao, P., Fitzgerald, T.L., **Zander, M**., Chan, C.-K.K., Manoli, S., Stiller, J., Batley, J. and Edwards, D. (2015) Identification and characterization of more than 4 million intervarietal SNPs across the group 7 chromosomes of bread wheat. *Plant Biotechnology Journal* **13**, 97-104

Chambers, K., Lowe, R., Howlett, B., **Zander, M**., Batley, J., Van de Wouw, A. and Elliott, C. (2014) Next-generation genome sequencing can be used to rapidly characterise sequences flanking T-DNA insertions in random insertional mutants of *Leptosphaeria maculans*. *Fungal Biology and Biotechnology* 1, 10.

Fraser-Smith, S., Czislowski, E., Meldrum, R.A., **Zander, M**., O'Neill, W., Balali, G.R. and Aitken, E.A.B. (2014) Sequence variation in the putative effector gene *SIX8* facilitates molecular differentiation of *Fusarium oxysporum* f. sp cubense. *Plant Pathology* **63**, 1044-1052.

Zander, M., Patel, D.A., Van de Wouw, A., Lai, K.T., Lorenc, M.T., Campbell, E., Hayward, A., Edwards, D., Raman, H. and Batley, J. (2013) Identifying genetic diversity of avirulence genes in *Leptosphaeria maculans* using whole genome sequencing. *Functional & Integrative Genomics* **13**, 294-308

Lai, K., Duran, C., Berkman, P.J., Lorenc, M.T., Stiller, J., Manoli, S., Hayden, M.J., Forrest, K.L., Fleury, D., Baumann, U., **Zander, M**., Mason, A.S., Batley, J. and Edwards, D. (2012) Single nucleotide polymorphism discovery from wheat next-generation sequence data. *Plant Biotechnology Journal* **10**, 743-749

Hayward, A., Dalton-Morgan, J., Mason, A., **Zander, M**., Edwards, D. and Batley, J. (2012) SNP discovery and applications in *Brassica napus*. *Journal of Plant Biotechnology* **39**, 1-12.

Book Chapters:

Patel, D.A., **Zander, M**., Dalton-Morgan, J. and Batley, J. (2015) Advances in plant genotyping: where the future will take us. In: Plant Genotyping (Batley, J. ed) pp. 1-11. Springer New York.

Edwards, D., **Zander, M**., Dalton-Morgan, J. and Batley, J. (2012) New technologies for ultra-high throughput genotyping in plant taxonomy. In: Molecular Plant Taxonomy (Besse, P. ed) pp. 151-175. Humana Press.

Conference Abstract:

(A) Oral Presentation

Zander, M., Patel, D.A., Van de Wouw, A., Lai, K., Lorenc, M.T., Hayward, A.C., Golicz, A., Chan, K., Bayer, P., Edwards, D., and Batley, J. (Jan 2014) New Insights into the Genome of the Plant Pathogen, *Leptosphaeria maculans*. Plant and Animal Genome Conference XXII, San Diego, California, USA.

Zander, M., Patel, D.A., Van de Wouw, A., Lai, K., Lorenc, M.T., Hayward, A.C., Golicz, A., Chan, K., Bayer, P., Hurgobin, B., Lee, J., Edwards, D., and Batley, J. (Sep 2015) Understanding the structure and variation within the genome of the pathogenic ascomycete fungus *Leptosphaeria maculans*. Pathogen Bioinformatics Workshop, Australian Plant Pathology Society Conference, Fremantle, Western Australia, AUS.

(B) Poster Presentation

Zander, M., Patel, D.A., Van de Wouw, A., Lai, K., Lorenc, M.T., Hayward, A.C., Golicz, A., Chan, K., Bayer, P., Edwards, D., and Batley, J. (Jan 2014) Understanding genome structure and diversity of the plant pathogen, *Leptosphaeria maculans*. Plant and Animal Genome Conference XXII, San Diego, California, USA.

Zander, M., Patel, D.A., Van de Wouw, A., Lai, K., Lorenc, M.T., Hayward, A.C., Golicz, A., Chan, K., Bayer, P., Edwards, D., and Batley, J. (March 2014) Understanding Genome Structure and Diversity of the Plant Pathogen, *Leptosphaeria maculans*. Crucifer Genetics Workshop and Brassica2014 Conference, Wuhan, China.

Publications included in this thesis

Zander, M., Patel, D.A., Tollenaere, R., Alamery, S., Lorberg, S.K., Edwards, D., Batley, J. (2015) Understanding Effector-Triggered Immunity of Plants to Pathogens: A Genomics Approach. Plant Biotechnology Journal (Submitted).

Incorporated in Chapter 1.

Contributor	Statement of contribution
Manuel Zander (Candidate)	Wrote the paper (40%)
Dhwani A. Patel	Wrote the paper (15%)
Reece Tollenaere	Wrote the paper (15%)
Salman Alamery	Wrote the paper (15%)
Sarah Kay Lorberg	Wrote the paper (15%)
David Edwards	Edited paper (40%)
Jacqueline Batley	Edited paper (60%)

The submitted manuscript of this publication can be seen in Appendix 1.

Zander, M., Patel, D.A., Van de Wouw, A., Lai, K.T., Lorenc, M.T., Campbell, E., Hayward, A., Edwards, D., Raman, H. and Batley, J. (2013) Identifying genetic diversity of avirulence genes in *Leptosphaeria maculans* using whole genome sequencing. *Functional & Integrative Genomics* **13**, 294-308.

Incorporated as Chapter 3.

Contributor	Statement of contribution
Manuel Zander (Candidate)	Designed experiments (60%) Conducted experiments (70%) Wrote the paper (90%)
Dhwani A. Patel	Conducted experiments (15%) Wrote the paper (10%)
Angela Van de Wouw	Edited paper (10%)
Kaitao T. Lai and Michal T. Lorenc	Conducted experiments (10%)
Emma Campbell	Conducted experiments (5%)
Alice Hayward	Designed experiments (10%) Edited paper (10%)
Dave Edwards	Designed experiments (10%) Edited paper (10%)
Harsh Raman	Edited paper (10%)
Jacqueline Batley	Designed experiments (20%) Edited paper (60%)

Contributions by others to the thesis

Chapter 4

Mr Pradeep Ruperao and Ms Agnieszka Golicz provided bioinformatics support and ran the SGSautoSNP and SnpEff pipelines on the sequence data. Ms Dhwani Patel and Dr Satomi Hayashi aided in the isolation of fungal DNA and in all aspects of genomic DNA sequencing of the samples.

Chapter 5

Dr Kenneth Chan, Ms Bhavna Hurgobin and Ms Jenny Lee developed the PAKAP pipeline for PAV prediction and provided ongoing bioinformatics support throughout the analysis of the data. Dr Satomi Hayashi, Ms Dhwani Patel and Ms Sarah Lorberg aided in the isolation of fungal DNA and in all aspects of genomic DNA sequencing of the samples.

All Chapters

Supervisors Prof. Dave Edwards and Prof. Jacqueline Batley provided assistance and guidance in experimental design and editing of writing.

Statement of parts of the thesis submitted to qualify for the award of another degree

None

Acknowledgements

First of all, I would like to thank the University of Queensland for providing me with the infrastructure and facilities to conduct my research and for being my home for the past 8 years. Particular acknowledgement goes to the Grains Research and Development Corporation (GRDC), the Scherago International organisation and School of Agriculture and Food Science (SAFS) for awarding me with funding for travel to conferences in the United States and in China. These opportunities have greatly enhanced my PhD experience.

Sincere appreciation goes to the many friends and colleagues that have made this journey so pleasurable and tolerable over the years. A special thanks to all the staff and students, past and present, of the CILR, especially to Dr Yu-Hsiang Lin, Dr Meng-Hang Lin, Dr Satomi Hayashi, Bethany Van Hameren and Elton Moraes for your friendship.

To all Batley lab members: past and present, who have taught me so much about science, companionship and about myself; the most sincere thanks. This includes Emma Gardiner and Dr Alice Hayward for teaching me most of my lab skills and having patience with me in my first proper lab experiences. A huge thanks to the Applied Bioinformatics Group of Prof. David Edwards, we couldn't do the things we do without you. A special thanks goes to Dhwani Patel, Reece Tollenaere and Sarah Lorberg for your constant encouragement and friendship, it made the road so much more fun.

Thank you Prof. Jacqueline Batley for the wonderful opportunities you've given me and for believing in me from the beginning, I have the utmost respect and admiration for you.

To my friends and family from the Russian Baptist Church in Brisbane, thank you for your prayers and support, especially Rev Emil Rahimov for your care. A special mention goes to Dr. Artak Amirbekjan for his informal mentorship along this journey.

I wouldn't be here without my parents Viktor and Irene and my sister Dorothea. They have believed in me, even when I didn't believe in myself. They know where I've come from and how much this means to all of us. To my lovely girlfriend Svetlana, you are amazing and beautiful, and you are truly worth fighting for.

Last but certainly not least, I would like to thank God, for giving my life meaning and purpose. Knowing you has changed my life and without you I would not be here today.

~Soli Deo Gloria~

Keywords

Blackleg, fungi, plant pathogen, avirulence, next generation sequencing, Brassica, genomics, SNP, presence absence variation, disease

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060408, Genomics, 60%

ANZSRC code: 060704, Plant Pathology, 20%

ANZSRC code: 060505, Mycology, 20%

Fields of Research (FoR) Classification

FoR code: 0604 Genetics, 60%

FoR code: 0607 Plant Biology, 20%

FoR code: 0605 Microbiology, 20%

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AvrLm	Avirulent towards *Rlm* gene
Bp	Base pair
CDS	Coding Sequence of gene
ETI	Effector-Triggered Immunity
LRR	Leucine Rich Repeat
Mb	Mega base (1 million bp)
MP	Mate Pair
NBS	Nucleotide Binding Site
NGS	Next Generation Sequencing
PAV	Presence/Absence Variation
PE	Paired End
PIC	Polymorphism Information Content
RIP	Repeat-Induced Point mutation
Rlm	Resistance to *Leptosphaeria maculans* gene
SC	SuperContig
SNP	Single Nucleotide Polymorphism
Ti/Tv	Transition to Transversion ratio

1. General Introduction

Sections of this Chapter have been submitted as a review paper in Plant Biotechnology Journal. Included as Appendix 1.

1.1. Introduction to *Brassicas* (Importance of *Brassica* crops)

The Brassicaceae family comprises some of the most economically important food crops in the world. Of particular significance are the *Brassica* species, which include more agricultural and horticultural crops than any other plant genus (Taiyan et al., 2001). The six major cultivated *Brassica* species exist as either diploids; *B. rapa* (AA genome), *B. nigra* (BB genome) and *B. oleracea* (CC genome) or allotetraploids, which contain two diploid sets of chromosomes; (*B. napus* (AACC), *B. juncea* (AABB) and *B. carinata* (BBCC)). The genetic relationship between these species was first discovered by Morinaga (1934) and visualised by Dr Woo Jang-choon in 1935 (U, 1935). These researchers concluded that the three cultivated allotetraploid species were derived from spontaneous interspecific hybridization between the three diploid progenitors (Figure 1.1).

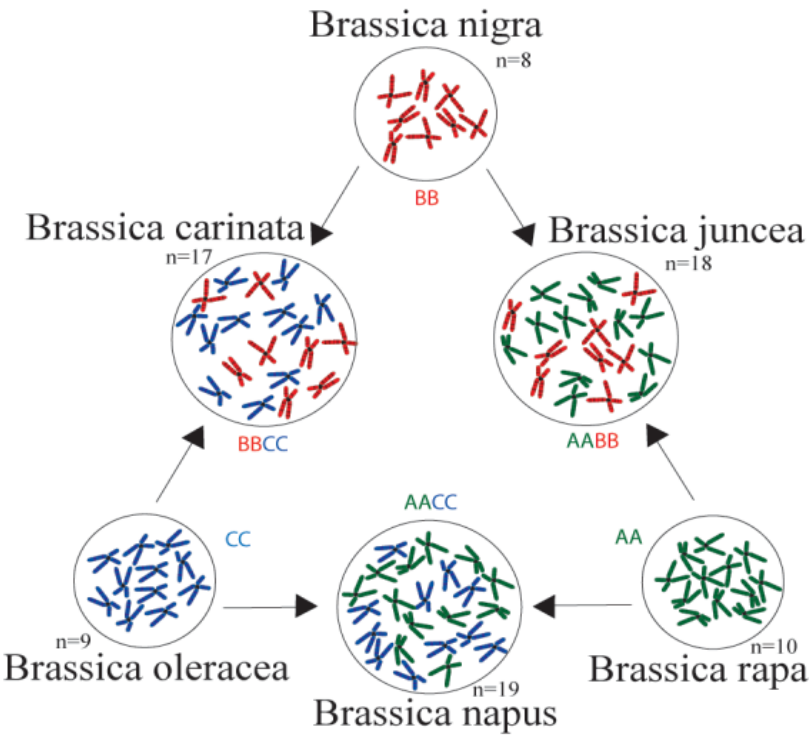

Figure 1.1: The Triangle of U, depicting the relationship between the *Brassica* genomes.

Brassica napus is primarily grown for its seeds which are processed to extract vegetable oil, known as canola oil, but *B. napus* varieties are also grown for the stem and leaves, which are commonly used in Eastern cuisine. The leaves and stem of *B. juncea* are widely used in African and Asian cooking and the seeds are used to produce brown mustard. However in recent decades *B. juncea* has also been bred as an oilseed crop due to its ability to grow in semi-arid territories (Edwards et al., 2007b). *Brassica rapa* is also grown as an oilseed, but is particularly popular in Asia as a cruciferous vegetable (e.g. chinese cabbage, bok choi, pak choi, turnips). *Brassica carinata* (Ethiopian mustard) is grown as an oilseed crop and leafy vegetable in Ethiopia, while *B. nigra* (black mustard) is grown as a condiment mustard seed. The most morphologically diverse cruciferous *Brassica* species is *B. oleracea,* which encompasses cabbages, broccoli, cauliflower and brussels sprouts.

Four of the 'U' triangle species are thus cultivated as oilseed crops. The oil extracted from the seed is widely used for both human consumption and industrial purposes. *Brassica napus* (canola/oilseed rape) is the most economically important of these *Brassica* oilseeds, and is responsible for the majority of the world's *Brassica*-derived vegetable oil supply, comprising 14% of total oilseed production globally (CCC, 2015; Raymer, 2002). Canola oil is used in the production of margarine and cooking oil due to its superior nutritional qualities, which include low saturated fat content and high levels of monounsaturated fat and omega-3 fatty acids. Rapeseed is also a source of oil extraction meal (Friedt and Snowdon, 2009) which is used as animal feed due to its high level of protein. Canola meal is the second largest source of protein meal in the world with 30.8 million metric tonnes produced in the year 2008/2009 (USDA, 2015). World production of rapeseed in 2013 was estimated at almost 62 million metric tonnes (Statista, 2015). By far, the largest exporters of rapeseed oil are China and Canada, with almost half of the world's rapeseed oil originating from these countries (UNFAO, 2015).

Canola is Australia's third largest broadacre crop, grown widely in the southern Australian states (Figure 1.2), and worth millions of dollars to the economy, with oil being the main product. Winter-growing varieties are grown in the wheat belt of Australia, often as a break crop for cereals such as wheat and barley. The estimated revenue from Australian canola is AUD 4.1 billion, considering current canola oil prices (AUD 1,050 per metric tonne crude canola oil) and the Australian volume of production (Indexmundi, 2015). The 2013/2014 growing season saw Australia produce a total of 3,900,000 tonnes of canola from 2,480

hectares (AOF, 2015), of which 3,488,000 tonnes was exported. Australia is the third largest exporter of canola, the successful growth of this crop is therefore important in terms of the Australian economy, maintaining food supply and optimising farming and production.

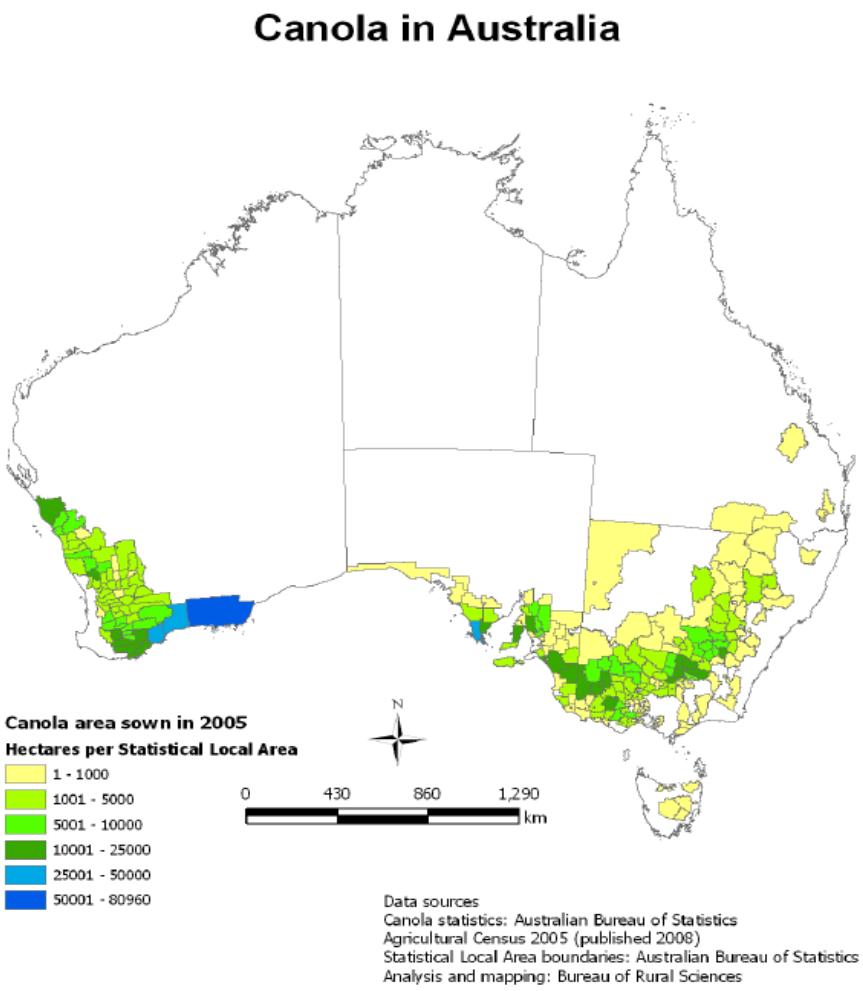

Figure 1.2: Areas in which Canola is grown in Australia (Seberry et al., 2008).

1.2. Blackleg Disease

In Australia, the most economically damaging pathogen of canola cultivars, as well as other *Brassica* species, is *Leptosphaeria maculans* (Howlett et al., 2001). It causes significant losses world-wide, particularly in Europe, Australia and North America (Fitt et al., 2006). *L. maculans*, a filamentous Ascomycete, is the causal agent of phoma stem

canker, commonly referred to as blackleg. It causes lesions on the stem of the plant, leading to poor growth and eventually plant death. In Australia alone, *L. maculans* infection is responsible for an estimated $100 million (AUD) in crop losses each year with losses ranging from 15 – 48 % (AOF 2010).

L. maculans fungal infection, particularly on cabbages (*B. oleracea*), has been on record since 1791 (Hammond and Lewis, 1987; Howlett et al., 2001). The fungus became a concern to canola crops only in the mid-20th century when *B. napus* became commercially important to many parts of the world. Heavy crop losses of up to 90% were recorded in the 1960s to 1970s in Australia (Rouxel and Balesdent, 2005), with major losses occurring by 1972, due to the increasing susceptibility of the crop (Salisbury et al., 1995).

This fungus is found in rapeseed-growing regions all around the world such as Canada and Europe. Increased production of rapeseed in these regions and Australia, has led to a rise in the severity of the disease (Kaur et al., 2009; Marcroft and Bluett, 2008). Furthermore, it has been noted that the variety of blackleg isolates found in Australia are the most virulent (Purwantara et al., 2000), with the susceptibility of the crop and the high availability of Australian inoculums creating conditions conducive to infection (McGee and Petrie, 1979). These Australian isolates can cause disease in naturally more resistant *Brassica* species including *B. juncea*, *B. nigra* and *B. carinata* (Purwantara et al., 2000; Purwantara et al., 1998).

The fungus can be transmitted via air-borne ascospores, entering through leaf stomata and the base of the stem, infecting the plant at both seedling and adult stages. The lifecycle of *L. maculans* includes phases of biotrophy, necrotrophy and saprotrophism (Figure 1.3). *L. maculans* ascospores enter the stomata of plants and form asexual fruiting bodies (pycnidia) in the biotrophic phase. The fungus begins to destroy plant tissue at the stem in the necrotrophic phase. After death or harvesting of the plant, the fungus remains on the decaying plant material and produces sexual fruiting bodies (pseudothecia) in the saprotrophic phase.

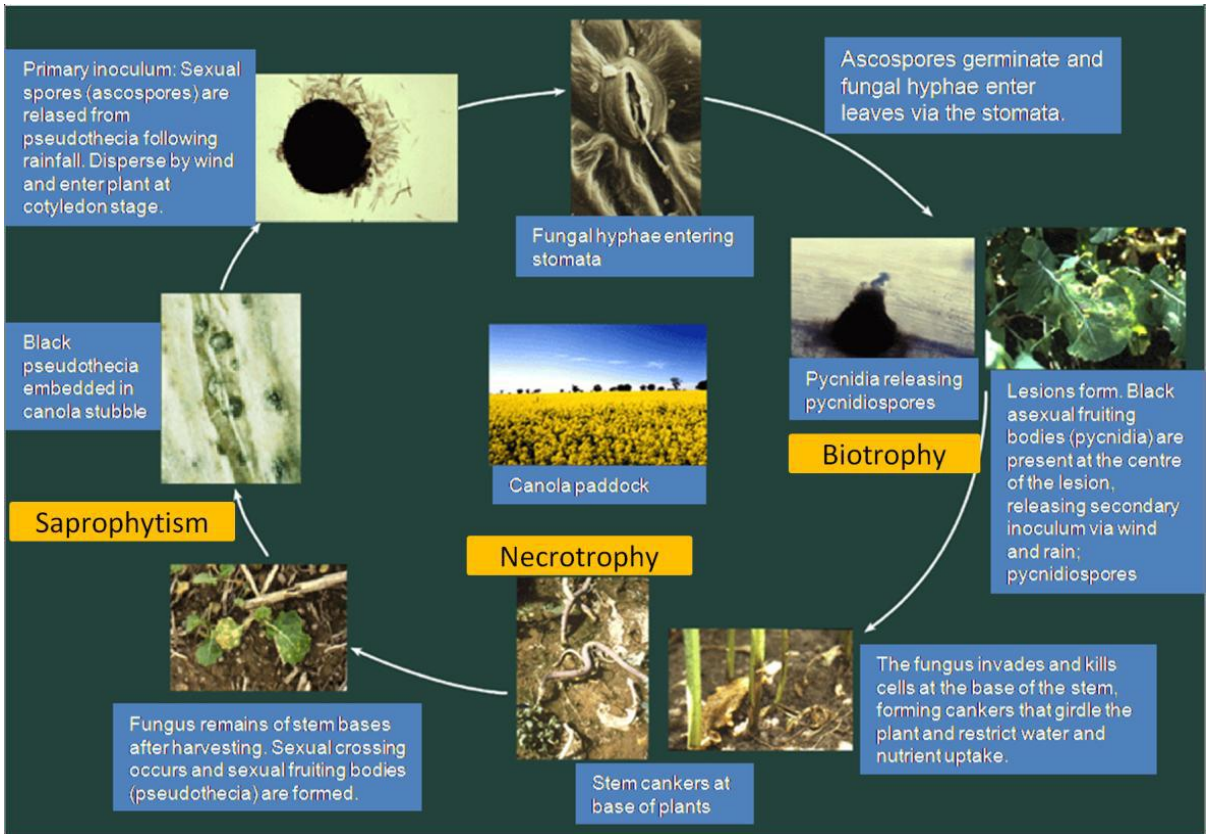

Figure 1.3: The lifecycle of *L. maculans*, adapted from Howlett (2015).

Methods of treatment for blackleg have so far focused on the use of fungicides, burning contaminated stubble, as well as rotating or changing growing fields. These approaches often have little success and negative environmental consequences, such as fungicide runoff into neighbouring ecosystems. They are also time consuming and can leave arable land unusable. Rotation of cultivars containing different resistance gene complements has been shown to be the most effective management method (Fitt et al., 2006).

Understanding the characteristics of the fungal pathogen *L. maculans* is vital in developing an effective and sustainable approach to the treatment of blackleg disease on *Brassica* species. In this regard, work is underway to identify and study resistance genes in *Brassica* species and apply this knowledge in the breeding and selection of resistant lines. This has become the main focus of many canola-breeding programs. Considering the vast threat imposed by this pathogen to the global rapeseed industry, it is imperative to gain more insight into the genetic composition of this fungus using novel approaches. Such information will assist in developing resistant *Brassicas* and eventually reduce crop losses.

1.3. Implementing Disease Resistance

The Australian canola industry has relied heavily on resistant cultivars that contain the *Rlm1* and *RlmS R* genes derived originally from interspecific breeding with *B. rapa* cv. Sylvestris. However, in 2003, just three years after these cultivars were released, this resistance failed and led to 90% crop losses in the Eyre Peninsula, South Australia (Van de Wouw et al., 2010). This indicated that the pathogen exhibited rapid adaptive evolution. Extensive sowing of cultivars with the same gene resistance likely imposed a selection pressure on the fungus to rapidly evolve towards virulence. This high evolutionary potential of *L. maculans* can be attributed not only to mechanisms of rapid genetic change including Repeat Induced Point mutations (RIP), but also to its large population size, sexual recombination in its prolific reproduction cycle and wide spread dispersal of ascospores, creating high gene flow (McDonald and Linde, 2002; Rouxel and Balesdent, 2005; Rouxel et al., 2003).

As mentioned earlier, crop rotation, stubble management by raking or burning and use of fungicides are being used for blackleg management (Hayward et al., 2012b). However, well thought out rotations of resistance genetics are deemed the best methods to combat blackleg infection (Marcroft and Bluett, 2008). It is recommended the cultivars containing a different combination of resistance genes be rotated every four years, in order to minimize selection pressure on the fungus and increase durability of cultivars (Howlett et al., 2001).

1.4. Gene-for-Gene Interaction, Effector-Triggered Immunity (ETI)

1.4.1. Effector-Triggered Immunity

Effector-triggered immunity (ETI), also known as the gene-for-gene hypothesis, is a pathogen race-specific immune response used to explain genetic interactions between host plants and their pathogens. This concept proposes that for each resistance gene in the host there is a specific effector gene corresponding to avirulence in the pathogen (Dangl and Jones, 2001). This was first proposed by Harold Flor in 1942 in the flax-flax rust pathosystem, which has become an important model system for ETI (Flor, 1942). The gene for gene interaction suggests direct or indirect recognition of pathogen avirulence (*Avr)*-encoded effectors by the protein product encoded by the corresponding resistance gene (*R* gene) in the host plant. In response to these effectors, resistance proteins in plants are triggered to activate defence mechanisms such as the hypersensitive response

(Heath, 2000). Often *R* genes encode nucleotide binding site - Leucine rich repeat (NBS-LRR) proteins leading to the induction of a signalling cascade and subsequent downstream defence responses (Bent, 1996). However, recent studies have concluded that the interaction between *R* and *Avr* genes is very complex and may involve multiple effectors to trigger resistance in the plant (Gassmann and Bhattacharjee, 2012).

1.4.2. Pathogen Effector Genes

In 1994 Oku defined (1) entering a plant, (2) overcoming host resistance and (3) initiating disease as the three abilities for microorganisms to parasitically adapt into plant pathogens (Ichinose et al., 2013; Oku, 1994). His observations were primarily based on infections caused by fungal pathogens, one of the many causal agents of diseases in plants. Understanding the role of secreted proteins or effectors that target plant defence responses is essential to gaining knowledge of how pathogens infect and colonise the host.

Effectors, or avirulence (*Avr*) genes, are phytopathogenic genes encoding proteins that are specifically recognised by host plants containing the 'matching' resistance (*R*) gene (Rouxel and Balesdent, 2005). The products of these genes are known as effector proteins, and these carry out functions such as plant defence suppression and host cell-wall modifications upon infection (Rouxel et al., 2011). Identifying novel effectors is a challenge due to the large portion that have novel sequences with no similarity to other effectors, other proteins in known databases and no relatedness to other species in the same genus or same species in different host ranges (Ellis et al., 2009). Many effectors possess a dual function of avirulence and virulence (White et al., 2000). Several effectors are recognised by host resistance (R) proteins, triggering the hypersensitive reaction (HR) in host cells during expression or secretion in host cells. This is their avirulence function caused due to the pathogen being avirulent on hosts that express the corresponding R protein (Ellis et al., 2009).

Plant pathogens have evolved to employ a variety of mechanisms that deliver effectors directly into host cells. Bacteria have been shown to utilise special secretory systems, like TS33, and fungi and oomycetes use structures known as haustoria (Hogenhout et al., 2009). Parasitic nematodes use stylets or special feeding tubes to deliver effectors into the plant vascular cell. There are selective fungal effector proteins, such as ToxA in *Pyrenophora tritici-repentis,* that translocate independently within the host cell by using a

plant-surface receptor binding to specialised motifs (Hogenhout et al., 2009). Effectors are also known to affect plant apoplastic defences. Fungi such as *Cladosporium fulvum*, which is an extracellular parasite of tomato, grow only in the apoplast and do not form haustoria. Apoplastic effectors are commonly known to protect the pathogen against plant hydrolytic enzymes such as glucanases, chitnases and proteases. *Phytophthora infestans* secretes protease inhibitors like EPIC2B and EPI1 and EPI10, which are cysteine protease and serine protease inhibitors respectively, protecting the pathogen against plant defence activity. *P. infestans* is also known to secrete host translocated or cytoplasmic effectors (Hogenhout et al., 2009).

1.5. Examples of Plant-Pathogen systems:

1.5.1. Flax and Flax Rust

The interaction between flax (*Linum usitatissimum, L. marginale* and other *Linum* species*)* and the pathogen flax rust (*Melampsora lini*) provided the earliest evidence of the genetic principle for plant disease resistance, being defined in 1956 by Flor et al. and remains one of the most important models for plant-pathogen interactions. It has been used to understand the genetic and molecular characteristics of host-pathogen interactions and how they develop in natural disease systems (Ravensdale et al., 2011). As an obligate biotroph, *M. lini* depends on living plant tissue for survival and extracts nutrients through specialised feeding structures known as haustoria.

Flax resistance (*R*) genes have been mapped to five distinct loci across the genome, referred to as the *K, L, M, N*, and *P* groups (Islam and Mayo, 1990; Lawrence et al., 2010). Within these loci, 30 genes conferring resistance to flax rust have been mapped and subsequently shown to be closely linked genes or allelic variants of the same gene. Of the nineteen *R* genes cloned, 11 are from group *L*, three from group *N*, two from group *P* and three from group *M*. The loci are complex, with each consisting of varying avirulence gene specificities. For example, a single gene at the *L* locus possesses 13 allelic variants, reacting to rust strains with different *Avr* genes. The *N* locus contains four, the *M* locus up to 15, and the *P* locus six to eight *R* gene paralogues, arranged in tandem at these locations, and more complex than the other loci (Ravensdale et al., 2011).

So far all the *R* genes in flax have been shown to encode TIR-NBS-LRR class proteins, with much of the variation between these occurring in the LRR domain (Ravensdale et al.,

2011), a horseshoe-shaped domain which seems to be critical in the recognition specificity between *R* and *Avr* proteins (Ellis et al., 1999). Polymorphisms of just 6 amino acids in the exposed residues within the LRR domains of *P* and *P2*, give these two *R* genes different avirulence gene specificities (Dodds et al., 2001). The role of the LRR domain in pathogen effector recognition has been observed, however evidence also suggests regions outside the LRR, such as the TIR domain, may further influence effector specificity (Luck et al., 2000), which has also been shown in the *N* resistance gene of the tobacco-Tobacco mosaic virus (TMV) pathosystem (Burch-Smith et al., 2007).

The effector genes of flax rust have been classified into four *Avr* families which contribute to approximately 30 different *Avr* specificities. All the identified *Avr* genes and their variants encode small secreted proteins, and the majority were identified by screening expressed rust genes during infection and colonisation of flax (Catanzariti et al., 2006; Dodds et al., 2004). These small secreted proteins appear to be translocated into plant cells and indicate that *Avr* protein recognition occurs inside the plant, as flax R proteins are cytoplasmic (Catanzariti et al., 2006; Dodds et al., 2004). Immunolocalisation studies have identified the presence of Avr proteins inside host cells, as well as evidence to suggest that the pathogen exploits plant-derived transport mechanisms to enter the cells (Rafiqi et al., 2010).

Flax rust effectors exhibit a high level of sequence polymorphism that is associated with a difference in recognition specificity (Dodds and Thrall, 2009). The four *Avr* families show no significant sequence similarity with each other as well as to other known proteins, indicating that they are unique and specific to this pathosystem (Catanzariti et al., 2006).

1.5.2. Rice and Rice Blast

Rice blast (*Magnaporthe grisea*) pathosystem is one of the most significant diseases of cereal crops in the world and has proven to be a useful model for ETI in both monocots and plants in general. *M. grisea* is an ascomycete fungus that infects both aerial and root tissue in cereals, primarily rice species (*Oryza* species). The genomes of the *O. sativa* ssp. *japonica* and *indica* contain approximately 80 rice blast *R* genes (Zhai et al., 2011). Of these, 13 have been isolated and characterised to date (Liu et al., 2010) and their interaction with rice blast *Avr* genes is still being elucidated, with direct R-AVR protein interaction so far studied in Pita:AVR-Pita (Jia et al., 2000) and Pizt:AVR-Pizt (Li et al., 2009b). Improvements have been achieved through the availability of whole genome

sequence for both host and pathogen species and will continue to improve the knowledge of this patho-system. Host resistance in rice contains both R gene mediated resistance, conferring complete resistance, and quantitative resistance which provide partial resistance in the field.

The R genes of rice have been shown to exist in clusters within the genome, such as those on chromosomes 6, 11 and 12 (Liu et al., 2010). Of the 13 cloned rice blast R genes, all except one (*Pi-d2*) encode NBS-LRR proteins, *Pi-d2* encoding a receptor kinase (Chen et al., 2006). As in other plant species, the LRR domain of rice R genes have been shown to form a direct interaction with the pathogen effector proteins, which has been identified between the rice *Pi-ta R* gene and the rice blast *Avr-Pita* (Jia et al., 2000).

The genome of *Magnaporthe* species has been shown to encode many secreted proteins which could be involved in the suppression of host defence responses. Indeed the The *AvrPiz-t* effector of *M. oryzae* induces suppression of PAMP-triggered immunity in rice, targeting the ubiquitin ligase APIP6 which in turn suppresses flg22 and chitin induction of reactive oxygen species (ROS) (Park et al., 2012). The *Avr* genes, which induce ETI in rice, vary greatly in structure and function. For example, the *ACE1* gene, avirulent towards the rice *Pi33 R* gene, encodes a hybrid polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS) and has been postulated to drive the biosynthesis of a secondary metabolite that in turn is recognised by the Pi22 protein (Bohnert et al., 2004).

1.5.3. Tomato and Fusarium Wilt

Fusarium wilt disease of Tomato (*Solanum lycopersicum*) is caused by the fungus *Fusarium oxysporum f. sp. Lycopersici (Fol).* Resistance genes in Tomato against *Fol* are known as *Immunity (I)* genes. So far the *I, I-1, I-2* and *I-3* genes have been identified. These genes were originally sourced from wild species and were successfully introgressed to create resistant tomato cultivars (Mes et al., 1999; Takken and Rep, 2010). Resistance genes *I* and *I-2* are both effective against the fungal avirulence gene *Avr1/Six4. I-3* is thought to interact with the *Fol Avr3/*SIX1 protein (Rep et al., 2005). To date, only one Tomato-Fusarium R gene, *I-2*, has been successfully cloned (Simons et al., 1998) and tomato lines transformed with *Fol Avr2* exhibit endogenous *I-2* expression in the leaves (Houterman et al., 2009; Takken and Rep, 2010). Unfortunately, proving a direct interaction between *Avr2* and *I-2* proteins through yeast two-hybrid methods has been unsuccessful, indicating that indirect signalling mechanism may be responsible for the

interaction (Takken and Rep, 2010). Similar to other pathosystems, the *I-2* gene encodes an NBS-LRR protein. Mutation of the nucleotide binding domain suggests that the active form of the protein requires the specific binding of a nucleotide for pathogen recognition (Takken and Rep, 2010). Unlike other R gene mediated defence responses where hypersensitive response (HR) is involved, *I* gene resistance utilises other mechanisms, including callose deposition, phenolic production, and deposition of other protective barrier compounds (Michielse and Rep, 2009; Takken and Rep, 2010).

During pathogenesis Fol secretes a multitude of proteins into the host, including effector proteins. Some of the *Fol* avirulence genes belong to a family which encode 'Secreted in Xylem' (SIX) proteins. There are 11 of these proposed, however, not all are thought to be true effectors based on the lack of interaction with host R genes (Gawehns et al., 2014; Takken and Rep, 2010). Of these, *SIX1/Avr3*, the *Avr* gene recognised by *R* gene *I-3*, is present in all *Fol* tested isolates and appears to be unique to this forma specialis, suggesting a dual, tomato-fusarium specific role (Rep, 2005; Rep et al., 2005). Three confirmed effectors interacting with host *R* genes are *Avr1/Six4, Avr2/Six3* and *Avr3/Six1* (Houterman et al., 2009; Houterman et al., 2007; Lievens et al., 2009; Rep et al., 2005). Recently, homologs of SIX1, SIX4, SIX8 and SIX9 were identified in an Arabidopsis infecting strain of Fol (Thatcher et al., 2012). Most SIX proteins appear to be unique to *Fusarium oxysporum*. However, recently, homologs of SIX6 were identified in several different microbes (Gawehns et al., 2014). A SIX6 knockout displayed inhibited virulence, with complementation restoring virulence, in two independent transformants suggesting an effector role for this protein (Gawehns et al., 2014). Furthermore, expression analysis show that SIX6 is not expressed on synthetic media but requires a live-host cell for expression as has been previously shown for *SIX1/Avr3* (Gawehns et al., 2014; van der Does et al., 2008a).

Although members of *Fusarium oxysporum f. sp. Lycopersici* have a polyphyletic origin all pathogenic strains require an 8 kb region for pathogenicity and specificity of this forma specialis (van der Does et al., 2008b). Within *Fol* mobile pathogenicity chromosomes have been described; when chromosome 14 was transferred into a non-pathogenic *Fol* this resulted in pathogenicity (Ma et al., 2010).

Due to the high reproduction rates and abundance of *Fusarium* in the soil there is a constant race between host and pathogen to overcome ETI. It was proposed that the

breakdown of single gene resistance is not dependant on loss or mutation of avirulence genes rather dependant on gene suppression (Rep et al., 2005). However, further studies have revealed this is through *Avr1* triggered suppression of *I-2* and *I-3* (Houterman et al., 2008; Lapin and Van den Ackerveken, 2013). It is proposed that Avr1 was acquired by Fol and suppression mechanism exploited in order to maintain other Avr content (Ma et al., 2010; Thatcher et al., 2012).

Futhermore, plant resistance utilising specific aspects of PAMP-triggered immunity (PTI) has been shown. For example two reports, Abdallah et al. (2010) and Khan et al (2011) successfully created *Fusarium oxysporum f. sp. Lycopersici* resistant GM lines through over expressing defensin genes, small cysteine rich proteins (Stotz et al., 2009). Other non-ETI related research targets in other microbial species, such as non-pathogenic *Fusarium* strains and *Bacillus spp.* have been used as biological controls against *Fol* (Ajilogba and Babalola, 2013). Despite these findings, *R* genes are still considered the most promising mechanism of resistance in plants.

1.6. Gene-for-Gene Interaction in the Brassica-Blackleg Pathosystem

It is currently known that *L. maculans* develops a gene-for-gene relationship with its *Brassica* plant host (Ansan-Melayah et al., 1998). Specific resistance genes in *Brassica* species, termed *Rlm* (*Resistance to Leptosphaeria maculans*) genes, recognise specific avirulence gene products in *L. maculans* (*AvrLm*). This means that if corresponding *AvrLm* and *Rlm* genes are present (e.g. *AvrLm1* and *Rlm1*), mechanisms for disease resistance are activated in the plant (Dangl and Jones, 2001) and the pathogen cannot cause symptoms. Therefore the outcome of infection depends on the presence of an avirulence (*AvrLm*) gene in the fungus and a resistance (*R*) gene in the plant.

1.6.1. Brassica Resistance genes

The plant-pathogen interaction between *Brassica* species and Blackleg disease, caused by *Leptosphaeria maculans*, has been shown to involve both qualitative and quantitative resistance (Balesdent et al., 2001; Balesdent et al., 2002). Qualitative resistance is race-specific and depends on the presence of a single resistance (*R*) gene in the plant corresponding to an avirulence (*Avr*) gene in the pathogen (Ansan-Melayah et al., 1998). Qualitative resistance is expressed from seedlings to adults in cotyledons and leaves (Delourme et al., 2006). Quantitative resistance is generally thought to be race non-

specific, mediated by many genes and expressed at the adult plant stage, conferring only partial resistance to the pathogen (Delourme et al., 2006; Rimmer, 2006).

The first genetic evidence for the gene-for-gene interaction between *Brassica* species and *L. maculans* was identified between the *BnRlm1* (*Resistance to Leptosphaeria maculans 1*) resistance gene in *B. napus* cv. Quinta and the corresponding avirulence gene *AvrLm1* (Ansan-Melayah et al., 1998). Fifteen specific resistance genes have been subsequently identified in the cultivated *Brassica* species (table 1.1) and are often clustered together on the genome (Delourme et al., 2006; Delourme et al., 2004; Rouxel et al., 2003).

Recently both the whole genome sequence of *L. maculans* (Rouxel et al., 2011) and several *Brassica* species, including *B. napus* (Chalhoub et al., 2014) and *B. rapa* (Wang et al., 2011) have been published. These developments will no doubt aid in understanding the disease interaction further from both the plant and pathogen research, as both *R* and *Avr* genes and other disease-related genes will be easier to discover and physically map to the genome.

Table 1.1: Relationship between *Brassica* resistance genes and *Leptosphareria maculans* avirulence genes in gene-for-gene interactions. (Adapted from Hayward et al. (2012b)).

Species	*L. maculans* Avirulence gene	Resistance gene	Location on chromosome	Reference
Brassica napus	-	*$LmFr_1$*	-	(Dion et al., 1995; Yu et al., 2005)
	-	*aRLMrb*	A7	(Zhu and Rimmer, 2003)
	-	*cRLMj*	A7	(Rimmer, 2006)
	-	*aRLMj*	A7	(Zhu and Rimmer, 2003)
	alm1	*LEM1=Rlm4 or 1?*	A7	(Ferreira et al., 1995; Mayerhofer et al., 1997; Pongam et al., 1998; Yu et al., 2005)
	AvrLm1	*Rlm1*	A7	(Ansan-Melayah et al., 1998; Balesdent et al., 2002; Brun et al., 2010)
	AvrLm2	*Rlm2*	A10	(Ansan-Melayah et al., 1998; Balesdent et al., 2002; Delourme et al., 2004; Ghanbarnia et al., 2014)
	AvrLm3	*Rlm3*	A7	(Balesdent et al., 2002; Delourme et al., 2004)
	AvrLm4	*Rlm4*	A7	(Balesdent et al., 2001; Brun et al., 2010; Delourme et al., 2004)
		cRLMm=Rlm4?	A7	(Rimmer, 2006)

		LmR1=Rlm4?	A7	(Mayerhofer et al., 1997)
		cRLMm=Rlm4?	A7	(Mayerhofer et al., 1997)
		cRLMrb=Rlm4?	A7	(Rimmer, 2006)
	AvrLm7	*Rlm7*	A7	(Balesdent et al., 2002; Delourme et al., 2004)
	AvrLm9	*Rlm9*	A7	(Balesdent et al., 2002; Delourme et al., 2004)
Brassica rapa	-	*Rlm8*	-	(Balesdent et al., 2002)
	AvrLm11	*Rlm11*	-	(Balesdent et al., 2013)
Brassica rapa sylvestris	-	*LepR1*	A2	(Yu et al., 2005)
Brassica rapa sylvestris	-	*LepR2*	A10	(Yu et al., 2005)
	-	*LepR3*	A10	(Rimmer, 2006)
	-	*LepR4*	-	(Yu et al., 2013)
Brassica juncea	-	*Rlm6=Jlm1*	-	(Balesdent et al., 2002; Brun et al., 2010; Chevre et al., 1996)
	AvrLm6	*Rlm6*		(Fudal et al., 2007)
	AvrLmJ1	*Rlm5? Rlm6?*	-	(Van de Wouw et al., 2014)
	-	*Rlm5*	-	(Balesdent et al., 2002)
Brassica nigra	-	*Rlm10*	-	(Delourme et al., 2006)

The first Blackleg resistance gene in a *Brassica* species, *LepR3*, was cloned in *B. napus* lines with introgressions from *B. rapa* ssp. *sylvestris* (Larkan et al., 2013). A second of the genetically mapped *Rlm* genes has also been sequenced and validated in *B. napus,* namely *Rlm2* (Larkan et al., 2015). This is an allelic variant of *LepR3*. New advances in Brassica genome sequencing and development of molecular markers have provided a basis for the identification of candidate genes. By performing sequence comparison of the blackleg *R* genes, and combining this with infection assays, using a range of *L. maculans* isolates that carry single (or very few) avirulence genes, the number and uniqueness of blackleg *R* genes can be confirmed. Recently two candidates for *Rlm4* were identified (Tollenaere et al., 2012) in *B. napus* using a molecular marker approach combined with phenotypic data.

1.6.2. Blackleg Avirulence genes

Previous mapping studies identified nine *AvrLm* and their corresponding *Rlm* genes. Of these, six *Avr* genes have been characterised, namely *AvrLm1* (Gout et al., 2006)*, AvrLm2* (Ghanbarnia et al., 2014)*, AvrLm6* (Fudal et al., 2007), *AvrLm4-7* (Parlange et al., 2009), *AvrLm11* (Balesdent et al., 2013) and *AvrLmJ1* (Van de Wouw et al., 2014). The products of these avirulence genes are involved in fungal virulence and are recognized by the products of the corresponding *R* loci in the plant; *Rlm1* (recognizes *AvrLm1*), *Rlm2* (recognizes *AvrLm2*), *Rlm6* (recognizes *AvrLm6*), and *Rlm4* and *Rlm7* (both recognize *AvrLm4-7*). There have been records of deletion of the *AvrLm1* locus in several *L. maculans* isolates, purported to be one mechanism by which the fungus adapts to *Rlm1* resistance (Fudal et al., 2009). The same has been noted for *AvrLm4-7* (Van de Wouw et al., 2014).

A further two *AvrLm* genes have been identified in *L. maculans* that have not previously been mapped with the initial nine *AvrLm* genes. These are *AvrLm11* (Balesdent et al., 2013) and *AvrLmJ1* (Van de Wouw et al., 2014). *AvrLm11* was identified on a minichromosome, which is frequently lost following meiosis (Balesdent et al., 2013). This gene interacts with a resistance gene termed *Rlm11* from *B. rapa,* however *Rlm11* is yet to be genetically mapped. *AvrLmJ1* is believed to be species specific in that it confers avirulence in *B. juncea* varieties (Van de Wouw et al., 2014).

All of the six characterised genes display features typical of effector genes in that they are expressed early in infection and secrete small proteins into the apoplast (Stergiopoulos

and de Wit, 2009). Rouxel et al. (2011) confirmed and physically mapped *AvrLm1, AvrLm4-7* and *AvrLm6* to the *L. maculans* genome. Previously, Fudal et al. (2009) used map-based cloning to clone *AvrLm1* and *AvrLm6.*

Data from Rouxel et al. (2011) revealed that these *AvrLm* genes are located in gene poor, AT-rich blocks of the fungal genome, referred to as 'lost in the middle of nowhere' (Balesdent et al., 2013, Gout et al., 2006). These AT-rich blocks, abundant in transposable elements (TEs), are affected by repeat induced point (RIP) mutations which create increased genetic diversity (Rouxel et al., 2011). By definition, a RIP mutation is a fungal-specific premeiotic repeat inactivation mechanism which causes nucleotide substitutions from C-to-T and G-to-A (Fudal et al., 2009; Rouxel et al., 2011). This RIP mechanism is postulated to intermittently cause SNP mutations in *Avr* genes, as these are often located in regions of RIP activity. These polymorphisms lead to altered *Avr* gene sequences that can no longer be recognised by *R* genes, thus providing the variation upon which selection can occur.

Indeed, RIP mutations at the *AvrLm6* locus of isolates containing this gene appear to be a mechanism of rapid evolution in response to resistance-gene selection pressure and are associated with virulence (Fudal et al., 2009). *AvrLm1* and *AvrLm6* are both positioned within a repeat-induced point mutation (RIP) affected *Avr1-2-6* region of the genome that contains two kinds of isochores. Isochores can be defined as long stretches of DNA homogenous in GC content, with abrupt changes in GC content from one segment to another (Parlange et al., 2009).

The identification of *AvrLmJ1* (Van de Wouw et al., 2014) further confirmed the genomic location of avirulence genes in *L. maculans*, showing that this gene is situated in a gene-poor, AT-rich region of the genome, with the closest genes being 57 and 33 kb away.

The interaction of *Brassica* resistance and blackleg avirulence gene proteins at the molecular level has not been studied so far. Determining the crystal protein structure of the resistance and avirulence genes will help in identifying novel *R* and *Avr* genes.

Recently, the crystal structure of *AvrLm4-7* was determined (Blondeau et al., 2015). The AvrLm4-7 protein, produced in *Pichia pastoris*, was found to be structurally unrelated to existing protein models. Loss of recognition by *Rlm4* is due to a single glycine to arginine substitution in the loop of the protein. More drastic modifications of the protein structure lead to a loss of recognition by the *Rlm7* gene from the plant host, including numerous

point mutations and gene loss, which lead to truncation of the protein (Blondeau et al., 2015; Daverdin et al., 2012). The crystal structures of the other known avirulence and resistance genes will likely follow. This may aid in identifying novel avirulence genes, as three dimensional fold can provide a new mechanism to search for these genes, as there is often very limited sequence similarity between them (Blondeau et al., 2015).

1.7. *L. maculans* Genetics and Genomics

Due to its rapid evolutionary capacity, several genetically variable isolates of *L. maculans* are usually present in the one environment (Grandaubert et al., 2014). In the past, genetic diversity studies of this pathogen have been conducted using amplified fragment length polymorphisms (AFLPs) as genetic markers (Barrins et al., 2004). Based on AFLP banding patterns, the genetic diversity of Australian isolates collected in the late 1980s was the same as that seen in Europe, but greater than that in Northern America. The study noted that a general uniformity in genetic diversity existed across isolates collected from different regions (East and the west coast of Australia). High genetic variation was found within the same population of isolates in the same region. However, later studies conducted using microsatellite and minisatellite markers showed genetic differences in isolates from Eastern and Western Australia, which were attributed to the presence of intervening arid desert acting as a natural barrier to gene flow (Hayden et al., 2007). The overall method of genetic spread of *L. maculans* in the environment is yet to be elucidated, as the latest studies have shown contrasting population structures in different regions (Dilmaghani et al., 2012, Travadon et al., 2011). A recent diversity analysis by Patel et al. (2015) indicates that the *L. maculans* population in Australia is panmictic, with high rate of sexual reproduction and evolutionary diversification within fungal populations.

1.7.1. Whole genome sequencing

With advances in genome sequencing technologies providing up to 600 Gbp sequence data per Illumina HiSeq run (Liu et al., 2012), SNP discovery from next generation sequencing data is now a convenient option. Re-sequencing is used to identify genetic variation between individuals, which can provide molecular genetic markers and insights into gene function (Imelfort et al., 2009). The process of whole genome re-sequencing using short read technologies involves aligning a set of literally millions of reads to a

reference genome sequence. Once this has been achieved, it is possible to determine variation in nucleotide sequence between the sample and the reference.

A substantial development in the study of this fungal pathogen was the completion of the *L. maculans* whole genome sequence in 2011 (Rouxel et al., 2011). The total genome assembly size is 45.12 Mb and is scaffolded into 76 supercontigs. Gene prediction analyses revealed 12,469 genes of which nearly 85% were shown to be expressed (Rouxel et al., 2011). The *L. maculans* 'brassicae' isolate v23.1.3 was used for the genome sequence as it harbours a number of avirulence genes which have been previously cloned. This effort will provide a vital resource in the future of *L. maculans* research by providing a reference genome to which markers, including SNPs, can be mapped to. In other plant pathogens to have their genomes sequenced, such as the wheat pathogen *Parastagonospora nodorum* (Hane et al., 2007) and the cereal pathogen *Fusarium graminearum* (Cuomo et al., 2007) the availability of a reference genome has been a major step forward. Whole genome re-sequencing will also be a cost-effective method to study diversity in this fungal genome, considering the relatively small size of the genome and given that newly generated sequence tags can be easily assembled using the reference genome. The availability of technologies such as RNA-seq has allowed for the study of transcriptomes and their role in disease initiation and progression, such as the Lowe et al. (2014) study, which identified the differences in gene expression upon infection of two *Leptosphaeria* species, namely *L. maculans* and *L. biglobosa*.

1.7.2. Single Nucleotide Polymorphisms (SNPs)

Single nucleotide polymorphisms (SNPs) are single nucleotide differences between the DNA sequences of individuals in a population and are currently one of the most popular markers for the fine mapping of heritable traits (Chagné et al., 2007). There are three categories of SNPs: transversions (C/G, A/T, C/A and T/G), transitions (C/T or G/A) and insertions/deletions, also known as indels. Most SNPs at any given site are bi-allelic, although tri-allelic and tetra-allelic SNPs also exist.

SNPs within fungal genomes are currently a novel and increasingly popular approach to study genomic evolution and genetic diversity (Appleby et al., 2009; Lai et al., 2015). The high heritability of SNPs makes them an excellent indicator of genetic diversity and phylogeny in pathogenic fungal species, such as *L. maculans*. This allows for the construction of high-density genetic maps which can provide a scaffold to map important

genetic traits (Duran et al., 2010a; Edwards and Batley, 2004; Edwards et al., 2007a). Furthermore, sequence variation can have a major impact on how the organism develops and responds to the environment. Several features of SNPs, such as stability during evolution, low mutation rate and their high density in the genome, make them the tool of choice for studying genetic diversity and phylogeny (Appleby et al., 2009). Current large-scale SNP genotyping assays, such as the Illumina GoldenGate assay, are cost-effective and flexible for analysing large numbers of SNPs across multiple individuals (Appleby et al., 2009). Comparative studies using both SNPs and microsatellites deemed the former a less error prone marker for linkage mapping in zebra finch (Ball et al., 2010). Other studies have concluded that when a large number of SNPs are available, they can be highly informative (Liu et al., 2005). To date, no previous studies have implemented the use of SNPs in *L. maculans*.

1.7.4. Presence-Absence Variations

Presence-absence variations (PAVs) are genomic regions of varying length that are present in one individual and absent in another, based on comparison of the DNA sequences. A number of studies have identified PAVs as a source of phenotypic variation, potentially contributing to the adaptability of an organism to different selective pressures (Feuk et al., 2006). PAVs have been identified in plant species, including rice (Kumar et al., 2007), maize (Belo et al., 2010), melon (Gonzalez et al., 2013), soybean (Lam et al., 2010) and *Arabidopsis thaliana* (Bush et al., 2014; Tan et al., 2012), as well as in bacteria (Araki et al., 2006; Francis and Tanaka, 2012). In *A. thaliana* as well as in *Cucumis melo* (melon) and other melon species, it was found that a large number of present/absent genes were involved in stress response and a significant portion of these were disease resistance genes (Gonzalez et al., 2013; Shen et al., 2006). In soybean, structural variants such as PAVs have been linked to nucleotide binding and biotic defense responses (McHale et al., 2012). PAVs exist in fungal species also, with PAVs being identified in the rice blast fungus *Magnaporthe oryzae* (Huang et al., 2014) as the potential mechanism of host immune response evasion upon infection. PAVs have also been seen in *L. maculans*, such as deletion of the *AvrLm1* and *AvrLm6* loci in *L. maculans* leading to virulence (Raman et al., 2012; Van de Wouw et al., 2010).

1.8. Aims and Significance of Research

The aim of this thesis is to increase the understanding of the structure and variation of the *L. maculans* genome. The advent of next generation sequencing technology has made genome re-sequencing a viable approach to comparing genomes of fungal species, such as *L. maculans*. For this reason, this technology has been implemented in the analyses of this thesis, as it provides a large amount of genome information for this purpose. The specific objectives for this thesis are as follows:

Chapter 3: To generate sequence data and analyse this to predict SNPs between two important *L. maculans* isolates used in the mapping of the *Rlm4* resistance gene in *B. napus*. These SNP markers are useful tools in describing and determining genetic diversity between these isolates, as selection pressures can be evaluated across the genome by analysing SNP numbers across SuperContigs for each dataset. Genes of interest can be studied to determine whether they are affected by SNP mutations in different isolates, and whether this alters their phenotype.

Chapter 4: To complete a larger SNP prediction using a 10 isolate set, with known *AvrLm* gene content. The analysis of the SNP affects will be performed, with emphasis on known *AvrLm* genes, as well as putative pathogenicity related genes.

Chapter 5: To implement a novel method of determining larger scale genome variations in *L. maculans*. Presence/absence variations (PAVs) are a relatively new area of study, which can provide information on larger structural variations, which can be in excess of 100 bp. The size difference to SNPs allows for the study of genes that are being lost or are highly diverged in a range of individuals.

Using these new molecular tools, a better understanding of disease processes in the Brassica-Blackleg pathosystem can be achieved. Understanding the inheritance and selective pressures working on pathogenicity genes can improve our understanding of blackleg disease. Therefore the main focus of this thesis is on the analysis of disease related genes, and the mechanisms that affect their recognition by the host. This knowledge can then be further utilised to aid in the testing and breeding of resistant plant lines in future. This will ultimately lead to improved crop protection results, with blackleg resistant plants providing less crop losses and improved yield.

2. Materials and Methods

2.1. Buffers/media

2.1.1. Buffers for DNA Extraction

The buffers for DNA extraction were made according to the protocol from Fulton et al. (1995), also known as the 'microprep protocol'. The final microprep buffer comprises several buffers which need to be combined to form the final mixture. The buffers are as follows.

The DNA extraction buffer consists of 0.35 M sorbitol, 0.1 M tris-base and 5 mM EDTA, adjusted to pH 7.5. The nuclei lysis buffer is comprised of 0.2 M tris, 0.05 M EDTA, 2 M NaCl and 2% CTAB. The sarkosyl solution has the concentration of 5% w/v.

The final Microprep buffer consists of 2.5 parts DNA extraction buffer, 2.5 parts nuclei lysis buffer, 1.0 part 5% Sarkosyl and 0.3 to 0.5 g sodium bisulfite per 100 mL buffer.

The individual buffers are stored at room temperature and are replaced every 6 months. The microprep buffer is only used on the day it is made, any remaining buffer is discarded.

2.1.2. V8 Juice Media

For the growth of *Leptosphaeria maculans*, V8 juice media was used (V8® Beverages, Campbell Australia Pty Ltd). The ingredients contained in V8 Vegetable juice (Original) are: Reconstituted vegetable juice blend (water and concentrated juices of tomatoes, carrots, celery, beets, parsley, lettuce, watercress, spinach), contains less than 2% of: salt, vitamin c (ascorbic acid), natural flavoring, citric acid. Prior to use in making media, the V8 juice is filtered through miracloth to remove the solid components of the juice. The filtered V8 juice is subsequently stored at -20°C in Schott bottles or in 40 mL aliquots in 50 mL falcon tubes.

2.1.2.1. Solid Media:

The most common culturing method of *L. maculans* is on solid V8 juice agar plates.

The recipe for 10% V8 juice agar is as follows (for 20% concentration, add 2X 40 mL V8 juice). The required amounts are 40 mL Campbell's V8 Juice (Original), 1.5 g Calcium Carbonate ($CaCO_3$) and 8 g Agar.

Notes:

These ingredients are placed in a 1 L Schott bottle and made up to 400 mL by adding laboratory-grade water (Milli-Q water (Millipore Corporation)). The pH is adjusted to 6.5 (using HCl) and the media is then placed in the autoclave for 20 minutes at 120°C. The V8 juice agar mixture is allowed to cool to 40-50°C before adding the antibiotics Rifampicin (10 mg/mL MeOH) and Chloramphenicol (30 mg/mL EtOH). The media is then poured into 90 X 14 mm petri dishes and left inside a laminar flow cabinet to dry with the lids left slightly open. One batch (400 mL) of V8 Juice agar usually generates around 20 petri dishes of media.

2.1.2.2. Liquid Media:

In order to grow fungal tissue for DNA extraction, liquid media is used to avoid agar contamination in downstream applications. The preparation of 10% V8 juice liquid media (for 20% concentration, add 2X 40 mL V8 juice) is described below.

Forty mL of V8 Juice (original) is added to a 1 L Schott bottle and subsequently filled to 400 mL with laboratory grade water (Milli-Q water). The solution is autoclaved for 20 minutes at 120°C. The V8 juice is allowed to cool to 40-50°C before adding the antibiotics Rifampicin (10 mg/mL MeOH) and Chloramphenicol (30 mg/mL EtOH). Roughly 50 mL of media is poured into 145 X 20 mm tissue culture dishes. One batch (400 mL) of V8 Juice liquid media usually generates around 8 tissue culture dishes of media.

2.1.3. TAE Buffer

TAE buffer is used for electrophoresis of the agarose gel and in making the agarose gel (see section 2.2). A 50X stock solution is made and diluted to a 1X working solution.

To make 1 litre of 50 X TAE Buffer solution, 242 g Tris Base (dissolved in water), 57.1 mL Glacial acetic acid and 100 mL 0.5 M EDTA (pH 8.0) are combined and the solution topped up with laboratory grade water (Milli-Q water) up to 1 litre.

The diluted 50X TAE buffer (50:1) results in a 1X TAE working solution which contains 40 mM Tris, 20 mM acetic acid and 1 mM EDTA.

2.2. Agarose Gel Electrophoresis

Agarose gel is used to separate nucleic acids, in this thesis primarily DNA. In this thesis it was performed after PCR in order to determine the size of the PCR product by comparing it to a DNA ladder with fragments of known length. It was also used to estimate the concentration of DNA after DNA extraction, by comparing the intensity of the sample bands to bands of a DNA ladder with known concentrations.

2.2.1. Agarose Gel Preparation

In this thesis, agarose gels were prepared using 1X TAE buffer, 1% Agarose and ethidium bromide at 0.5 µg/mL.

Notes:

The 1X TAE buffer and 1% Agarose are combined in a conical flask and placed in the microwave until the agarose is fully dissolved. The solution is allowed to cool to ~60°C and then the ethidium bromide is added. The solution is mixed by swirling the contents in the conical flask. The final solution is poured into a gel mould and a comb is placed inside to form the wells that the DNA can be added into.

2.2.2. Gel Electrophoresis

Gel electrophoresis was performed to separate DNA fragments. Prior to electrophoresis, 6X loading dye, containing the dyes bromophenol blue and xylene cyanol FF (Thermo Scientific ©), was added to the sample at a concentration of 1X.

Depending on the size of the gel and the voltage, the running time can vary from 20 to 60 minutes. Generally a voltage of 80 – 100 Volts and electric charge of 400 milliampere (mA) was used.

2.2.3. Gel Imaging

After the DNA has been electrophoresed on the agarose gel, the gel is imaged using a Gel Documentation System (Gel-doc). This uses UV light which causes the ethidium bromide, bound to the DNA, to fluoresce. The system used in this thesis is from Major Science™ (SmartView Pro 1200 Image System) which utilises a computer program that visualises and saves the gel pictures for viewing on the computer.

2.3. *L. maculans* growth conditions

2.3.1. Solid Media Culture for Growth and Storage

Leptosphaeria maculans fungal isolates were grown using the media described in section 2.2.1. The following guidelines were used to optimise growth.

Media plates were sealed with Parafilm (Parafilm M®) and placed in a growth cabinet, kept at room temperature. A 12 hour dark and 12 hour light cycle was implemented, with light provided by one white light (GE Tri-tech F18T8/865) and by one black light (Crompton F18T8 BLB), the black light to increase spore production. Two weeks were required for the fungal growth to cover the entire petri dish. At intervals of 1 inch, 6 mm filter discs (Grade AA, Whatman®) were placed on the plate to ensure a source for future fungal growth.

2.3.2. Liquid Culture for DNA Extraction

After a period of 2-3 weeks, the isolates were regrown on 10% V8 juice liquid media to obtain better DNA yield from tissue after DNA extraction. Each 145 X 20 mm tissue culture dish containing 50 ml of media (as in section 1.2.2.) was inoculated with a filter disc or plug (0.5 cm square of agar from previous plates) and grown under the same conditions as solid media (section 3.1.). After 2-3 weeks of regrowing the isolates on liquid media, fungal tissue was harvested from the plates.

An autoclaved piece of Miracloth (Merck Millipore) was placed in a funnel and the plate contents were drained through it, with the leftover fungal tissue in the cloth being squeezed to get rid of excess liquid. This tissue was stored in falcon tubes at -80 °C for later DNA extractions.

2.3.3. Liquid Culture for Spore Collection

To obtain spores with which plant inoculations can be performed, fungal tissue was grown according to section 2.3.1. To harvest spores the following procedure was followed.

Solid media plates were flooded with water and the surface scraped using a scalpel or L-shaped cell spreader. An autoclaved piece of Miracloth (Merck Millipore) was placed in a funnel and the liquid from the plate was drained through it to filter out large fungal pieces.

The flow through was poured into 1.5 mL eppendorf tubes and placed in the -20°C freezer within an hour to prevent spore germination. Prior to storage, spores can be quantified using a microscope counting chamber (hemocytometer) and a light microscope.

2.3.4. Fungal Storage

2.3.4.1. Cultures

L. maculans cultures were stored as dried filter disks (Grade AA, Whatman®) for long term storage and placed on fresh plates to re-culture. They were also stored as small agar plugs cut from plates with fungal growth and subsequently stored in sterile water. The cultures were renewed once every 1 - 2 years to ensure the viability of the fungus.

2.3.4.2. Spores

L. maculans spores were stored at -20°C until they were used for infection of plants. The spores were not re-frozen after thawing, as the spores may not infect following repeated freeze/thaw cycles. *L. maculans* spores germinate within the first 4 hours of being thawed so plants were inoculated at this time. Spores were kept for a maximum of 1 year, as they do not germinate and cause disease symptoms optimally after this time.

2.4. DNA Extraction

2.4.1. Microprep DNA Extraction, Phenol/Chloroform

The microprep protocol for DNA extraction described by Fulton et al. (1995) was followed for fungal genomic DNA extraction, with the following modifications. After incubation at 65 °C, 750 µL of Phenol: Chloroform was added to each sample and centrifuged for 5 mins at 10,000 rpm. The DNA samples were resuspended in 50 µL of DNase-RNase free water, quantified using a Nanodrop (ND-1000 spectrophotometer) and visualized for integrity on a 1% TAE-agarose gel containing ethidium bromide, described in section 2.2. For subsequent reactions that are sensitive to phenol contamination, the Phenol: Chloroform step was replaced with Chloroform: Isoamyl.

Notes:

Some issues that arose during DNA extraction include the abundance of secondary metabolites within extracted DNA samples. This was remedied using the silica-based clean up method described in section 2.6.2. The secondary metabolites interfere with

sequence library generation reagents and therefore must be removed to avoid this. It was found that harvesting of younger fungal tissue provided better quality DNA with less secondary metabolites and contaminants than older tissue.

2.4.2. DNA Extraction using Kit

For some fungal samples that were showing high amounts of contamination and low yield using the microprep DNA extraction protocol, a DNA extraction kit was used. The DNeasy® Plant Mini Kit (QIAGEN©) was used according to manufacturer's instructions with the only exception being that the fungal tissue was first ground using a mortar and pestle and then added to the QIAschredder column to begin the extraction.

2.5. DNA Quantification and Quality Control

Quantification and quality control of DNA was performed following DNA extraction procedures in order to determine the concentration and integrity of the DNA before use in PCR, sequencing and Next Generation Sequencing reactions.

2.5.1. Gel Electrophoresis

DNA samples were combined with 6X loading dye and gel electrophoresis performed according to section 2.2.2. Generally a 1% TAE-agarose gel was prepared. The resulting gel image can provide information about the integrity and the concentration of the DNA e.g. if the DNA sample appears as a smear on the gel, it is most likely highly degraded. This can affect downstream applications and DNA extractions may need to be repeated.

2.5.2. DNA Quantification

For accurate quantification of DNA samples, the Qubit® fluorometric assay was used (Life Technologies). This employs a flurometric reagent that binds only to double-stranded DNA, avoiding the overestimation that occurs with spectrophotometric methods such as the NanoDrop technology (Thermo Scientific ©). Two reagents were used here, namely high sensitivity (HS) for DNA samples of low concentration (below 10 ng/ µL), and broad range (BR) for samples with higher concentration (10 – 500 ng/ µL).

2.6. PCR methods

2.6.1. Primer Design

For PCR primer design the Primer3 primer tool was used (Koressaar and Remm, 2007; Untergasser et al., 2012). Primers were designed to have annealing temperatures between 58°C and 65°C and lengths of between 15 and 28 bp. For a primer pair, the temperature difference was selected to be no more than 2-3°C. Primers were tested to minimise self-complementarity and primer hybridisation.

2.6.2. PCR Product Purification

Prior to sequencing, the PCR products were purified using the silica-based method of Boyle and Lew (1995), with the following modifications. The silica dilution was made mixing 100 mg of silica dioxide in 1 ml of UltraPure™ H2O (GIBCO-Invitrogen, DNAse RNAse free). The solution was left to settle overnight. After removing the supernatant the process was repeated over two hours. The final solution contained the purified Silica and upto 1 mL of UltraPure™ H2O. Nine g of 6M NaI (MW-149.9) was diluted in 10 ml of UltraPure™ H2O, wrapped in foil and stored at 4°C. The wash buffer was prepared as listed in Boyle et al (1995). Three times the volume of PCR product of 6M NaI was added to the tube and incubated at 55°C for 5 mins, with hand mixing every 2 mins. 10-15 µL of silica was then added and incubation and mixing was repeated. This was followed by centrifugation at 13,200 rpm for 30 s. The supernatant was carefully removed and 0.5 mL of wash buffer was added. The pellet was completely broken up by pipetting or vortexing. Centrifugation was repeated and the supernatant was removed. Care was taken to completely discard all wash buffer by repeat centrifugation and the remaining liquid was left to air dry for 10-15 mins. The DNA was eluted in 20 µL of UltraPure™ H2O. The DNA was electrophoresed on a 1% TAE-Agarose gel containing ethidium bromide (0.5 µg/mL) for visual quantification (see section 2.5.1). For future use of DNA, the tube was centrifuged at 13,200 rpm for 5 mins prior to use.

2.6.3. PCR Product Sequencing

The purified PCR products were sequenced using the Sanger sequencing method at the Australian Genome Research Faculty (AGRF). The samples were prepared according to DNA sample preparation instructions listed by the AGRF. Each reaction mixture was temperature cycled and contained 6-12 ng of PCR product and 0.8 pmol/µL of the specific

primer in 12 µL with H2O. The precise guidelines for sequencing sample submission can be found at www.agrf.org.au. The purified PCR products were Sanger-sequenced using Big-Dye 3.1 (PerkinElmer, Waltham, MA), using forward and reverse PCR primers, and analysed using an ABI3730xl.

2.7. Next-Generation Sequencing

2.7.1. Sequence Library Generation and Genomic Sequencing

The sequencing library generation kits used in this thesis were manufactured by Illumina (San Diego, USA). All library preparation was performed in our laboratory facilities.

The majority of the genomic sequencing was performed on the Illumina GAIIx and Illumina HiSeq platforms, provided by the Australian Genome Research Facility (AGRF). The Illumina MiSeq platform was also used, which was performed in our laboratory.

2.7.2. Reference Genome

The JN3 (also known as v23.1.3) genome was used as the reference in this study (Rouxel et al. 2011). It is 45.12 Mbp in size and has been scaffolded onto 76 SuperContigs (SCs). These range in size from SC0 at 4,258,568 bp to SC75 at 491 bp. The first 31 SuperContigs (SC0 to SC31) are above 100,000 bp in length and represent the majority of the genome at ~95 %. The main focus of the analyses in this thesis therefore focuses primarily on these contigs. SC30 comprises all of the assembled *L. maculans* mitochondrial DNA (Rouxel et al. 2011). There are 12,611 genes that have been annotated on the reference genome, many of which do not have a functional annotation.

2.8. Bioinformatics tools

2.8.2. Geneious Software Package

For visualising sequences from Sanger sequence and Next Generation Sequencing platforms, the Geneious© (Biomatters Limited) software package was used. This includes alignments of DNA sequence to a reference gene sequence and to the *L. maculans* reference genome (Rouxel et al. 2011).

2.8.3. SNP Prediction

2.8.3.1. SOAP Mapping

The SOAPaligner 2 software (Li et al., 2009a) was used to map DNA sequence reads to the *L. maculans* ('brassicae') isolate v23.1.3 reference genome (Rouxel et al., 2011). The SOAP –r 0 parameter was used to allow a sequence read to align to the genome only once, as well as the –M 2 parameter, allowing only two mismatches to the reference genome.

2.8.3.2. SGSautoSNP

The software SGSautoSNP (Lorenc et al., 2012) was used for SNP prediction between the *L. maculans* isolates with default parameters.

3. Identifying genetic diversity of avirulence genes in *Leptosphaeria maculans* using whole genome sequencing

Published in Functional and Integrative Genomics (2013), 13: pp. 295-308

The published version of this paper can be accessed here:

http://link.springer.com/article/10.1007%2Fs10142-013-0324-5

3.1. Introduction

The most economically damaging pathogen of *Brassica* species, particularly canola (*Brassica napus*; rapeseed; oil-seed rape), is *Leptosphaeria* maculans (Howlett et al., 2001). *L. maculans*, a filamentous ascomycete, is the causal agent of phoma stem canker, commonly referred to as blackleg. In late stages of infection, it spreads through the stem vasculature causing lesions and leading to poor growth, lodging and eventually plant death. This fungus is found in canola-growing regions worldwide such as Australia, Canada and Europe. Increased production of canola in these regions has led to a rise in the severity of the disease. In Australia alone, *L. maculans* infection is responsible for an estimated Australian $100 million in crop losses each year, with average losses ranging from 15–48 % and significant efforts are underway to improve resistance to this disease (Kaur et al., 2009; Tollenaere et al., 2012).

Understanding the characteristics of *L. maculans* is vital for developing an effective and sustainable approach to the management of blackleg disease on *Brassica* species. Given the threat imposed by this pathogen on the global rapeseed industry, it is imperative to gain more insight into the genetic structure of this fungus. Such information will assist in developing resistant Brassicas and eventually reduce crop losses. The completion of the *L. maculans* genome sequence (Rouxel et al., 2011) was a significant development in the study of this fungal pathogen and provides a reference genome to which molecular markers can be physically mapped. This has been highly useful in other plant pathogens with sequenced genomes, such as the wheat pathogen *Parastagonospora nodorum* (Hane

et al., 2007) and the cereal pathogen *Fusarium graminearum* (Cuomo et al., 2007). Importantly, a reference genome also allows mapping of whole genome re-sequencing data, which is becoming a high-throughput, cost-effective method to study genome-wide diversity, particularly for the relatively small, lower complexity genomes of many fungal species. By re-sequencing the genome of different *L. maculans* isolates, variations in genome sequence and structure can be elucidated.

Previously, genetic analyses of blackleg populations were performed using AFLP (Barrins et al., 2004), microsatellite and minisatellite markers (Hayden et al., 2007). Single nucleotide polymorphisms (SNPs) are increasingly popular for studying genomic evolution and diversity (Batley and Edwards, 2007; Duran et al., 2009b; Lai et al., 2012). Sequence variation can have a major impact on how an organism develops and responds to the environment. The high heritability of SNPs makes them an excellent indicator of genetic diversity and phylogeny in pathogenic fungal species, such as *L. maculans*. Additional features of SNPs such as stability during evolution, low mutation rate and their high density in the genome also make them the tool of choice for such analyses (Appleby et al., 2009; Duran et al., 2010b; Hayward et al., 2012b).

The usefulness of SNPs for various applications depends on their genomic location and environment. Genic SNPs are those that have been identified within expressed sequences, most commonly from available EST databases (Duran et al., 2009a; Erwin et al., 2007; Love et al., 2005; Love et al., 2004; Love and Edwards, 2007) and, more recently, next generation transcriptome sequencing data (e.g. mRNAseq) (Batley et al., 2007; Edwards, 2007). These SNPs can be synonymous, resulting in the same amino-acid being translated, or non-synonymous, where a different amino-acid is incorporated into the resulting gene product. Non-synonymous SNPs within transcribed genes can be responsible for phenotypic change, whereby they alter protein structure or function to affect an organism's development or response to environment. Such SNPs that are linked directly to gene function and quantifiable phenotypic change are known as 'perfect' markers.

Due to their location, genic SNPs have an increased likelihood of being deleterious, or associated with a fitness cost for an organism. Genic SNPs are often selected against, which can be observed by the lower frequency of non-synonymous to synonymous base changes in gene regions. SNPs also tend to be more frequent outside of transcribed

regions due to the increased 5-methylcytosine (5meC) abundance, resulting in a greater likelihood of C to T mutations over evolutionary time due to amplified cytosine deamination (reviewed in Edwards et al. (2007c)). SNPs that are linked to traits under selection are highly valuable for identifying genetic loci that contribute to phenotypic variation based on linkage disequilibrium (LD).

Current large-scale SNP genotyping assays, such as the Illumina GoldenGate assay, are cost-effective and flexible for analysing large numbers of SNPs across multiple individuals. To date, no previous studies have implemented the use of SNPs in genetic analyses of *L. maculans*.

Advances in genome sequencing technologies have revolutionised plant and fungal genomics (Edwards et al., 2013; Lee et al., 2012; Marshall et al., 2010). They have made genome sequencing, re-sequencing and SNP discovery highly accessible, high-throughput and cost-effective (Batley and Edwards, 2009; Hayward et al., 2012a; Hayward et al., 2012c). The process of whole genome re-sequencing involves aligning millions of short sequence reads generated on a single next-generation sequencing run to a reference genome sequence. Once this has been achieved, it is possible to identify genetic variation between individuals, which can be linked to variation in phenotype to provide molecular genetic markers and insights into gene function (Batley and Edwards, 2009; Imelfort et al., 2009).

Here, we describe the re-sequencing of two *L. maculans* isolates for the identification of 21,814 SNPs. We demonstrate the application of a novel SNP calling method, SGSautoSNP (Lorenc et al., 2012) and its robustness and sensitivity in identifying polymorphisms in *L. maculans*. We describe the use of these SNPs for phylogenetic analysis, genome analysis, including SNP properties and density in relation to genomic position and predicted function.

3.2. Experimental methods

3.2.1. Fungal samples

The isolates 04MGPP021 and 06MGPP041 (Raman et al., 2012) were used for next-generation sequencing, initial SNP prediction and validation. For simplicity, these are referred to as isolates 21 and 41, respectively herein. A further set of 22 isolates from

various spatiotemporal sources, were used for validation of the SNPs and initial assessment of polymorphism (table 3.1).

Table 3.1: Information about all isolates used in the study.

Isolate number	Reference number	Year cultured	Brassica species isolate cultured from	Cultivar/line isolate cultured from	Resistance genes in cultivar[a]	Stubble collection site	State	Reference
Reference	v23.1.3.	Mid-1990s	N/A[b]	N/A	N/A	Europe	N/A	(Balesdent et al., 2001)
PO21	04MGPP021	2004	B. napus	AG-Emblem	Rlm4, Rlm9	Bordertown	South Australia	
PO41	06MGPP041	2006	B. napus	Skipton	Rlm4, Rlm9	Lake Bolac	Victoria	
1	04MGP-P029	2004	B. napus	TI1 Pinnacle	Rlm3, Rlm9	Geelong	Victoria	
2	07VTJH020	2007	B. juncea	JC05007	Unknown	Horsham	Victoria	
3	09SMJ087	2009	B. juncea	EXCEED™ OasisCL	Unknown	Kaniva	Victoria	
4	05MGPP033	2005	B. napus	Skipton	Rlm4, Rlm9	Yeelanna	South Australia	(Van de Wouw et al., 2010)
5	89/C16	1988	B. napus	Unknown	N/A	Millicent	South Australia	
6	04MGP-S006	2004	B. napus	Surpass400	Rlm1, RlmS	Eyre Peninsula	South Australia	
7	10SMJ041	2009	B. juncea	EXCEED™ OasisCL	Unknown	Tamworth	New South Wales	
8	06MGPP019	2006	B. napus	ATR-Beacon	Rlm3, Rlm4	Wagga Wagga	New South Wales	
9	LM300	2002	B. napus	TI1 Pinnacle	Rlm3, Rlm9	Mt. Barker	Western Australia	
10	LM592	2003	B. napus	TI1 Pinnacle	Rlm3, Rlm9	Mt. Barker	Western Australia	
11	06S012	2006	B. napus	ATR-Beacon	Rlm3, Rlm4	Bordertown	South Australia	
12	06J085	2006	B. napus	Unknown	N/A	Horsham	Victoria	
13	05P032	2005	B. napus	Skipton	Rlm4, Rlm9	Yeelanna	South Australia	(Van de Wouw et al., 2010)
14	IBCN13 (D1)	1991	B. napus	Unknown	N/A	Mt. Barker	Western Australia	(Purwantara et al., 2000)
15	IBCN15 (D2)	1988	B. napus	Unknown	N/A	Streatham	Victoria	(Purwantara et al., 2000)
16	IBCN16 (D3)	1988	B. napus	Unknown	N/A	Mt. Barker	Western Australia	(Purwantara et al., 2000)
17	IBCN17 (D4)	1988	B. napus	Unknown	N/A	Millicent	South Australia	(Purwantara et al., 2000)
18	IBCN18 (D5)	1988	B. napus	Unknown	N/A	Penshurst	Victoria	(Purwantara et al., 2000)

19	IBCN75 (D6)	1987	B. napus	Unknown	N/A	Mt. Barker	Western Australia	(Purwantara et al., 2000)
20	IBCN76 (D7)	1987	B. napus	Unknown	N/A	Mt. Barker	Western Australia	(Purwantara et al., 2000)
21	D8	2005	B. napus	Surpass501TT	Rlm1, RlmS	Mt. Barker	Western Australia	(Marcroft et al., 2012)
22	D9	2005	B. napus	ATR-Beacon	Rlm3, Rlm4	Mt. Barker	Western Australia	(Marcroft et al., 2012)
23	PHW1223 (D10)	1987	B. napus	Unknown	N/A	Mt. Barker	Western Australia	(Purwantara et al., 2000)

N/A not applicable

IBCN numbers represent the IDs of “International Blackleg Collection Network” isolates

a Resistance genes as determined by Marcroft et al. (2012)

b Isolate v23.1.3 is the result of a series of in vitro crosses between European field isolates (Balesdent et al., 2001). Not all data on these isolates were available

3.2.2. Growth conditions

The *L. maculans* isolates were grown and cultured as described in section 2.3.2.

3.2.3. Microprep DNA extraction

DNA was extracted as described in sections 2.4.1. and 2.4.2.

3.2.4. Next generation sequencing

Sequence libraries were prepared for the isolates 21 and 41 using the Illumina paired end (PE) and mate pair (MP) Genomic DNA Sample Prep Kits (Illumina, San Diego, USA) as described in section 2.7.1. One PE and one MP library were made for isolate 41, as well as one MP library from isolate 21. The insert sizes for the libraries were selected to be ~500 bp for PE libraries and ~2,500–3,000 bp for MP libraries. The Illumina GAIIx platform provided by the Australian Genome Research Facility (AGRF) was used to generate paired end and mate pair short sequence reads of 100 bp length. The three libraries were each run individually on separate lanes of the sequencer. Upon completion of sequencing, the sequence reads were quality trimmed to ensure accurate mapping and SNP prediction.

3.2.5. SNP prediction

SNP prediction was performed as described in section 2.8.3.

3.2.6. SnpEff variant analysis

The SnpEff variant annotation tool (Cingolani et al., 2012) was used to predict the effect of the identified SNPs on the annotated genome of *L. maculans* within different genomic DNA sequences, including putative exons, introns, and gene upstream and downstream sequences. The patterns of codon usage and the ratio of transitions/transversions resulting from SNPs were also recorded.

3.2.7. SNP validation

In order to validate the SNP prediction and ensure no ascertainment bias in the isolates used for SNP identification, a selection of 20 SNPs, with a SNP score of 2 or more from SGSautoSNP, were chosen for PCR and sequencing (table 3.2). These were located on different SuperContigs of the v23.1.3 reference genome ranging from the largest (SC0, 4.26 Mbp) to one of the smallest (SC60, 6,199 bp). These were subsequently annotated as SNP 1–20 (table 3.2). Primers were designed to achieve a product size of 200–400 bp.

PCR reactions were carried out in 20 µL containing 1X reaction buffer with 20 mM MgCl2 (Scientifix, VIC, Australia), 10 mM dNTPs (Scientifix, VIC, Australia), 2 units of iTAQ DNA polymerase (Scientifix, VIC, Australia), 0.1 mM of each primer and 20 ng of DNA template. The cycling conditions were as follows: 94 °C for 3 min, followed by 30 cycles of 94 °C for 30 s, 58 °C for 40 s and 72 °C for 30 s and a final extension at 72 °C for 4 min. PCR products were visualized on a 1 % TAE–agarose gel containing ethidium bromide (0.5 µg/mL), see section 2.5.1. The GeneRuler™ 1 kb DNA Ladder or the GeneRuler™ 100 bp DNA Ladder were used as size standards (both Thermo Scientific, MA, USA; 0.5 µg/µL, 50 µg).

PCR products were silica-purified (Boyle and Lew, 1995) and eluted in 20 µL of DNAse RNAse free UltraPure™ H2O (GIBCO-Invitrogen, CA, USA), see section 2.6.2. An aliquot was then run on a 1 % TAE–agarose gel containing ethidium bromide (0.5 µg/mL) for visual quantification (as described in section 2.2.1 and 2.2.2). For all downstream applications, tubes were centrifuged at 13,200 rpm for 5 min prior to use to pellet silica beads. Purified PCR products were directly sequenced using the PD Sanger sequencing

service at the Australian Genome Research Faculty (AGRF), Brisbane, as described in section 2.6.3. Geneious Pro (Drummond et al., 2012) was used to align and assemble all sequenced SNP regions per isolate to the reference SuperContig used to identify the SNPs.

3.2.8. Phylogenetic analysis

Data from SNPs 1, 2, 3, 5, 6, 7, 8, 9, 11, 14, 15, 16, 17 and 18 for each isolate, including the reference, was used to produce an unrooted dendrogram computed based on the Euclidean distance of the isolates, using the R project for statistical computing (The R Foundation for Statistical Computing version 2.13.0, 2009). Isolate 7 was eliminated from the tree due to insufficient sequence data. Both alignments and assemblies were scanned manually to detect novel SNPs in the flanking regions of the existing annotated SNPs. A 'dudi.pca' principal component analysis was conducted using the R project for Statistical Computing (The R Foundation for Statistical Computing version 2.13.0 2009).

Table 3.2: Predicted SNPs chosen for validation.

SNP no.	Super Contig	Position	SNP score	Base change	Forward Primer (5´ to 3´)	Reverse Primer (5´ to 3´)
1	0	369650	15	A/G	CATACTGCAGATTGAGGATGC	GCTACGTTTGCCCTCATTC
2	20	433691	8	C/T	ACGCCCTCGGTTAGCTC	ATCGGTCTCGGAGTGGTC
3	25	259169	5	A/T	GAGGCTTGAAATAGTGAGGTACG	CATCCATCACAGCGCAG
4	30	115	6	T/G	GGAGCTAGTAAGCGATGTCG	AGGTAAACCATAGGTTACCATACCC
5	5	996624	4	A/G	CGTCAACTCTACCGACTCG	ACGCTGCCAGCTTGG
6	60	506	2	A/T	GCTCTTGGGAACCTATCTTCC	CGACGCTGACGTGCC
7	9	1524477	8	C/T	CGACTCTGCAAGACACCG	CCTGATGGGACTTGAGCC
8	9	789721	11	A/G	CTTGTATCGAGATAGGTGGACTG	GGTGGATTACTCTACTTGGAGGC
9	2	838930	13	A/C	AGGGATTGTACACTGTGATTACCAG	GATGAGCGTGGACTCGG
10	4	1281059	15	C/T	CATACACTGCAACTTGATGGATG	TGAATGTTCTTGCGTTGCC
11	4	1362010	23	G/T	GCCGAGCACAGAAGAGC	CGTTGAGGTTAGAGGTGGAGTG
12	8	169231	14	C/T	CCCGAGCCACTACCACTAC	GCGATATAGGAGAACATGTGCC
13	2	889136	13	A/G	CTAGAGAACATGGACGGTACTCTAG	ACACACTCTTAGTCTCACGCAC
14	8	1296096	14	A/G	CTTGGACAATAGAGACATTCAGTG	GGGGCGGCTGGATTG
15	12	352844	10	A/G	GGGCAAGAGAGCGAATG	GCACAAACACCGTAGGCC
16	12	838085	18	C/G	TCGCCCTGTATAGCTATTTCG	GCATCACTGTATACCGCCG
17	12	1107690	5	C/G	GCGGGTTGAGGTTGCC	CGCACGGGTGAAGGAAG
18	14	1268037	11	A/G	CGGCGCGTCAAAGAC	CCTCACCGCTGTTGGC
19	14	852465	6	G/T	CGTCGTCGGCTACCACAG	GCACATGCAATCGCAAGAG
20	3	562718	5	C/T	CGTCTCACGACCCTTCATG	CTCGTCCACCTCTACAACAACG

3.3. Results

3.3.1. Sequencing

To generate a SNP resource for *L. maculans* genome analysis we performed whole genome next generation sequencing (NGS) on two blackleg isolates. In total, 76,313,272 short sequence reads of 100 bp were generated, of these 29,051,680 reads were from isolate 21 (average insert size 3,000 bp) and 47,261,592 reads were from both libraries of isolate 41 (19,120,582 from MP (average insert size 2,500 bp) and 28,141,010 from PE (average insert size 480 bp). Predicted genome coverage for these isolates was high when sequence tags were aligned to the reference genome (Rouxel et al., 2011): 140× coverage for isolate 41 and 86× coverage for isolate 21. The minimum genome coverage was 4× and the maximum was 602× with the mean coverage being 28.308×. The highest coverage was seen in SuperContigs 1–40. No significant difference in coverage was observed between AT and GC rich regions, with the average read coverage being uniform across the genome.

3.3.2. SNP discovery

SNP prediction using SGSAutoSNP (Lorenc et al., 2012) revealed 21,814 SNPs between isolates 21 and 41. This indicates an abundance of SNPs in the *L. maculans* genome, with an average of one SNP every 2,065 bp. SGSAutoSNP records only those SNPs where the polymorphism is present in a minimum of two unique, overlapping sequence reads per individual, minimising false-positive SNP calls due to possible sequence errors. To further provide a measure of confidence in SNP prediction this software also generates a SNP score: representing the minimum number of sequences per individual containing the polymorphism. In this study, the SNP score ranged from 2 to 43, with an average of 6.96.

For each SNP, the directionality of the change cannot be inferred from the data, thus polymorphisms were grouped alphabetically, that is, A > G and G > A were grouped as A > G and so on (table 3.3). A greater number of transitions (A > G or C > T) (17,663) than transversions (A > C, A > T, C > G or G > T) (4,151) were identified. This was consistent across all SuperContigs, with an average ratio of 4:1 transitions to transversions (table 3.3).

Table 3.3: SNP details for all SNPs detected.

SNP type	Base change	SNP count[a]	Percent[b]
Transition	A > G	8,905	40.82
Transition	C > T	8,758	40.15
Transversion	A > C	1,088	4.99
Transversion	A > T	1,021	4.68
Transversion	C > G	1,038	4.76
Transversion	G > T	1,004	4.6

[a] Number of each SNP type predicted
[b] Percent value overall

3.3.3. SNP validation

As an initial validation of the predicted 21,814 SNPs between isolates 21 and 41, 20 SNPs were selected for amplification and direct Sanger sequencing of PCR products. These SNPs were chosen to represent a range of SGSautoSNP scores, including the lowest possible score of 2, up to 23, with an average of 10.5. Amplification was successful for 18 (90 %) of the predicted SNPs and 14 of these were amplified and sequenced in a further 22 isolates (table 3.1) to check for ascertainment bias.

Of the 18 SNPs amplified in isolates 21 and 41 all were shown to be true polymorphisms, following alignments of the obtained sequences to the *L. maculans* whole genome shotgun sequences using Geneious Pro (Drummond et al., 2012). This gave an accuracy of 100 % for SGSAutoSNP-based SNP prediction. Of the 14 SNPs which were subsequently analysed across 22 samples, six were within coding regions and eight within intergenic regions. Eight of the 14 SNPs were transitions, five were transversions, and one was a tri-allelic variant which was both a transition and a transversion. The polymorphism information content (PIC) of these, representing the usefulness of each SNP as a marker for population genetic and phylogenetic analysis, ranged from 0.06 to around 0.5 for different SNPs with an average of 0.33. This indicates that the tested SNPs provide a good level of polymorphic information for use in downstream phylogenetic studies. There was no

observable difference in the PIC values according to SNP type (transition/transversion) or location in the genome (table 3.4). Similarly no observable difference in PIC values could be seen based on the SNP score.

Table 3.4: Properties of the validated SNPs.

SNP number	SuperContig	Position[a]	Genic location	Base changes	PIC[b]	SNP Score
SNP1	0	369,650	CDS–Syn	G > A	0.4352	15
SNP2	20	433,691	CDS–Syn	C > T	0.1472	8
SNP3	25	259,169	Intergenic	A > T	0.4862	5
SNP5	5	996,624	CDS–Nonsyn	G > A	0.1472	4
SNP6	60	506	Intergenic	A> T/G	0.4966	2
SNP7	9	1,524,477	Genic–Syn	C > T	0.3085	8
SNP8	9	789,721	Intergenic	A > G	0.0671	11
SNP9	2	838,930	Intergenic	A > C	0.3299	13
SNP11	4	1,362,010	Intergenic	G > T	0.4898	23
SNP14	8	1,296,096	Intergenic	A > G	0.4959	14
SNP15	12	352,844	CDS–Nonsyn	G > A	0.4445	10
SNP16	12	838,085	Intergenic	G > C	0.2188	18
SNP17	12	1,107,690	Intergenic	C > G	0.4133	5
SNP18	14	1,268,037	CDS–Nonsyn	G > A	0.0643	11

CDS coding region of a gene, *PIC* polymorphism information content, *Syn* synonymous change, *Nonsyn* non-synonymous change

[a]Nucleotide position on the supercontig

[b]Botstein et al. (1980)

3.3.4. Phylogenetic analysis

In order to determine the suitability of the SNPs for phylogenetic analysis and to understand the genetic relationships among the *L. maculans* isolates studied, cluster analysis was performed, based on the combined sequences of the 14 SNP loci. Isolate 7 failed to amplify and was discarded from further analysis. The Euclidean distances between pairs of accessions were then used to generate a dendrogram (Fig. 3.1). On the

phylogenetic tree, as indicated in Fig. 3.1, seven of the nine isolates collected from Western Australia grouped together in one clade. The phylogenetic tree indicates that based on genetic distance, isolates 2 and 3 are identical with these SNPs. These isolates were both collected from *Brassica juncea*. Isolates '21' and '41' that the SNP discovery was performed on are considerably genetically different. Isolate 14 is shown to be an outlier. No association can be observed between isolate similarity and place of collection, year of collection or cultivar source.

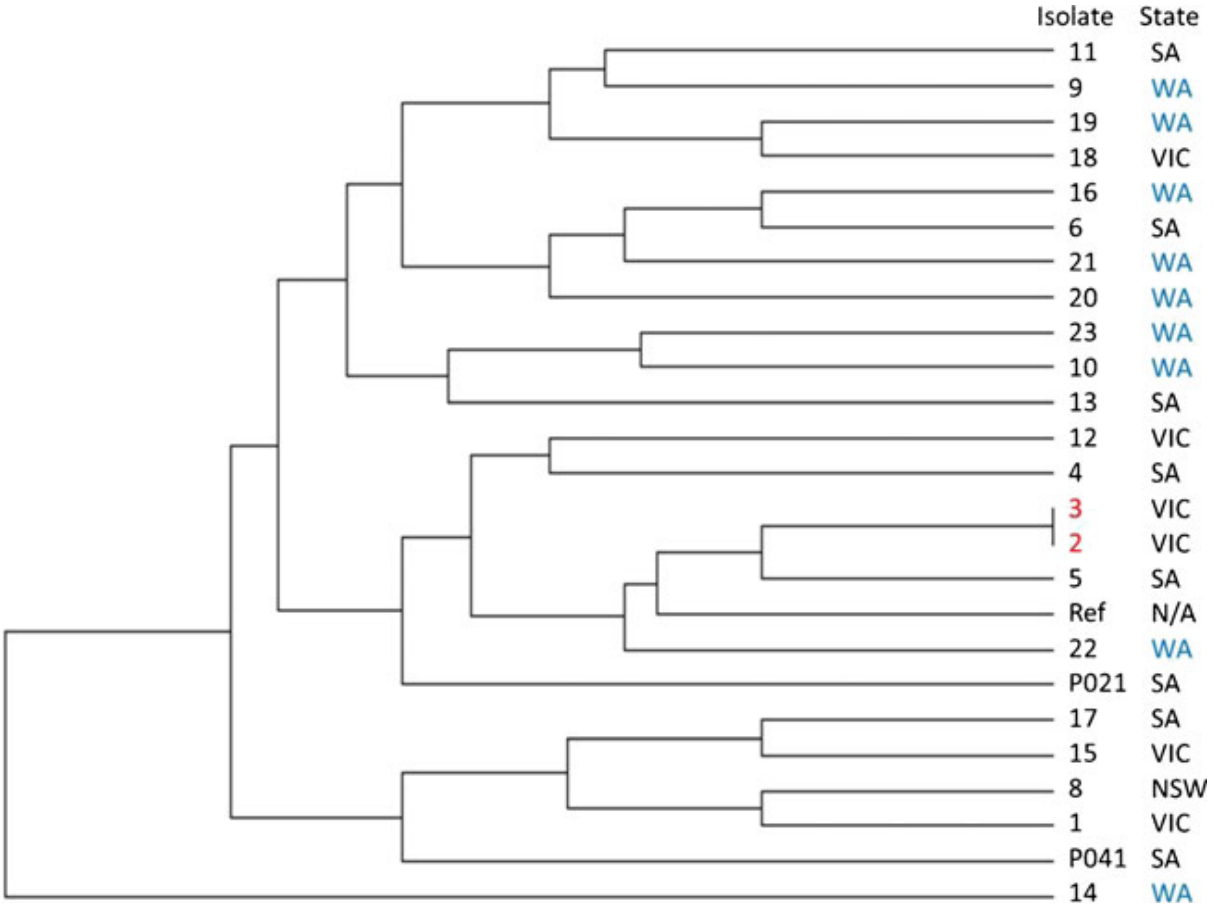

Figure 3.1: Dendrogram showing the relationship between 24 Australian isolates as well as the reference (*Ref*) *L. maculans* isolate, v23.1.3. *Highlighted in red* are two isolates cultured from *B. juncea* rather than *B. napus*. *Highlighted in blue* are the isolates that were cultured from sites in West Australia.

The results from the principal component analysis confirmed the results obtained from the phylogenetic analysis. There is a random spread of isolates along the two principal component axes, with no relatedness between isolates. Isolates 21 and 41 are again shown to be different, and 14 is an outlier (Fig. 3.2). In contrast, isolates 2 and 3 group

slightly differently. Axis 1 represents principal component 1 and axis 2 represents principle component 2.

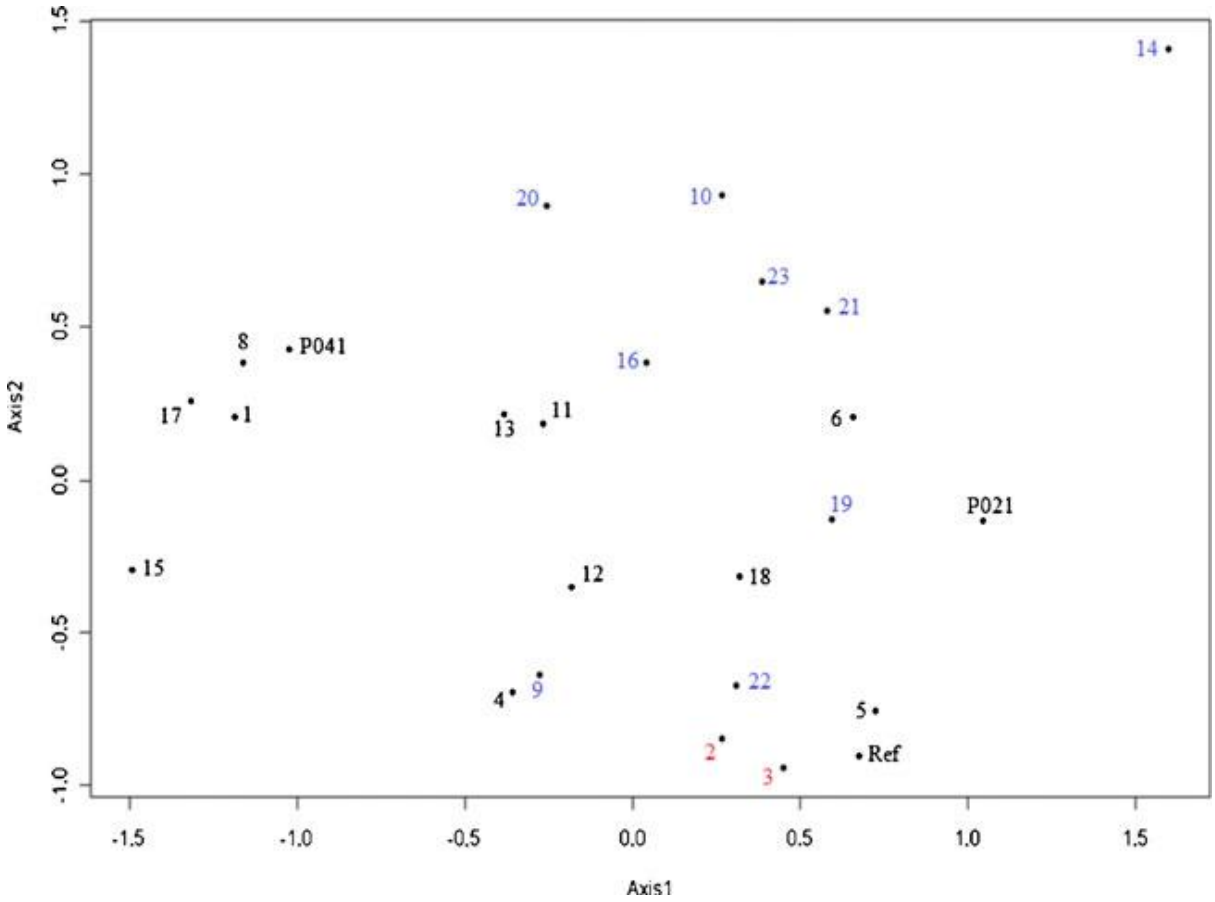

Figure 3.2: PCA showing relationship between blackleg isolates used in this study. *Highlighted in red* are two isolates cultured from *B. juncea* rather than *B. napus*. *Highlighted in blue* are the isolates that were cultured from sites in West Australia. *Ref* refers to the reference isolate, v23.1.3.

3.3.5. SNP properties

Analysis indicates that there is, for the most part, a consistent average SNP frequency across the genome of *L. maculans*, with the majority of SuperContigs over 500,000 bp in length ranging from one SNP every 1,166 bp (SuperContig 22) to one SNP every 3,168 bp (SuperContig 10). Exceptions to this finding are SuperContigs 30 and 37, which display an unusually low number of polymorphisms - being one SNP every 38,715 bp and 17,397 bp, respectively (Fig. 3.3).

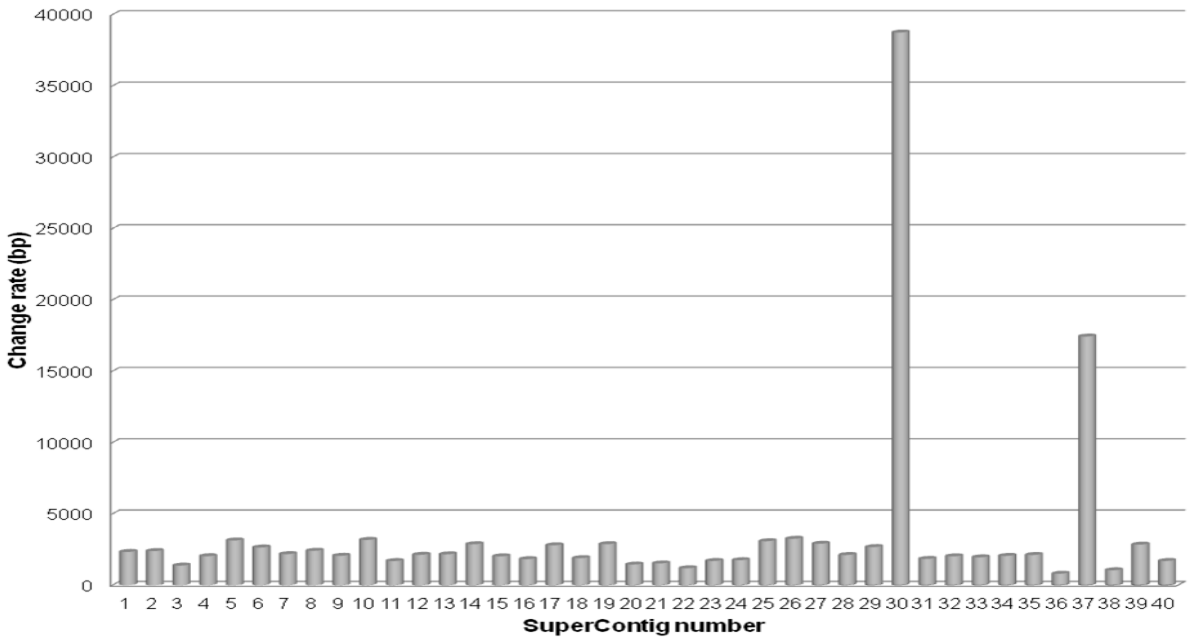

Figure 3.3: SNP density within the *L. maculans* genome. SuperContigs >20,000 bp are included. Change rate denotes the frequency of SNP changes per base pair (e.g. one change every 3,000 bp).

To predict the possible functional effect of SNPs in *L. maculans*, SnpEff was used to determine modifying SNP effects, based on the predicted coding regions of the annotated genome. The total number of effects that the predicted SNPs have on the genome of *L. maculans* was 58,434, with a ratio of missense/silent mutations of 1.0263. The predicted modifying effects of these SNPs is summarised in table 3.5.

Table 3.5: Modifying SNP effects.

Type (alphabetical order)	Count	Percent
Downstream	22,583	38.65
Intergenic	10,234	17.51
Intron	825	1.41
Non Synonymous Coding	1,782	3.05
Splice Site Acceptor	5	0.01
Splice Site Donor	5	0.01
Start Lost	8	0.01
Stop gained	39	0.07
Stop lost	8	0.01
Synonymous coding	1,743	2.98
Synonymous stop	9	0.02
Upstream	21,193	36.27

The highest number of SNPs can be seen in upstream and downstream regions (~75 %) of predicted genes (Fig. 3.4). Coding regions are less affected with only 6.03 % of modifying effects being predicted within coding regions.

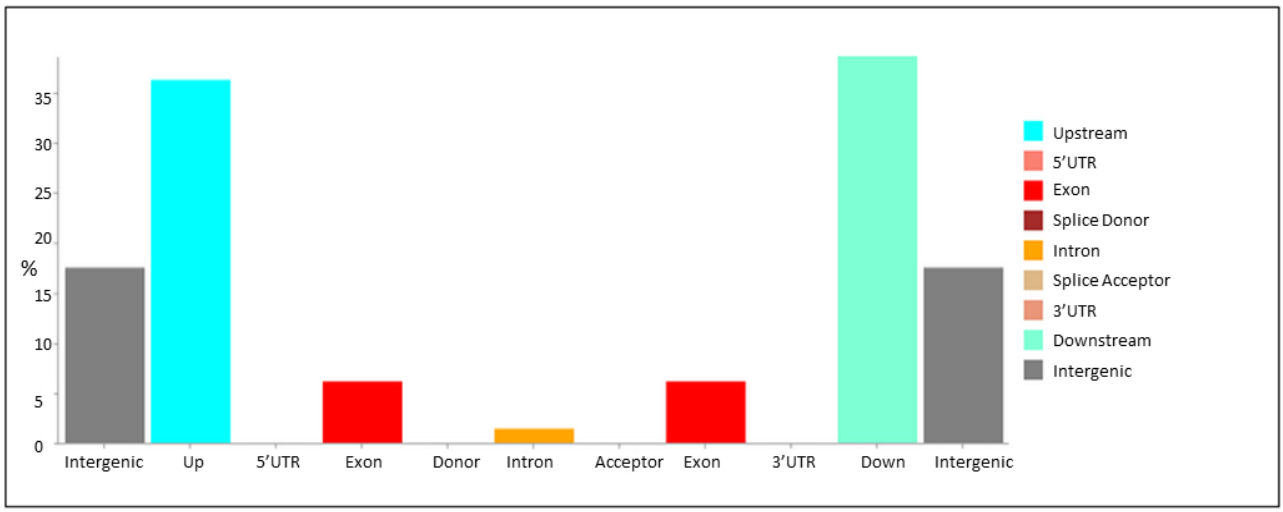

Figure 3.4: SNP distribution variation within the *L. maculans* genome. Percentage (*%*) of SNPs within different genomic regions.

3.3.6. SNP density

SNP density can be seen to vary across the SuperContigs (Fig. 3.5), including in flanking regions to *AvrLm* genes. SC6 shows a low SNP frequency between bases 190,000 and 245,000 (Fig. 3.5). This region contains ~22 predicted genes, including deoxyribonuclease tatD gene, asparagine synthetase, exportin-1 and transport protein sec23, all of which play important roles in cell function and metabolism. Isolates 21 and 41 were highly similar to the reference in all these genes, further highlighting the likely conservation of these putative genes. No difference in sequence read coverage was seen between any of these regions (30–50× coverage throughout), further suggesting that these polymorphisms are valid. The high peak seen at base 970,000 of SC6 shows a SNP frequency of 41 SNPs/10 kb, which is around 1 SNP every 250 bp. This region contains no predicted genes and may show this level of diversity due to low selection pressure in this region. SC12 exhibits a SNP frequency of up to 16 SNPs/10 kb at the *AvrLm4-7* gene locus.

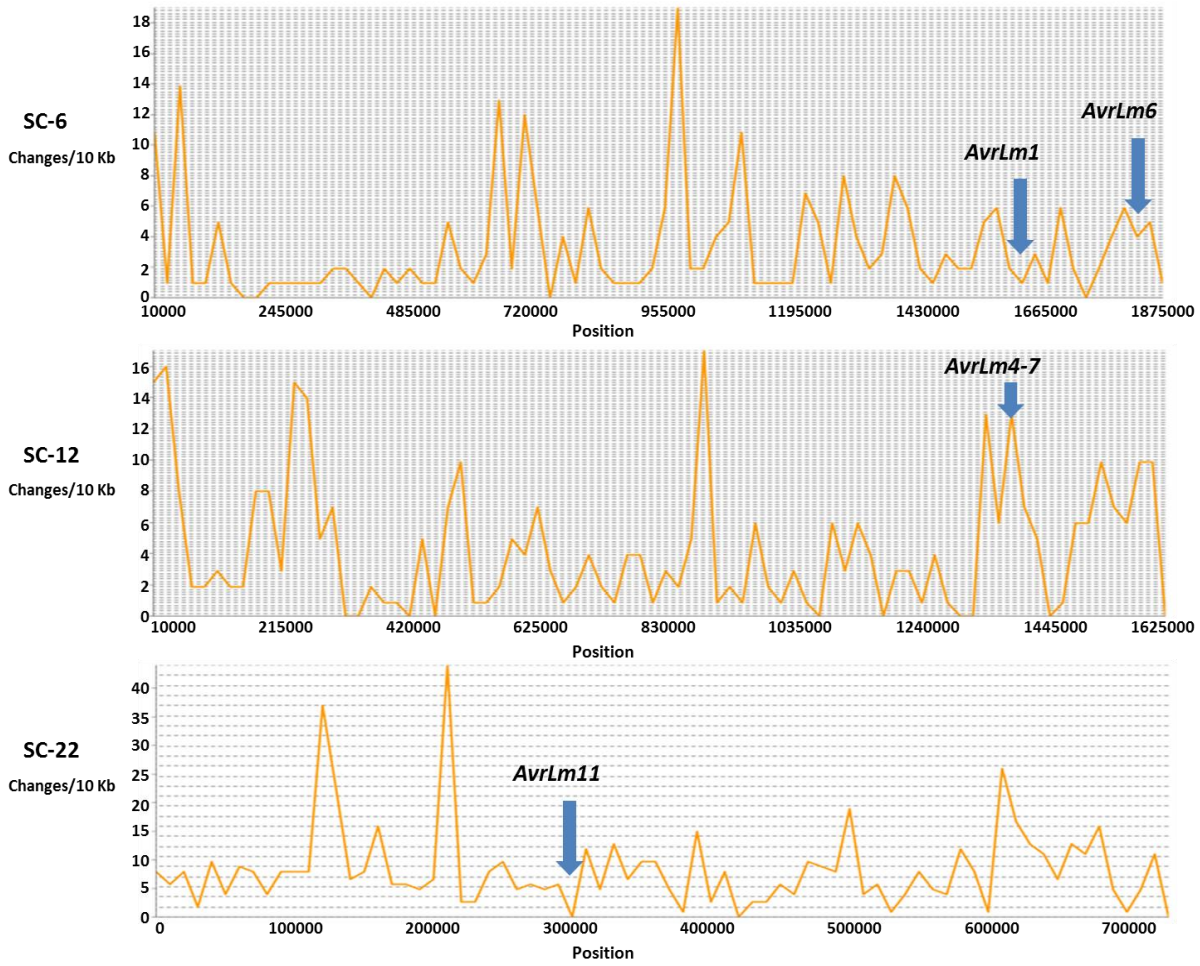

Figure 3.5: SNP frequency (changes/10 kb) plotted against SuperContig position. Position of *AvrLm* genes is shown.

3.3.7. *AvrLm* gene regions

Avirulence genes are a sub-class of effector proteins that elicit host immune responses and are involved in resistance and susceptibility of the host to a pathogen such as *L. maculans*. Alignment of sequence tags from isolates 21 and 41 to the reference genome allowed comparison of previously published avirulence genes in *L. maculans*: *AvrLm1* (Gout et al., 2006), *AvrLm4-7* (Parlange et al., 2009), *AvrLm6* (Fudal et al., 2007) and *AvrLm11* (Balesdent et al., 2013). A recently documented C-G SNP in the *AvrLm4-7* gene at base 358 (Van de Wouw and Howlett, 2012) was found between isolates 21 and 41 using the sequence tags generated in this study (Fig. 3.6). In this study, isolate 41 was found to contain the virulent C allele (arginine amino acid residue) while isolate 21 has the

avirulent G allele (glycine amino acid residue). An additional previously characterised SNP (Daverdin et al., 2012) was identified in the coding region of the *AvrLm4-7* gene at base 256 (Fig. 3.6). The SNP lies in the first of two exons of the *AvrLm4-7* gene and changes the base at this locus from a negatively charged D (aspartic acid) to a neutral N (asparagine) in isolate 41.

Sequence comparison of the genomic region surrounding the *AvrLm6* gene showed an absence of sequence tags for isolate 41 in the coding region. Isolate 41 has previously been shown, to lack the functional *AvrLm6* gene while isolate 21 possesses it (Raman et al., 2012). Alignment of the non-coding area around this gene has shown that sequence tags from both *L. maculans* isolates are present in the close proximity of this gene. An alignment of the sequence tags of isolates 41 and 21 showed that no tags matched the *AvrLm1* locus. The recently identified *AvrLm11* gene (Balesdent et al., 2013) was present in both isolates with 100 % sequence identity to the reference genome. This has not been previously characterised in Australian isolates.

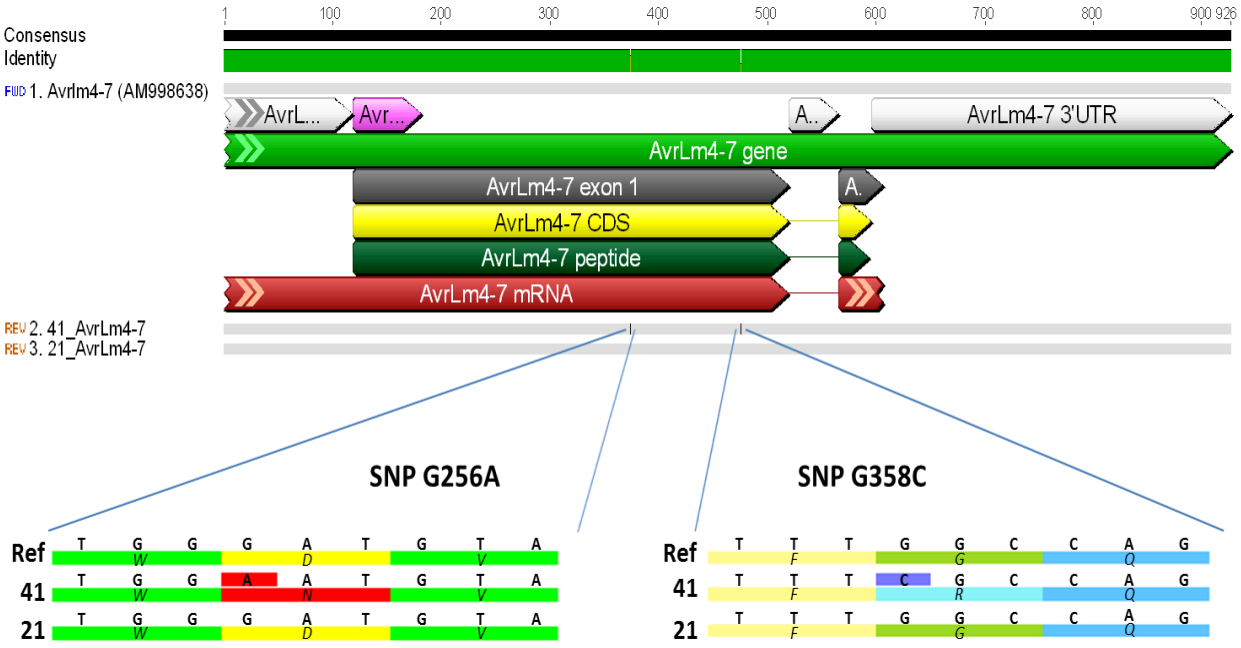

Figure 3.6: Alignment of *AvrLm4-7* genes from the reference, isolate 21 and isolate 41. The two SNPs in isolate 41 are shown.

3.3.8. Housekeeping gene comparison

Sequence comparison within putative housekeeping genes, between isolates 21 and 41 showed a high level of conservation in these isolates compared to the reference. A number of these genes are seen in SuperContig 17 around position 400,000. Isolates 21 and 41 show 100 % sequence identity to the reference sequence in the following putative (annotated as 'similar to') genes: actin gene, Hsp70 nucleotide exchange factor fes1, monosaccharide transporter and major facilitator superfamily multidrug-resistance gene. Overall, SuperContig 17 has a lower level of SNP frequency than the neighbouring SuperContigs, indicating that perhaps a range of important conserved genes are located on this scaffold.

3.4. Discussion

Our results demonstrate that NGS technology and automated SNP prediction is a viable option for genotyping efforts in *L. maculans*. With new barcoding methods that allow the sequencing of multiple samples in a single lane, NGS can prove a cost-effective method of producing whole genome sequence data in small genomes such as the ~45.12 Mbp genome of *L. maculans*. We demonstrate how these data can be used for genome analysis.

3.4.1. SNP prediction and validation

Combined, the three sequence libraries provided genome coverage of 226×. This is many times the coverage necessary for high confidence SNP calling, with current research suggesting that 10–30× coverage for assembled genomes is sufficient (HGP, 2013). High levels of coverage also allow for confidence in SNP prediction, as SGSautoSNP will disregard loci where a difference in SNP call occurs in reads from the same isolate. Furthermore, re-sequencing of genomes with high coverage can potentially provide information on large-scale recombination events, as well as copy number variation and presence/absence variation (Li et al., 2012).

The frequency of SNPs between the two *L. maculans* isolates is consistent with previous SNP identification studies in fungi; for example, average SNP density was one SNP every 3,449 bp in *Fusarium graminearum*, a plant pathogenic fungus that causes head blight disease in wheat and barley (Cuomo et al., 2007). The –r 0 parameter in the SOAPaligner

ensures that reads align only once and do not align equally well to multiple regions of the genome, thus minimising false SNP calls in repetitive regions. This provides confidence in the SNP prediction, further making sure that SNPs are only predicted in single-copy regions of the genome.

A higher number of SNPs was observed in upstream and downstream regions of predicted genes, rather than coding regions. We expect a higher level of conservation within coding regions (Castle, 2011), as well as even higher conservation at splice sites and start/stop codons. This can also be seen in these data with only 0.11 % of modifying SNPs in these regions. The transition to transversion ratio was shown to be similar to SNP prediction in *Candida albicans*, where out of 561 SNPs predicted, 437 were transitions and 126 were transversions (Forche et al., 2004). Repeat-induced point mutations (RIPs) may play a role in the high transition to transversion ratio. Previous studies identified regions of high RIP mutation levels in *L. maculans*, whereby G/C bases are mutated to A/T bases as a protective mechanism against the accumulation of repetitive elements in the fungal genome (Fudal et al., 2009; Rouxel et al., 2011). The greater number of transitions than transversions may also reflect the high frequency of C to T mutation following methylation (Coulondre et al., 1978).

SNP prediction by SGSAutoSNP was found to be correct even for the lowest SNP score of 2, whereby a minimum of two aligned sequence reads per isolate (minimum of four from the two isolates) aligned across the polymorphism. This can be useful in non-model organisms where sequence coverage may not be sufficient for genome assembly. On the other hand, high levels of coverage were also shown to provide accurate prediction, with a SNP score of 23 (SNP11) successfully predicted and validated.

The high SNP validation rate seen here is consistent with previous applications of SGSAutoSNP, whereby SNPs from canola and wheat Illumina whole genome sequence reads were predicted with an accuracy of over 96 % and 93 %, respectively (Lorenc et al., 2012). We have now shown that SGSAutoSNP can be successfully implemented in predicting SNPs in the fungal genome of *L. maculans*. The genomes of these organisms vary greatly in size (*L. maculans*, 45.12 Mb; *Triticum aestivum* (wheat), 16,000 Mb) and in many orders of complexity, with the *L. maculans* genome being haploid and the wheat genome a hexaploid. However, the method described in this research proves to be a robust method of predicting SNPs and thus creating novel molecular markers in small

fungal genomes. A previous study in the ascomycete plant pathogen *Ophiognomonia clavigignenti-juglandacearum* (~16 Mbp) identified only 298 SNPs within eight isolates using Roche's GS Mapper, of which only 45 could be found to be true polymorphisms (Broders et al., 2011).

The main advantage of the SGSautoSNP prediction method over existing SNP callers is that it does not consider the reference for SNP discovery (Lorenc et al., 2012). The reference is needed only to assemble the sequence reads, SNPs are called between the assembled reads. Based on the validation results, the SNP calling was successful in large contigs (>4 Mbp) as well as small contigs, such as SuperContig 60 (6,199 bp). Alignments of small contigs, ESTs and other sequence data could be used in combination to predict SNPs using SGSautoSNP.

Genomic regions of significantly reduced SNP densities can indicate sequence stringency based on high functional conservation of coding or regulatory regions between otherwise diverse individuals. For example, such regions may represent blocks of housekeeping genes or chromosomal regulatory regions required for an organism's reproduction and/or survival (She et al., 2009). SuperContig 30 contains all of the assembled *L. maculans* mitochondrial DNA (Rouxel et al., 2011) thus the low diversity between the two sequenced isolates, compared to the rest of the genome, could be explained given that mitochondrial DNA is generally highly conserved within a species (Kocher et al., 1989) and encodes vital metabolic genes. A more likely explanation, however, is that more than one copy of mitochondrial DNA is present in one or both of the sequenced *L. maculans* isolates. This is known as heteroplasmy and can be seen in humans (Payne et al., 2012), plants (Tian et al., 2006), animals (Kmiec et al., 2006), as well as fungi (Lesemann et al., 2006). This may explain why the SGSautoSNP software does not call SNPs in these regions, as its stringency does not allow for differences between reads of the same isolate. This generally works to eliminate false SNP calls due to sequence error; however, it may in this case lead to the masking of SNPs between isolates due to the presence of SNPs within an isolate.

3.4.2. Population structure and avirulence gene evolution

Interestingly during SNP validation, isolate 10 was unique in displaying a third allele at the SNP6 locus, making this SNP a tri-allelic polymorphism, thought to be a relatively rare occurrence, with one study showing only ~0.2 % of SNPs were tri-allelic (Huebner et al.,

2007). Recent SNP research conducted in the human genome has found that tri-allelic SNPs within populations are present more than twice as often as expected (Casci, 2010). This isolate also displays unique polymorphisms for two other SNPs, in an otherwise synchronised polymorphic set. Such recurring differences indicate that isolate 10 possesses a number of isolate specific alleles.

Few relationships between isolates and collection parameters could be observed. The isolates from Western Australia tended to group together in one clade of the phylogenetic tree. This is consistent with the findings of Hayden et al. (2007) who also observed geographical separation between western and eastern Australian *L. maculans* populations, using microsatellite and minisatellite markers. In addition, the isolates collected from *B. juncea* grouped together. Relatively low sample sizes in this study may preclude any definitive interpretation of the relationships between specific populations. In all, the predicted SNPs were highly conserved and polymorphic in the different population isolates that were included in the analysis. These isolates are from geographically distant locations (up to 3,000 km, from Western Australia to Victoria) and isolated from different growing seasons (from 1987 to 2009). This indicates that these SNPs could be useful in tracking the spatio-temporal spread of loci, and potentially linked traits, in the Australian population.

These initial data suggest that SNPs (RIP-induced or non-RIP-induced), PAVs or partial deletions within genes may all play a role in the virulence and avirulence phenotypes seen in previous studies. Looking at these changes using re-sequencing data can be a method to identify previously uncharacterised *AvrLm* genes in *L. maculans*. Identifying regions of the genome that are present in one isolate yet not in the other may help to determine those areas where the stringency of the alignment software cannot allow alignment of sequence reads due to a high number of polymorphisms, such as in the *AvrLm6* gene region.

Smaller variations, such as SNPs can be rapidly analysed and compared in the re-sequenced individuals by simply mapping the sequence reads to the reference genome. The use of the SGSautoSNP software in this study has identified previously characterised SNPs in *AvrLm4-7*, further highlighting the usefulness of this approach. This can prove to be a simple method of genotyping individuals in important loci of the genome, given that whole-genome sequencing is becoming increasingly accessible and affordable. An interesting observation is the differing SNP frequency within the immediate location of the

AvrLm genes. The *AvrLm1, AvrLm6* and *AvrLm11* genes exhibit a relatively low SNP frequency within their genomic environment, whereas *AvrLm4-7* resides in a region of higher SNP frequency. This may be due to the high selective pressure from the corresponding plant resistance gene acting at the *AvrLm4-7* locus, as both isolates were collected from *B. napus* cultivars containing the *Rlm4* gene. Prior to 2002, 53 % cultivars grown in Australia contained *Rlm4*, decreasing to the current 29 %. Furthermore, *Rlm4* is the most common resistance gene in Australian cultivars (Marcroft et al., 2012). This will have led to strong selection pressure towards the *AvrLm4* locus and therefore higher SNP frequency. The mechanism of *AvrLm* adaptation to resistance gene pressure may play a role in sequence divergence around these genes.

Two previously reported SNPs in the *AvrLm4-7* gene (Daverdin et al., 2012; Van de Wouw and Howlett, 2012) were found to be polymorphic between isolates 21 and 41 using the sequence tags generated in this study. One SNP is known to cause a change from virulence to avirulence against the resistance genes *Rlm4* in *B. napus* (Parlange et al., 2009; Van de Wouw and Howlett, 2012). Consistent with this, the two isolates differ in their infection success in *B. napus* plants containing the *Rlm4* resistance locus, with isolate 41 being virulent and isolate 21 being avirulent (Raman et al., 2012). The polymorphisms lead to a difference in effector/avirulence protein structure and/or function in the *Brassica* host. It may be possible that the two SNPs (bases 256 and 358) show some association, and it is likely that they both play a role in the documented virulence/avirulence phenotypes.

The lack of reads mapping in the *AvrLm6* gene may be due to a deletion or substantial polymorphism in the gene compared to the reference genome preventing the reads from mapping. Substantial polymorphism has previously been shown for isolate 41 in the *AvrLm6* gene region (Van de Wouw et al., 2010); with 43 documented RIP-like mutations, compared to the reference. As no tags were present for the *AvrLm1* locus, this indicates that this allele is most likely not present in these isolates and supports previous observations, using a differential set of *B. napus* lines, that isolates 21 and 41 do not possess the *AvrLm1* gene (Harsh Raman, personal communication). The recently identified *AvrLm11* gene (Balesdent et al., 2013) was present in both isolates with 100 % sequence identity to the reference genome. *Rlm11* has not been commercially released in Australian canola cultivars and therefore the gene is unlikely to be under selection pressure in Australian populations.

3.5. Conclusions

This study demonstrates the applicability of whole-genome re-sequencing in small fungal genomes such as *L. maculans*. It allows for SNP prediction and validation, and the use of these novel markers as a tool for molecular and phylogenetic analysis. Re-sequencing can also be used to validate known mutations, as well as identify putative mutations with modifying effects on the genome. Furthermore, combining SGSautoSNP and SnpEff, we demonstrate a useful pipeline in determining whole-genome polymorphic trends, such as the transition to transversion ratios, polymorphism directionality and the SNP density across chromosomes or contigs.

4. SNP Prediction and Analysis in Ten Differential Isolates

4.1. Introduction

As described in chapter 3, Single Nucleotide Polymorphisms (SNPs) are an increasingly popular molecular marker source due to their high heritability, low mutation rate and high density within genomes. The genomic evolution and diversity of SNPs within Australian *L. maculans* isolates has been recently studied (Patel et al., 2015). Fifty nine blackleg isolates were genotyped using an Illumina GoldenGate 384 SNP assay, designed using SNPs predicted by Zander et al. (2013) and described in Chapter 3. SNPs from across the genome were selected for this assay (described in Patel et al. (2015), in order to determine the extent of diversity within a range of Australian *L. maculans* isolates, as well as providing a validation of the SNP prediction from Zander et al. (2013). The results indicate that the population structure of *L. maculans* is panmictic, meaning that there exists a high rate of sexual reproduction and evolutionary diversification in this pathogen. The Australian isolates of *L. maculans* appear to be a part of a single population, with only minor genetic links between subpopulations based on their proximity to each other (Patel et al., 2015).

The aim of the work presented in this chapter was to conduct a larger-scale SNP prediction across multiple *L. maculans* isolates from a differential set. This differential set represents a variety of isolates collected in Australia from canola stubble from cultivars containing different resistance genes and is routinely used as a tool to determine plant resistance, as they display a wide range of virulence/avirulence towards *Brassica* species and cultivars. The *AvrLm* gene complement of these isolates is known and they have been used in previous studies (Van de Wouw et al., 2010; Van de Wouw et al., 2014). The host resistance (*R*) genes interact with pathogen avirulence (*Avr*) gene products in a process known as effector-triggered immunity (ETI). In order for the host plant to initiate the appropriate defence response, it must be able to recognise the pathogen. Specifically, *R* genes can recognise pathogen *Avr* gene products, known as effector proteins, and thus

combat the spread of blackleg disease. If the *R* gene target (*Avr* effector gene product) cannot be recognised, virulence and disease follows.

As previously discussed, six *Avr* genes have been characterised, namely *AvrLm1* (Gout et al., 2006), *AvrLm2* (Ghanbarnia et al., 2014), *AvrLm6* (Fudal et al., 2007), *AvrLm4-7* (Parlange et al., 2009), *AvrLm11* (Balesdent et al., 2013) and *AvrLmJ1* (Van de Wouw et al., 2014).

The following analyses describe the effects of SNPs on the *L. maculans* genome, and their distribution within it. Of particular interest are disease related genes such as avirulence genes.

4.2. Materials and Methods

4.2.1. Fungal Isolates Used

The isolates used in this study were selected from a group of differential *L. maculans* isolates from a collection used to determine *Rlm* gene presence in canola crop species (table 4.1). These differential isolates were supplied by Prof. Barb Howlett and Dr Angela Van de Wouw at the University of Melbourne and are routinely used in many projects to test a range of *Brassica* varieties for resistance to *L. maculans*. The isolates therefore have a known avirulence gene complement *(AvrLm* genes) and have been used in previous studies (Van de Wouw et al., 2010; Van de Wouw et al., 2014). This is important because it allows us to look for sequence variations between isolates and then attempt to correlate this to the *AvrLm* gene compositions, allowing the possibility of identifying novel *AvrLm* gene candidates. Avirulence/virulence information for *AvrLm11* is not known for these isolates as no cultivars containing *Rlm11* are available to us. The isolates used can be seen in table 4.1. The JN3 Reference details are included as a guide.

Table 4.1: *L. maculans* isolates used in this analysis.

Diff number	Other names	*AvrLm* Gene											
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Sylvestris**	**Hyola50**	**Mustang**
D1	IBCN13	Vir	Avr	Vir	Vir	Avr	Avr	Vir	Vir	Avr	Avr	Avr	Avr
D2	IBCN15	Vir	Vir	Vir	Vir	Avr	Avr	Vir	Avr	Vir	Avr	Avr	Avr
D3	IBCN16	Vir	Vir	Vir	Vir	Avr	Avr	Vir	Avr	Vir	Vir	Avr	Vir
D4	IBCN17	Vir	Vir	Vir	Avr	Avr	Avr	Avr	Avr	Vir	Avr	Avr	Avr
D5	IBCN18, M1	Avr	Avr	Vir	Avr	Vir	Vir	Avr	Vir	Vir	Avr	Avr	Avr
D6	IBNC75	Avr	Vir	Vir	Vir	Avr	Avr	Vir	Avr	Vir	Avr	Avr	Avr
D8		Vir	Vir	Vir	Vir	Avr	Vir	Avr	nd	Vir	Vir	Avr	Vir
D9		Vir	Vir	Vir	Vir	Avr	Avr	Avr	nd	Vir	Vir	Avr	Avr
D10	PHW1223	Vir	Vir	Vir	Vir	Avr	Avr	Vir	Avr	Avr	Avr	Avr	Avr
D12	04MGPS012	Vir	Vir	Vir	Avr	nd	Vir	Avr	nd	Vir	Vir	nd	Avr
JN3 Reference	v23.1.3	Avr	Vir	Vir	Avr	Avr	Avr	Avr	Avr	Vir	nd	nd	nd

Avr means the isolate is avirulent for the corresponding resistance gene e.g. *AvrLm1* can be recognised by *Rlm1*

Vir means the isolate is virulent and cannot be recognised by the corresponding resistance gene, thus causing infection

Nd means there is no data for this resistance gene in this isolate

Hyola50 and ***Mustang*** are Australian canola varieties with to date unidentified resistance

4.2.1. DNA Extraction and Quality Control of DNA

DNA extractions and quality control for the fungal isolates used in this study were performed according to methods described in section 2.4 and 2.5.

4.2.2. Sequence Library Preparation and Genomic Sequencing

The paired-end libraries were made using either the TruSeq or TruSeq Nano Library Preparation (Prep) Kits, following manufacturer's instructions (Illumina), as described in section 2.7.1. We aimed for a 300-500 bp insert size for these libraries. Libraries for isolates D1 and D2 were created using the TruSeq Prep Kit and isolates D3 – D12 using the TruSeq Nano Prep Kit. All library preparation was performed in our laboratory facilities.

Isolates D1 and D2 were sequenced using the Illumina GAIIx platform, with 200 cycle reagents (resulting in 100 bp PE sequence reads). The D10 isolate was sequenced on the Illumina MiSeq platform, using 500 cycle reagents (resulting in 250 bp PE sequence reads). The remaining isolates (D3, D4, D5, D6, D8, D9 and D12) were sequenced using the Illumina HiSeq platform (resulting in 100 bp PE sequence reads).

4.2.6. SNP prediction

SNP prediction was performed as described in section 2.8.3.

4.2.7. SnpEff variant analysis

The SnpEff variant annotation tool (Cingolani et al., 2012) was used to predict the effect of the identified SNPs on the annotated genome of *L. maculans* within different genomic DNA sequences, including putative exons, introns, and gene upstream and downstream sequences. The patterns of codon usage and the ratio of transitions/transversions resulting from SNPs were also recorded.

4.3. Results

4.3.1. Sequencing

The Illumina whole genome sequencing results are presented in table 4.2. The read length for each isolate was 100 bp, apart from D10 for which 250 bp read length was used. Considerable coverage was achieved for all the isolates, ranging from 26.7x for D10 to 266.3x for D1, with an average coverage of 99.65x.

Table 4.2: Whole genome sequencing output.

Isolate Name	Total PE reads	Read Length (in bp)	~ Insert size (in bp)	Total # bp	~ Coverage
D1	59,493,424	100	200	12,017,671,648	266.3x
D2	11,492,603	100	250	2,320,000,000	51.4x
D3	17,778,999	100	580	3,555,799,800	78.8x
D4	22,401,329	100	500	4,480,265,800	99.2x
D5	22,935,182	100	500	4,587,036,400	101x
D6	19,967,228	100	500	3,993,445,600	88.5x
D8	28,463,112	100	500	5,692,622,400	126x
D9	16,293,997	100	530	3,258,799,400	72.2x
D10	2,402,994	250	550	1,206,302,988	26.7x
D12	19,507,705	100	520	3,901,541,000	86.4x

PE reads are 'Paired-End' sequence reads, the output from the Illumina sequencing platforms
Insert size is the size of the fragment used for sequencing
bp stands for 'Base-pair'
Total # bp is the total number of reads (double the PE number) multiplied by the read length
Coverage; the 'x' refers to the 'times' coverage, which is how many times the total bp output can cover the 45.12 Mbp *L. maculans* genome

4.3.2. SNP Discovery using SGSAutoSNP

The SGSAutoSNP pipeline described by Lorenc et al. (2012) resulted in the prediction of 47,097 SNPs between the differential isolates. There is an average of one SNP every 953 bp across the first 40 SuperContigs (SCs), which comprise ~99% of the genome and the majority of coding sequences. These SNP change rates can be seen in table 4.3.

Table 4.3: SNP change rate (the length of the SC in bp divided by the number of SNPs).

SuperContig	Length (in bp)	No. of Changes	Change rate
SC0	4,258,568	4,483	949
SC1	3,378,610	3,078	1,097
SC2	2,939,989	2,583	1,138
SC3	2,348,246	2,425	968
SC4	1,918,205	1,529	1,254
SC5	1,869,450	1,759	1,062
SC6	1,888,674	2,331	810
SC7	1,769,547	1,659	1,066
SC8	1,809,296	1,511	1,197
SC9	1,772,623	1,727	1,026
SC10	1,758,670	1,658	1,060
SC11	1,590,160	1,255	1,267
SC12	1,631,710	1,472	1,108
SC13	1,634,580	1,807	904
SC14	1,533,332	1,242	1,234
SC15	1,560,629	1,637	953
SC16	1,397,653	1,747	800
SC17	1,445,693	1,380	1,047
SC18	1,351,976	1,773	762
SC19	1,186,800	1,647	720
SC20	1,087,932	1,732	628
SC21	1,020,521	900	1,133
SC22	731,443	1,013	722
SC23	521,426	624	835
SC24	475,869	605	786
SC25	318,058	236	1,347
SC26	261,540	260	1,005
SC27	250,629	645	388
SC28	236,098	702	336
SC29	200,940	233	862
SC30	154,863	1	154,863
SC31	143,268	134	1,069
SC32	87,679	246	356
SC33	65,326	289	226
SC34	58,596	221	265
SC35	79,158	193	410
SC36	35,372	93	380
SC37	52,193	88	593
SC38	23,239	40	580
SC39	22,454	105	213
SC40	21,590	34	635
Total	**44,892,605**	**47,097**	**953**

SC20 had the lowest change rate of any SC > 1 Mbp in length with 1 SNP every 628 bp. The lowest change rate of all supercontigs was 1 SNP every 213 bp on SC39. By far the highest change rate (the fewest SNPs) was seen on SC 30 (which contains all mitochondrial DNA) with only a single SNP across its 154,863 bp length.

A comparison of these results to the change rates observed in chapter 3, indicate that the general trend is similar across SC0 – SC20 (figure 4.1). Change rates in chapter 3 however, showed a less uniform change rate across the genome, with strong strong variation between high and low change rates. From SC21 – SC40, a higher variation between the change rates in both studies was observed (a higher fluctuation between high and low change rates).

SuperContig 37 is the most notable example of this, with the change rate in chapter 3 being one SNP every 17,397 bp, whereas this analysis yielded one SNP every 593 bp. SC30 has a very high change rate (low number of SNPs) in both studies with only one SNP predicted in this study and four SNPs predicted in chapter 3 (see figure 4.2), giving a change rate of 154,863 and 38,715 respectively.

The SNP numbers in figure 4.2. indicate that there is a general correlation between SNP numbers between the two studies, with chapter 4 SNP numbers (10 Isolates) being roughly double those of chapter 3 (2 Isolates) SNP numbers.

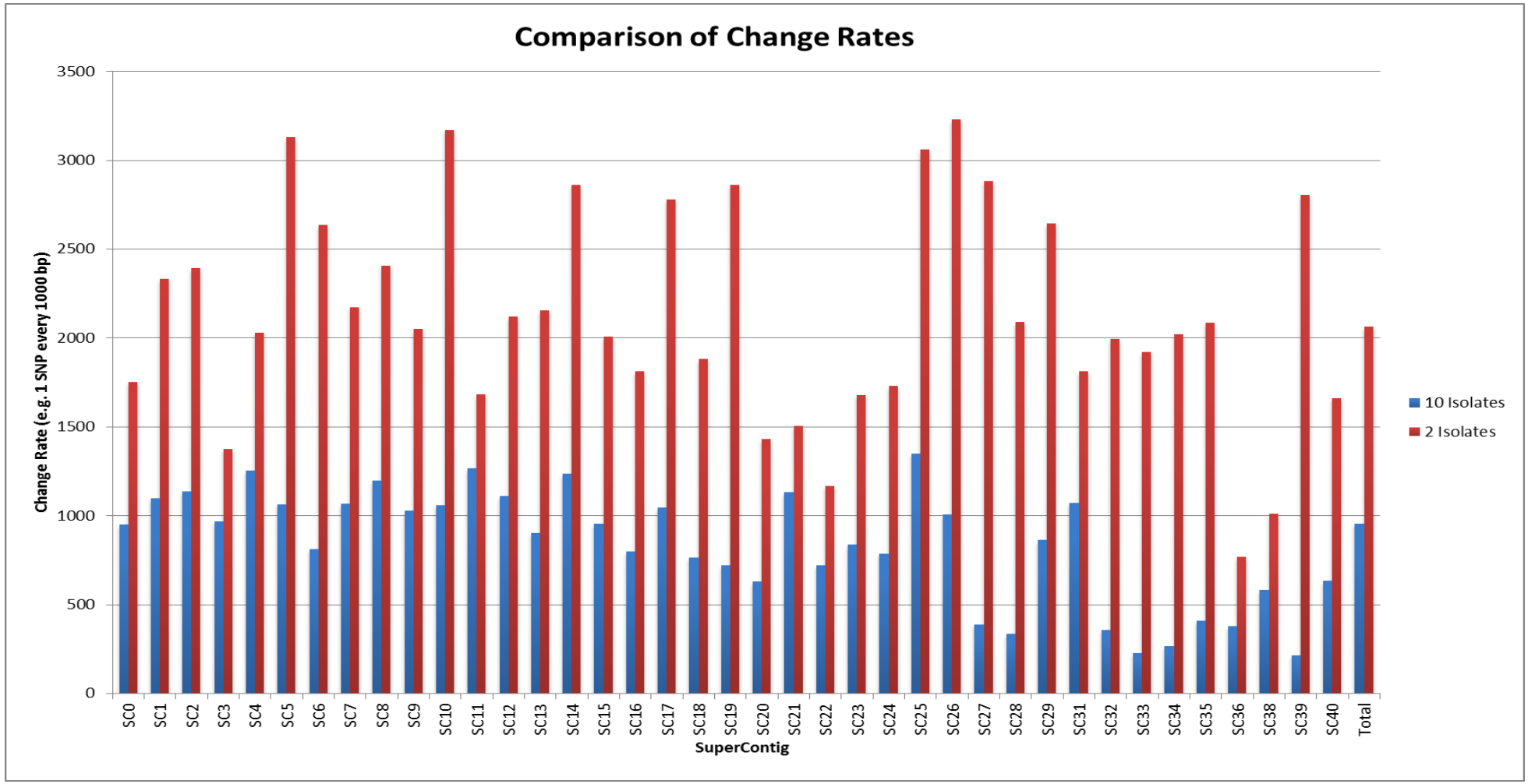

Figure 4.1: Change rate comparison between SNP prediction in 2 isolates (Chapter 3) and 10 isolates. SC30 and SC37 were excluded from this graph as their change rates were over 10 times higher than the average.

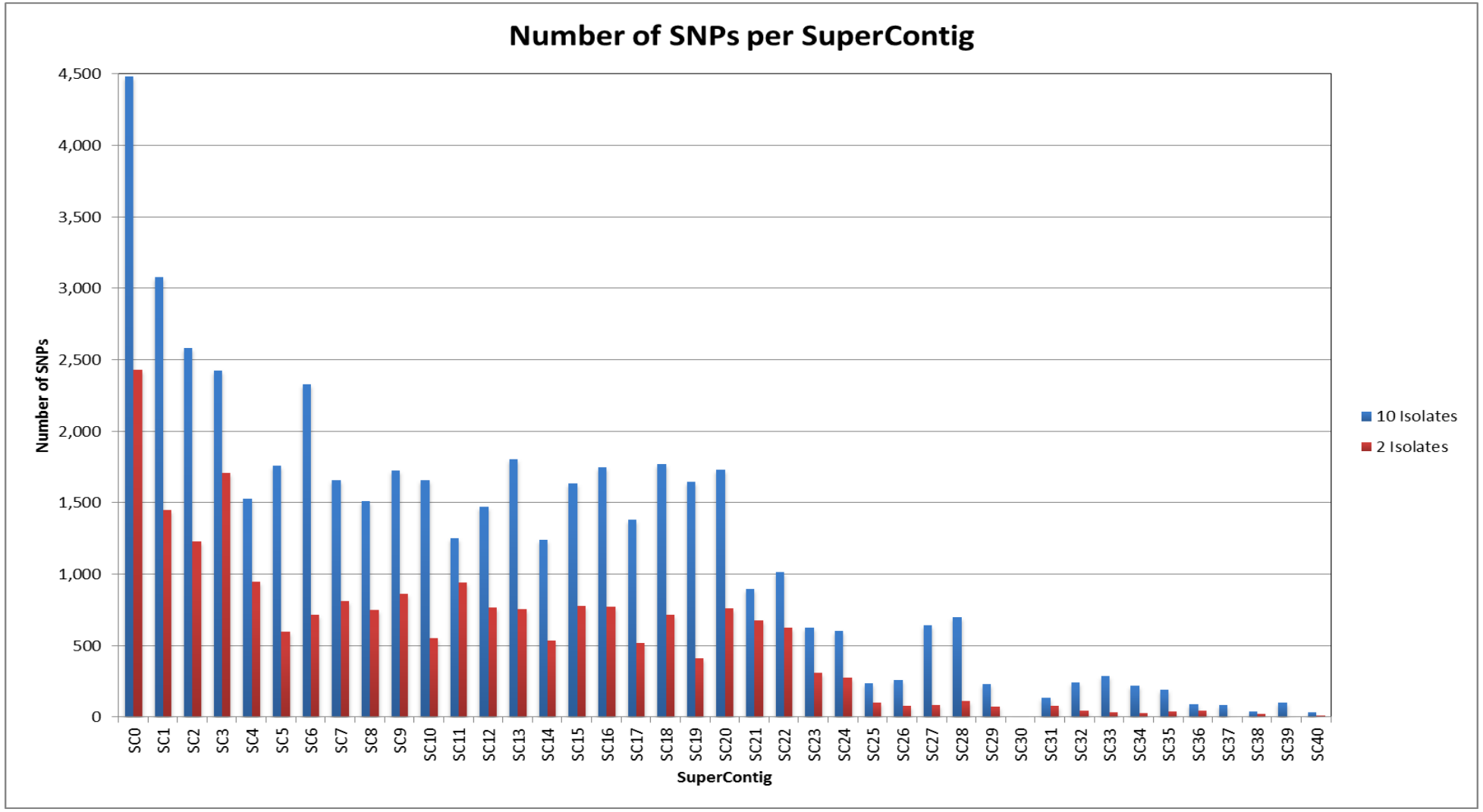

Figure 4.2: Number of SNPs per SuperContig from SNP prediction using 2 isolates (chapter 3) and 10 isolates (chapter 4).

Table 4.4: SNP numbers between differential isolates.

	D12	D10	D1	D2	D3	D4	D5	D6	D8	D9
D12	0	5,776	11,090	10,271	10,708	10,759	11,099	9,205	11,157	10,449
D10		0	5,943	5,643	5,998	6,226	6,313	5,885	6,174	6,226
D1			0	10,468	10,639	11,156	11,752	9,511	11,660	10,854
D2				0	9,950	10,211	10,825	8,835	10,749	10,020
D3					0	10,688	11,202	8,751	10,087	9,539
D4						0	11,348	9,771	11,147	10,414
D5							0	9,779	11,523	10,875
D6								0	9,384	8,578
D8									0	10,117
D9										0

The SNP numbers between the differential isolates are fairly consistent with the average number of SNPs being 9,735 SNPs (table 4.4). The isolate with the highest average SNP number compared to the others is D1 with an average of 10,862 SNPs. The lowest is D10, with an average of 6,051 SNPs compared to the other isolates. Isolate D6 also had slightly lower SNP numbers compared to other isolates. The average SNP number between D6 and other isolates was 8900 SNPs.

Isolates D1 and D5 have the highest number of SNPs between each other with 11,752 SNPs predicted. There is however a high number of SNPs between the majority of datasets, indicating that the isolates do not seem to be closely genetically related to each other.

4.3.3. Transitions and Transversions

A greater number of transitions (39,003) than transversions (8,094) were observed with a total transition/transversion (Ti/Tv) ratio of 4.8188. The trend in this case indicates a higher abundance of G>A and C>T transitions than the opposite direction for this transition (table 4.5). This is consistent with the Ti/Tv ratio of the previous study in chapter 3 (Zander et al., 2013).

Table 4.5: Details for Transitions and Transversions.

Table 3 SNP details for all SNPs detected			
SNP type	**Base change**	**SNP count[a]**	**Percent[b]**
Transition	A > G	8,557	18.17
Transition	G > A	10,795	22.92
Transition	C > T	10,954	23.26
Transition	T > C	8,697	18.47
Transversion	A > C	824	1.75
Transversion	C > A	1,077	2.29
Transversion	A > T	1,045	2.22
Transversion	T > A	1,096	2.33
Transversion	C > G	1,081	2.30
Transversion	G > C	1,012	2.15
Transversion	G > T	1,101	2.34
Transversion	T > G	858	1.82

[a] Number of each SNP type predicted
[b] Percent value of overall SNP

4.3.4. Overall SNP Effects

The overall effect of the predicted SNPs within the *L. maculans* genome was determined using SnpEff. This can be seen in table 4.6.

Table 4.6: SNP Effects by Type represented as a heat map (The red background indicates more changes, the green less changes).

Type		
Type (alphabetical order)	**Count**	**Percent**
Downstream	37,446	35.17%
Intergenic	26,363	24.76%
Intron	1,376	1.29%
Non-synonymous Coding	2,696	2.53%
Non-synonymous START	2	0.00%
Splice site acceptor	13	0.01%
Splice site donor	8	0.01%
START lost	10	0.01%
STOP gained	75	0.07%
STOP lost	14	0.01%
Synonymous Coding	2,213	2.08%
Synonymous STOP	11	0.01%
Upstream	36,257	34.05%

The regions directly upstream and downstream of coding sequence i.e. genic regions, contain the overwhelming majority of SNPs (69.2%). Upstream and Downstream regions were designated as 5,000 bp up or downstream of the gene annotation. Intergenic SNPs account for about a quarter of total SNPs. This means only 6.03% of SNPs are found within coding regions, of which 1.29% are within introns. The categories with the lowest number of changes were the non-synonymous STARTs and Splice site acceptors. There were 1,948 gene annotations with at least one non-synonymous SNP change within coding regions. The genes with at least five non-synonymous SNPs (29 gene annotations) within coding regions are included in table 4.7.

The non-synonymous coding, start lost, stop gained and stop lost changes are the most likely to have deleterious effects on protein translation and function. These account for 2.62% of SNPs predicted.

Table 4.7: Gene annotations with over five non-synonymous SNPs within coding regions.

Gene Annotation	Putative Function	Super Contig	Length (GENE)	Bases affected (EXON)	Count (NON SYNONYMOUS CODING)	Bases affected (SPLICE SITE DONOR)	Count (STOP GAINED)	Count (SYNONYMOUS CODING)
Lema_G103060.1	-	17	5048	19	13	0	4	2
Lema_G121880.1	-	24	2645	17	11	0	1	5
Lema_G082060.1	similar to AflR	13	1317	19	10	0	4	5
Lema_G056140.1	-	5	2015	10	8	1	0	2
Lema_G043030.1	-	4	2123	11	7	0	1	3
Lema_G081650.1	-	13	2457	16	7	0	0	9
Lema_G082050.1	-	13	3345	17	7	0	2	8
Lema_G104210.1	-	16	3476	10	7	0	0	3
Lema_G104300.1	-	16	2514	7	7	0	0	0
Lema_G106780.1	-	18	3485	10	7	0	0	3
Lema_G108890.1	-	18	2373	15	7	0	0	8
Lema_G109740.1	-	18	2809	7	7	0	0	0
Lema_G002260.1	-	0	2715	9	6	0	1	2
Lema_G009140.1	-	0	5306	10	6	0	0	4
Lema_G012010.1	-	0	6276	7	6	0	0	1
Lema_G029930.1	similar to TPR domain protein	2	1749	13	6	0	0	7
Lema_G041660.1	dicer-like protein 2	4	4581	8	6	0	0	2
Lema_G082090.1	similar to fatty acid synthase beta subunit	13	5834	13	6	0	0	7
Lema_G113900.1	-	20	9885	7	6	0	0	1
Lema_G121460.1	-	24	2184	7	6	0	0	1
Lema_G014740.1	-	1	6913	9	5	0	1	3
Lema_G036000.1	-	3	1482	6	5	0	0	1

Lema_G038060.1	-	3	1688	6	5	0	0	1
Lema_G049600.1	-	6	3273	10	5	0	1	4
Lema_G053740.1	-	5	2217	6	5	0	0	1
Lema_G056490.1	-	8	2997	6	5	0	0	1
Lema_G072900.1	-	10	7356	8	5	0	0	3
Lema_G081980.1	-	13	1633	13	5	0	4	4
Lema_G090610.1	-	15	2346	5	5	0	0	0

A total of 29 genes showed higher than five non-synonymous SNP changes within coding regions. SuperContig 13 had the highest number of affected genes with five (17.2%). All but three of the listed genes (89.6%) also contained synonymous SNP changes within coding regions and 31% gained a stop codon. The number of bases affected within exons is determined by non-synonymous coding, synonymous coding and splice site donor SNPs, along with stops gained within coding regions.

The majority of the gene annotations (86.2%) revealed no gene function information. The four putative functional annotations were similar to AflR, similar to TPR domain protein, dicer-like protein 2 and similar to fatty acid synthase beta subunit. The AflR protein in *Aspergillus* species is involved in the activation of aflatoxin-related genes which are implicated in plant infection (Lee et al., 2006). Aflatoxins are a mycotoxin that can also affect animal and human health (Liu and Chu, 1998). The tetratricopeptide repeat (TPR) domain is a structural motif that is present in a wide range of proteins, often forming scaffolds that mediate protein-protein interactions (Blatch and Lassle, 1999). The dicer-like proteins are involved in regulation and initiation of defence responses towards viruses and transposons (Kadotani et al., 2004). The fatty acid synthase beta subunit has the biological functions of oxidation reduction processes and fatty acid biosynthesis in fungi (Jenni et al., 2007).

The SNP variation and distribution within the genome can also be seen in figure 4.3. There is a clear trend that shows a significantly higher number of SNPs within Upstream, Downstream and Intergenic regions compared to coding regions.

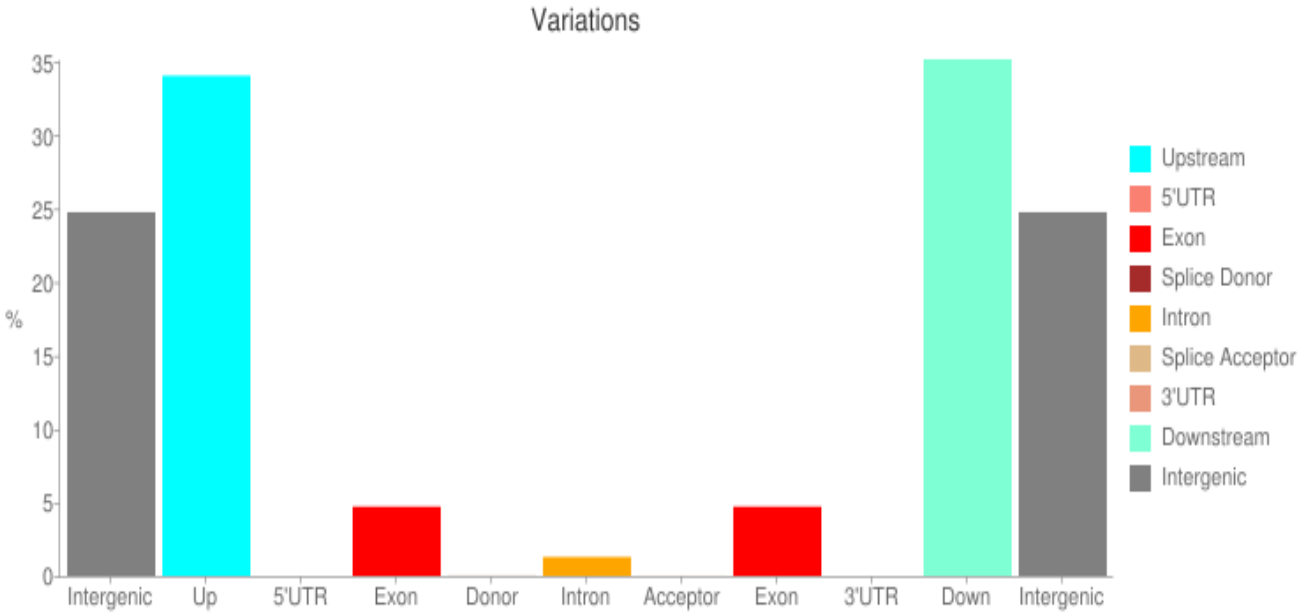

Figure 4.3: Distribution of SNPs within the *L. maculans* genome. Percentage of SNPs within different genomic locations.

The exons contain only 4.72% of the SNP effects, with splice site SNPs (splice site acceptors and donors) only accounting for 0.02% of SNPs combined.

4.3.5. SNPs in AvrLm Genes

A number of SNPs were identified in the regions surrounding the characterised *AvrLm* genes, namely *AvrLm1, AvrLm2, AvrLm4-7, AvrLm6, AvrLmJ1* and *AvrLm11*. Of these six genes, all contained at least one SNP within an exon, except *AvrLm6,* which has been shown to be present/absent in avirulent/virulent *L. maculans* isolates. The SNP alleles for the avirulence genes in each differential isolate can be seen in table 4.8.

Missing values for *AvrLm4-7* indicates high sequence divergence and double virulence towards *Rlm4* and *Rlm7*. The *AvrLm4-7* SNP (position 1,374,707), highlighted in green, is the SNP resulting in virulence toward *Rlm4* plant resistance. This confirms the differential table, as isolates D8 and D9 are virulent for *AvrLm4* but avirulent for *AvrLm7*. Isolates D4,

D5 and D12 are the only isolates avirulent at the *AvrLm4* locus (table 4.1). The result for isolate D6 suggest this is the wrong prediction in this case, as only one sequence read was seen mapping to that locus.

Table 4.8: Avirulence gene SNPs within differential isolates. SNPs highlighted in green are known to affect avirulence/virulence in these avirulence genes. Grey boxes indicate missing values.

***AvrLm* Gene**	***AvrLm1***	***AvrLm2***	***AvrLm2***	***AvrLmJ1***	***AvrLmJ1***	***AvrLm11***	***AvrLm4-7***	***AvrLm4-7***	***AvrLm4-7***	***AvrLm4-7***	***AvrLm4-7***
SNP Allele	A/T	A/C/G	A/G	C/G/T	A/G	C/T	A/G	A/G	C/T	C/G	A/T
SuperContig	6	6	6	7	7	22	12	12	12	12	12
SC Position	1,607,437	1,887,678	1,887,679	983,151	983,230	294,318	1,374,328	1,374,508	1,374,512	1,374,707	1,375,102
Ref	T	A	A	T	G	C	G	G	C	C	A
Avriulent	T	G	G	T	G	N/A	G	G	C	C	A
D1	-	A	A	C	G	C	-	-	-	-	-
D2	-	C	A	T	G	C	-	-	-	-	-
D3	-	A	A	G	G	T	A	A	T	-	-
D4	-	C	A	C	G	C	G	G	C	C	A
D5	A	G	G	C	A	C	G	G	C	C	A
D6	T	A	A	T	G	C	-	A	T	C	-
D8	-	C	A	G	G	C	G	G	C	G	T
D9	-	C	A	T	G	C	G	G	C	G	A
D10	-	A	A	C	G	C	-	-	-	-	-
D12	-	C	A	T	G	C	G	G	C	C	A
SNP score	37	21	19	159	58	27	8	2	2	176	61

Alleles that match differential isolate table

Alleles not matching differential data

Alleles different from reference *AvrLm* gene

The SNP in *AvrLm1* confirms the differential data available to us (table 4.1). The only isolates known to contain *AvrLm1* are D5 and D6, which are the only ones to have sequence coverage for this gene (the rest have no coverage, seen in the grey shading). D5 contains the A allele, which varies from the avirulence allele of T in both the reference and in isolate D6. The effect of this amino acid change, from an isoleucine to a lysine, is unknown.

A novel *AvrLm11* SNP was predicted in isolate D3, however its effect is unknown as *AvrLm11* is believed to be made virulent by the loss of the minichromosome, on which it resides (Balesdent et al., 2013). As we have no access to *Rlm11* cultivars, screening for the *AvrLm11* avirulence gene has not been carried out for the differential isolates, and therefore no conclusions can be drawn from this data at the moment. The *AvrLmJ1* SNPs at positions 983,151 and 983,230 have been previously described (Van de Wouw et al., 2014).

The *AvrLm2* results confirm the differential table (table 4.1.) for these isolates, with the exception of D1, which showed the virulent allele for these SNPs (see figure 4.4). The two SNPs predicted in this analysis confirm the results of Ghanbarnia et al. (2014), which had previously described these as being causative SNPs in *AvrLm2* virulence/avirulence.

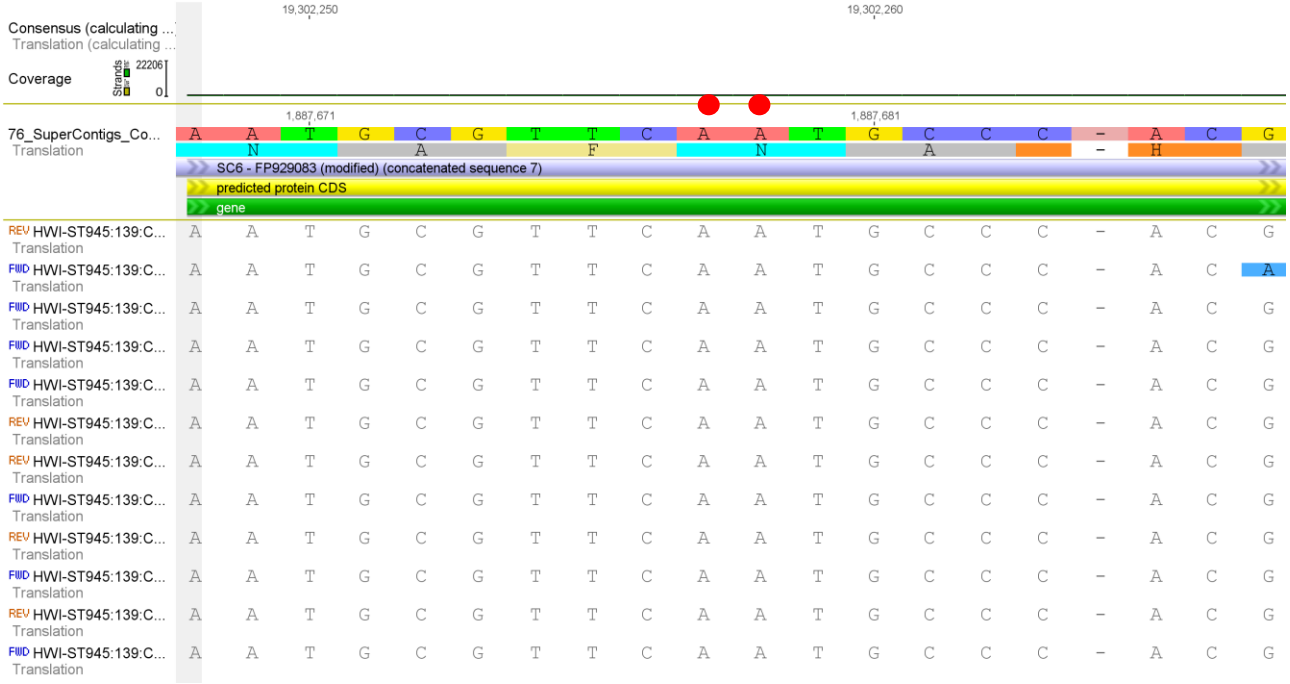

Figure 4.4: D1 sequence reads mapped to *AvrLm2.*

The SNP values for isolate D1 at the *AvrLm2* loci (positions 1,887,678 and 1,887,679) were confirmed by viewing the annotated sequence alignment file. Here D1 sequence reads are mapped to the reference genome. The SNPs are highlighted with red circles. There are many reads mapping in this position confirming the result.

4.3.6. SNPs within Small Secreted Proteins (SSPs)

A number of SNPs were identified within small secreted proteins (SSPs) of *L. maculans*. These were chosen from supplementary data provided by Rouxel et al. (2011) and only those SSPs with both transcriptomic support and overexpression at 7 or 14 dpi (days post inoculation) were selected (supplementary table 4, Rouxel et al. (2011)). This gives confidence that these proteins are being expressed upon infection of the host. Of 116 gene annotations that fit these criteria, 19 SSPs showed SNPs within exons in the differential isolates. Of these, three were the known avirulence genes *AvrLm2, AvrLm4-7* and *AvrLm11*, which had been analysed in the avirulence gene SNP search. The remaining 16 SSPs were named SSP1 – SSP16 and are presented in table 4.9.

The SNP alleles for the SSP genes in the differential isolates are presented in table 4.10. Two thirds (66.6%) of the SNPs within these genes were found to be transitions, common among avirulence gene SNPs. Of 21 SNPs, 9 were A/G transitions and 5 were C/T transitions.

Confidence for correct SNP prediction is high, as the SNPs within SSP1, SSP2 and SSP3 all seem to be inherited in haplotype blocks, with isolates containing one SNP always containing the others. Interestingly, the SSP1 gene seems to match the *AvrLm5* profile based on the differential table (table 4.1), although further work needs to be undertaken to confirm this.

Table 4.9: SSP1 – SSP16 and avirulence genes with genome positions and annotations.

Name	Annotation	SC number	Position Start	Position Stop
AvrLm1	Lema_G049660.1	6	1,607,018	1,607,681
AvrLm2	Lema_G049950.1	6	1,887,282	1,887,980
AvrLm4-7	Lema_G086290.1	12	1,374,257	1,375,182
AvrLm6	Lema_G049940.1	6	1,839,097	1,839,677
AvrLm11	Lema_G119060.1	22	294,086	294,450
AvrLmJ1	Lema_G070880.1	7	982,889	983,113
		Revised Gene Annotation		983,314
SSP1	Lema_G003630.1	0	1,237,353	1,237,934
SSP2	Lema_G013380.1	1	390,339	391,007
SSP3	Lema_G119130.1	22	436,526	436,999
SSP4	Lema_G119620.1	23	109,110	110,169
SSP5	Lema_G021120.1	1	2,476,727	2,477,421
SSP6	Lema_G028710.1	2	1,708,344	1,708,944
SSP7	Lema_G038290.1	4	69,860	70,849
SSP8	Lema_G043210.1	4	1,412,535	1,413,274
SSP9	Lema_G065940.1	9	1,242,629	1,243,051
SSP10	Lema_G074890.1	10	857,938	859,079
SSP11	Lema_G075980.1	10	1,151,853	1,152,706
SSP12	Lema_G083940.1	12	646,591	646,977
SSP13	Lema_G093990.1	14	91,631	92,329
SSP14	Lema_G103890.1	16	426,650	427,405
SSP15	Lema_G106060.1	16	1,224,896	1,225,660
SSP16	Lema_G113200.1	20	228,262	228,840

Revised Gene Annotation refers to the re-annotation of this gene as *AvrLmJ1* by Van de Wouw et al. (2014), which found a START codon upstream of the original annotation from Rouxel et al. (2011).

Table 4.10: SNP alleles for SSP genes in differential isolates.

SSP Name	SuperContig	Position	Allele	Ref	D1	D2	D3	D4	D5	D6	D8	D9	D10	D12	SNP score
SSP1	0	1,237,478	A/C	C	C	C	C	C	A	C	C	C	C	C	5
		1,237,484	C/T	C	C	C	C	C	T	C	C	C	C	C	3
		1,237,608	A/G	A	A	A	A	A	G	A	A	A	A	A	71
SSP2	1	390,613	G/T	G	T	T	G	T	G	T	T	G	G	T	45
		390,614	A/G	A	G	G	A	G	A	G	G	A	A	G	44
SSP3	22	436,548	A/G	G	A	A	G	A	G	A	G	G	G	G	297
		436,769	C/T	T	C	C	T	C	T	C	T	T	T	T	275
SSP4	23	109,914	A/G	G	G	G	G	G	A	G	G	G	G	G	71
		109,943	C/T	C	C	C	C	T	C	C	C	C	C	C	80
SSP5	1	2,476,807	C/T	C	C	T	C	C	C	C	C	C	C	C	43
SSP6	2	1,708,676	A/G	A	A	A	A	G	A	A	A	A	A	A	69
SSP7	4	70,340	G/T	T	T	T	T	G	G	T	T	T	T	T	132
SSP8	4	1,412,682	A/G	G	A	G	G	G	A	G	G	G	G	G	273
SSP9	9	1,242,947	A/G	G	G	G	G	G	G	G	A	G	G	G	91
SSP10	10	858,918	C/T	T	T	C	C	C	T	C	C	C	T	C	255
SSP11	10	1,152,614	A/C	C	C	C	C	A	C	C	C	C	C	C	70
SSP12	12	646,911	A/G	G	G	G	G	G	G	G	A	G	G	G	93
SSP13	14	91,847	A/G	A	G	G	G	G	G	G	A	G	A	A	186
SSP14	16	426,850	C/G	G	C	C	G	G	G	G	G	G	G	G	176
SSP15	16	1,225,394	A/T	A	A	A	T	T	T	T	T	A	A	T	198
SSP16	20	228,795	C/G	G	C	G	G	C	C	C	C	G	G	G	259

Ref refers to the reference genome allele

Cells are coloured to highlight the differences between isolates at these alleles

4.4. Discussion

The results of this SNP analysis reveal a high degree of polymorphisms within the *L. maculans* genome. These SNPs provide a valuable marker resource that can be used to understand overall genomic diversity, as well as targeted analysis for genes of interest.

4.4.1. SNP Numbers

The sequencing outputs, ranging from 26.7x to 266.3x coverage across the 10 datasets, gave confidence that the SNP prediction would be true. These coverage values are sufficient for consistent read coverage across the genome and for SNP prediction. Furthermore, high coverage allows confidence in the SNP prediction, as SGSautoSNP disregards SNPs that contain differing sequence within a single isolate.

The SGSAutoSNP method revealed 47,097 SNPs between the 10 isolates. This is more than double the number of predicted SNPs described in the previous study (Chapter 3, Zander et al. (2013)), which revealed 21,814 SNPs between two *L. maculans* isolates. This is to be expected, as there are 10 isolates which are being compared to each other, as opposed to only 2. This highlights the ascertainment bias which occurs when only 2 samples are utilised for SNP discovery.

4.4.2. SNP Numbers between Isolates

The SNP numbers between isolates was generally quite consistent with an average of almost 10,000 SNPs between isolates (table 4.4). D10 SNP numbers were significantly lower than those of the other isolates. This can be explained by the fact that D10 has a much lower sequence coverage than the nine other isolates, which can affect the number of SNPs called.

The D6 dataset also showed slightly lower SNP numbers than the average. This dataset showed signs of contamination, as both the quality and the mapping of sequence reads to the reference genome was lower than the other isolates (data not shown).

The highest number of SNPs predicted, were between isolates D1 and D5 with 11,752 SNPs between them. This indicates that they have a high amount of sequence divergence between them, which could be seen in the phylogenetic tree in chapter 3 (figure 3.1, D1

being isolate 14 and D5 being isolate 18 in that analysis). Overall the number of SNPs between isolates was high, indicating a high amount of genetic diversity between isolates.

4.4.3. Change rates across the genome

SNP numbers across the SCs was generally quite even, with an average of one SNP every 953 bp, as seen in table 4.3. This indicates that a high number of polymorphisms exist across the entire *L. maculans* genome, with a fairly even spread. Examples of the SNP frequency in other fungi are *Saccharomyces cerevisiae* with 1 SNP every 357 bp (Schacherer et al., 2009) and *Fusarium graminearum* with one SNP every 3,449 bp (Cuomo et al., 2007). It must be said however, that these numbers inevitably depend on the quality and reliability of the SNP prediction.

The exception to this is SC30 which only contains a single SNP across its entire length of 154,863 bp. This SC contains the mitochondrial DNA of *L. maculans*, which is known to be generally highly conserved between species (Kocher et al., 1989). Another possible explanation could be the presence of more than one copy of mitochondrial DNA in each isolate, known as heteroplasmy (Lesemann et al., 2006), which was discussed in chapter 3. This may be likely, as there are even fewer SNPs within SC30 in this chapter when 10 isolates are compared, which has many times the overall coverage and 8 more isolates than the previous analysis in chapter 3.

A comparison to the change rates in chapter 3 indicates a similar spread of change rates across the genome. The exception is SC37 which has a change rate almost 30 times lower than that observed in chapter 3. This SC contains unassembled rDNA sequences which are affected by RIP mutations (Rouxel et al., 2011). It is likely that the isolates 21 and 41, described in chapter 3, have considerably less diversity within this genomic region, than across all 10 isolates, and therefore less SNPs are present. The genomes of these two isolates may have closer ancestral linkage to each other, and therefore exhibit higher sequence similarity.

4.4.4. Transition/Transversion Ratio

The transition/transversion (Ti/Tv) ratio of the predicted SNPs is ~4.82, which is slightly higher than the Ti/Tv ratio of ~4.2, observed in chapter 3. The effect of RIP mutations, which results in higher AT content, may be the cause of the higher than average Ti/Tv ratio

in *L. maculans*. Ti/Tv ratios differ greatly in fungi, with values ranging from 5.1 in *Tricholoma matsutake* (Xu et al., 2007), to 3.47 in *Candida albicans* (Forche et al., 2004).

4.4.5. SNP Effects

Non coding genomic regions contained the majority of the SNPs predicted in this analysis with almost 70% of SNPs being located 5,000 bp up or downstream of a gene annotation (see table 4.6 and figure 4.3). Intergenic SNPs account for about a quarter of total SNPs. Only a small percentage of SNPs are located within coding regions (~6%) and of these, only 2.62% are predicted to have strong deleterious effects such as SNPs present in non-synonymous coding, start lost, stop gained and stop lost categories. This is expected, as these regions are generally well conserved to avoid changes in protein translation and negative effects on protein function.

Almost 2,000 gene annotations showed at least one non-synonymous SNP change within coding regions. A selection of these can be viewed in table 4.7., which lists those gene annotations with over five non-synonymous SNPs within coding regions. The majority of these genes also contained synonymous SNP changes and a number of gained stop codons was also observed, indicating that these genes seem to be highly variable across the differential isolates. Interestingly, five (17.2%) of the affected genes are located on a ~60,000 bp region of SuperContig 13 (3.65% of contig length). This area seems to exhibit high diversity in the differential isolates and contains several secondary metabolite genes involved in plant infection processes, such as a gene similar to AflR (mycotoxin activation) and another with similarity to the beta subunit of fatty acid synthase. In *Aspergillus* fungal species, the fatty acid synthases produce a toxic secondary metabolite related to aflatoxin (regulated by AflR), which affects infection processes in this organism (Brown et al., 1996).

Two other genes with functional annotations were affected by non-synonymous SNP changes, namely one containing a TPR domain and another with similarity to dicer-like protein 2. The dicer-like proteins regulate and initiate defence responses toward viruses and transposons (Kadotani et al., 2004). *L. maculans* contains a high number of transposable elements so it may be the case that these genes are under selection pressure within the genome. The majority of genes (86.2%) in table 4.7., do not have putative functional annotations which makes determining the impact of SNP changes difficult.

4.4.6. *AvrLm* SNP Effects

Adding confidence to the SNP prediction are the SNPs called within the known *AvrLm* genes (table 4.6). This method correctly predicted two *AvrLm2* SNPs described by Ghanbarnia et al. (2014). These two polymorphic sites, G^{397} -> A/C^{397} (SNP1 + 2) and G^{398} -> A^{398} (SNP3), were the only SNPs that were responsible for loss of *Rlm2*-mediated recognition specificity (Ghanbarnia et al., 2014). Both lead to a change in the Gly^{133} amino acid, which results in changes to Ser^{133}, Arg^{133} and Asp^{133}. All the differential isolates confirmed their previous annotation as virulent/avirulent except isolate D1. It has been thought to be avirulent, however in this case the virulent alleles were identified (figure 4.4). The D1 isolate may need to be re-screened for the *AvrLm2* gene to confirm the results from this analysis, as the high sequence read coverage for this gene gives a high level of confidence that the SNP prediction is true.

A single SNP was identified within the coding sequence of the *AvrLm11* gene. Isolate D3 showed a based change from C to T at position 294,318, resulting in a synonymous change from Leu to Leu. The *AvrLm11* SNP was predicted in isolate D3, however its effect is unknown as *AvrLm11* is believed to be made virulent by the loss of the minichromosome on which it resides (Balesdent et al., 2013). Interestingly, the *AvrLm11* gene has so far been present in all sequenced isolates, with no loss of the minichromosome detected.

The *AvrLmJ1* gene, described by Van de Wouw et al. (2014) also contains SNPs responsible for altered virulence phenotypes towards *B. juncea* varieties. This method successfully identified the SNP at R^{29} (Base 87, position 983,230) which causes a premature STOP codon within the *AvrLmJ1* gene and results in virulence towards *B. juncea* (Van de Wouw et al., 2014). Another predicted SNP at position 983,151 (also discussed by Van de Wouw et al. (2014)), shows no correlation to virulence towards *B. juncea* and does not correspond to any other avirulence genes, based on the differential table (table 4.1).Interestingly, the SNP at position 983,230 matches *AvrLm5* based on the phenotype of the differential isolates. Indeed, Van de Wouw et al. (2014) hypothesised that the *AvrLmJ1* and *AvrLm5* genes may be one and the same. The same isolates as those in the Van de Wouw et al. (2014) study were used in this SNP prediction. SGSautoSNP thus confirms the results of that study and indicates that SNP prediction is accurate. Further analysis should be undertaken to surmise whether *AvrLm5* and *AvrLmJ1* are indeed the same gene.

Of the four SNPs predicted within the genic region of *AvrLm4-7*, only one was in the exon region. Previously described by Parlange et al. (2009), this SNP causes a base change from G^{358} to C^{358}, corresponding to the amino acid change from G^{120} to R^{120}. This confirms the findings of Parlange et al. (2009), which revealed that the G^{358} position is responsible for *Rlm4*-meditated recognition in the host.

Isolates D8 and D9 both have the virulent allele at the G^{358} position, as they are both avirulent towards *Rlm7* but virulent towards *Rlm4*. Isolates D4, D5 and D12 are all avirulent towards *Rlm4* and *Rlm7,* and the SNP results confirm this and the accuracy of the differential isolate data. The only outlier is isolate D6 which only had a single read that mapped to this position, likely resulting from sequence error or mismapping of the read. Given the high coverage across the genome, a large number of reads would be expected to map to this gene if it was a true SNP prediction.

The SNP data for *AvrLm4-7* in isolates D1, D2, and D10 are missing, suggesting that the gene sequence may be highly diverged in these isolates, which prevents read mapping. If a read contains more than 2 SNPs compared to the reference genome, the SOAP program prevents mapping and thus SNP calling in these instances. In the case of zero coverage for the *AvrLm4-7* gene, a complete or partial deletion of the *AvrLm4-7* locus may be the cause, leading to virulence towards both the *Rlm4* and *Rlm7* resistance genes. This was seen for the isolates D1, D2, and D10, with no SNPs called within the *AvrLm4-7* gene, confirming their double virulence towards *Rlm4* and *Rlm7,* which can be seen in table 4.1. The D3 isolate contained some SNPs, however missing values were seen in the *AvrLm4-7* gene sequence, also confirming its double virulence (table 4.1). These results seem to be consistent with the study by Parlange et al. (2009), which hypothesised that virulence toward *Rlm7* was caused by partial or complete deletion of the *AvrLm4-7* gene.

There was no sequence coverage for isolates D1, D2, D3, D4, D8, D9, D10 and D12 at the *AvrLm1* locus, which is consistent with *AvrLm1* being absent in these isolates, based on previous studies and testing of these differential isolates (table 4.1). The isolates known to contain *AvrLm1*, D5 and D6, were the only ones to have sequence coverage for this gene (confirming table 4.1), although D5 contains the A allele, which varies from the avirulence allele of T in both the reference and in isolate D6. The effect of this amino acid change, from an isoleucine to a lysine, is unknown, however the mechanism of virulence towards *Rlm1* has been found to be a complete loss of the gene (Fudal et al., 2009).

4.4.7. SSP Genes - Putative *AvrLm* genes

The SSP genes provide a potential resource for putative avirulence gene identification. They represent avirulence gene candidates because they are small, secreted proteins (a defining characteristic of avirulence genes in *L. maculans* so far), have transcriptomic support and are overexpressed upon infection (supplementary table 4, Rouxel et al. (2011)). Interestingly, two thirds of the SNPs within the SSP genes are transitions, which is a common SNP change caused by RIP mutations. Indeed, transitions are very common in avirulence genes, with 8/11 (~73%) of avirulence gene SNPs in table 4.8. being transitions. Giving more confidence to the SNP prediction is the observation that SNPs within SSP1, SSP2 and SSP3 all seem to be inherited in haplotype blocks, with isolates containing one SNP always containing the others. By comparing the differential table with the SNP profiles of the SSP genes, it was hoped that some of the SSP genes would match one of the avirulence genes. The only SSP gene that matches an avirulence gene profile is the SSP1 gene which matches the *AvrLm5* profile. However this needs to be further evaluated in order to confirm these results, by designing infection assay experiments, such as those using tobacco infiltration for transient expression of this gene (Sparkes et al., 2006). This is complicated by the fact that the *Brassica Rlm5* gene has not been cloned so far.

4.5. Conclusions

A high number of SNPs were predicted using NGS data in this study. These SNPs provide and vast and valuable marker resource for further study of diversity and genes of interest in *L. maculans,* such as those involved in plant infection processes and secondary metabolite synthesis. The SNP prediction in this analysis, showed success in predicting the correct *AvrLm* gene complement in most differential isolates, as well as identifying a number of novel SNPs in these genes. Further work on the SSP genes should be conducted in order to test these genes as avirulence gene targets.

5. Identification of Presence/Absence Variations (PAVs) from *Leptosphaeria maculans* NGS Data

5.1. Introduction

Single nucleotide polymorphisms (SNPs) have become increasingly popular for analysing genetic variations between individuals of the same species or family (Duran et al., 2009b; Edwards et al., 2012; Rafalski, 2002). For example as described in this thesis, Zander et al. (2013) developed a set of SNP molecular markers with which to study *L. maculans* genetic diversity and population structure (Patel et al., 2015). Growing evidence suggests that along with SNPs larger, structural variations such as presence/absence variants (PAVs) are causative of phenotypic variation. This has been shown in a number of plant species, including rice (Kumar et al., 2007), maize (Belo et al., 2010), soybean (Lam et al., 2010) and *Arabidopsis thaliana* (Tan et al., 2012), as well as in bacteria (Francis and Tanaka, 2012). A gene ontology analysis of PAVs in *A. thaliana*, found that a large number of present/absent genes were involved in stress response and a significant portion of these were disease resistance genes (Shen et al., 2006). PAVs have also been seen in fungi, such as deletion of the *AvrLm1* and *AvrLm6* loci in *L. maculans* leading to virulence (Raman et al., 2012; Van de Wouw et al., 2010). A study of the rice blast fungus *Magnaporthe oryzae*, found that avirulence genes showed a high prevalence of presence/absence polymorphisms and were generally associated with diverse repeat sequences (Huang et al., 2014). The relatively high rate of nonsynonymous base changes in repeat-affected regions, which leads to alteration of gene function, inadvertently results in the pathogen evading host immune responses upon infection.

The Presence Absence Variation (PAV) prediction method used here has the goal of identifying regions of genomic sequence that are differentially present between datasets. It employs a Kmer based approach that can identify regions that are lost/missing between different genomes, or highly diverse in their sequence. This method uses smaller portions

of all the available reads (Kmers) to compare these to all the other smaller portions of the data set, which requires significantly less processing power than comparing whole sequence reads to each other. The whole sequence read that contains the Kmer is then pulled into a new file that then contains complete PAV reads, not just the Kmer sequences alone.

Current approaches used to detect PAVs require a reference genome to which sequence reads can be mapped-to and compared (Tan et al., 2012; Wang et al., 2014). The novel approach presented here, termed Present/Absent K-mer Analysis Pipeline (PAKAP), has the added benefit of predicting these polymorphisms without the need for a reference genome. This allows for the use of this method in species which do not yet have an available reference genome and importantly to study sequences that are not present in the reference, such as genes that are lost or highly variable between individuals.

The *L. maculans* genome reveals some interesting characteristics of this fungal pathogen. The genome contains a large number of repeat sequences that make up one third of the genome (Rouxel et al., 2011). These repeats cause a segmentation of the genome which can be seen in the existence of isochores. Isochores are defined as long stretches of DNA which are homogenous in GC content, with abrupt changes in GC content from one segment to another (Parlange et al., 2009). Segments of low GC content (AT-rich) generally carry an abundance of repeat sequences, whereas higher GC content segments include most of the gene sequences.

Repeat-induced point mutation (RIP) is a type of pre-meiotic repeat inactivation mechanism that causes substitutions from C to T and G to A and has played a role in *L. maculans* genome amassing a large proportion of repeats (Idnurm and Howlett, 2003). This RIP mechanism is postulated to intermittently cause SNP mutations in *Avr* genes, as these are often located in regions of RIP activity. These polymorphisms may then be responsible for altered *Avr* gene sequences that can evade recognition by *R* genes. The *AvrLm1*, *AvrLm2* and *AvrLm6* genes are positioned within a RIP affected region of the genome and the *AvrLm1, AvrLm4-7* and *AvrLm6* genes were physically mapped to the *L. maculans* genome by Rouxel et al. (2011) after being identified previously through map-based cloning (Fudal et al., 2007; Gout et al., 2006; Parlange et al., 2009). Ghanbarnia et al. (2014) more recently characterised *AvrLm2* and confirmed its position in the same

genomic region as *AvrLm1* and *AvrLm6*, as previous mapping studies had suggested (Balesdent et al., 2002). *AvrLm11* was found to be located on a conditionally dispensable chromosome (CDC) with a high enrichment of transposable elements (Balesdent et al., 2013). *AvrLmJ1* is involved in avirulence towards *Brassica juncea* cultivars, but displays similar characteristics as the other avirulence genes in that it encodes a small secreted protein (SSP) and is cysteine rich (Van de Wouw et al., 2014), indeed all the avirulence genes identified to date, display features typical of effector genes in that they are expressed early in infection and secrete small proteins into the apoplast (Stergiopoulos and de Wit, 2009). Previous insights have shown that the *AvrLm1* and *AvrLm6* avirulence genes can be present/absent in *L. maculans* isolates, making this the likely cause of virulence, since recognition of the avirulence gene cannot occur (Raman et al., 2012; Van de Wouw et al., 2010; Zander et al., 2013). A recent analysis of *L. maculans* sequence data confirms these findings (Golicz et al., 2015). In contrast, the *AvrLm4-7* gene is known to become virulent through SNPs in the coding sequence (Parlange et al., 2009). Similarly, *AvrLm2* also evades recognition by host resistance genes through three identified SNPs which coincide with the virulent phenotype (Ghanbarnia et al., 2014).

The Presence/Absence method was first implemented to predict Presence/Absence Variations (PAVs) between two *L. maculans* isolates D1 and D5. A further analysis describes a larger study of PAVs between 10 isolates, and how this relates to avirulence genes and other disease related genes; in particular the significance in understanding the genome of *L. maculans*.

The SNP prediction in chapters 3 and 4 indicates that considerable sequence differences exist between the differential isolates. This chapter expands on the insights from the earlier chapters to consider presence/absence variations (PAVs) in *L. maculans* isolates in order to understand the distribution and frequency of PAVs in more depth, as well as their effect on the genome.

5.2. Materials and Methods

5.2.1. Isolates Used in Analyses

The isolates used in this study were selected from a group of differential *L. maculans* isolates from a collection used to determine *Rlm* gene presence in canola crop species, as described in section 4.2.1.

The isolates used can be seen in table 5.1. Avirulence/virulence information for *AvrLm11* is not known for these isolates and has therefore been excluded from the table. The JN3 Reference details are included as a guide in determining *Avrlm* gene presence/absence in the *L. maculans* genome.

5.2.2. DNA Extraction Quality Control of DNA

DNA extractions for the fungal isolates used in this study were performed according to methods described in section 2.4. Quality control of DNA samples was performed as described in section 2.5.

5.2.3. Sequence Library Preparation and Genomic Sequencing

5.2.3.1. PAV Validation

Paired-end sequencing libraries, with a 300-500 bp insert size, were prepared from the *L. maculans* DNA using the TruSeq Library Preparation Kit, following the manufacturer's instructions (Illumina), as described in section 2.7.1. Isolates D1 and D5 were sequenced using the Illumina GAIIx platform, with D1 being sequenced using 200 cycle reagents (resulting in 100 bp PE sequence reads) and D5 using 300 cycle reagents (resulting in 150 bp PE sequence reads).

5.2.3.2. Ten Differential Isolates

The paired-end libraries were made using either the TruSeq or TruSeq Nano Library Preparation (Prep) Kits as described in section 4.2.3.

5.2.4. Reference Genome

The reference genome used in this study is described in section 2.7.2.

Table 5.1: *L. maculans* isolates used in this analysis.

Diff number	Other names	*AvrLm* Gene											
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Sylvestris**	**Hyola50**	**Mustang**
D1	IBCN13	Vir	Avr	Vir	Vir	Avr	Avr	Vir	Vir	Avr	Avr	Avr	Avr
D2	IBCN15	Vir	Vir	Vir	Vir	Avr	Avr	Vir	Avr	Vir	Avr	Avr	Avr
D3	IBCN16	Vir	Vir	Vir	Vir	Avr	Avr	Vir	Avr	Vir	Vir	Avr	Vir
D4	IBCN17	Vir	Vir	Vir	Avr	Avr	Avr	Avr	Avr	Vir	Avr	Avr	Avr
D5	IBCN18, M1	Avr	Avr	Vir	Avr	Vir	Vir	Avr	Vir	Vir	Avr	Avr	Avr
D6	IBNC75	Avr	Vir	Vir	Vir	Avr	Avr	Vir	Avr	Vir	Avr	Avr	Avr
D8		Vir	Vir	Vir	Vir	Avr	Vir	Avr	nd	Vir	Vir	Avr	Vir
D9		Vir	Vir	Vir	Vir	Avr	Avr	Avr	nd	Vir	Vir	Avr	Avr
D10	PHW1223	Vir	Vir	Vir	Vir	Avr	Avr	Vir	Avr	Avr	Avr	Avr	Avr
D12	04MGPS012	Vir	Vir	Vir	Avr	nd	Vir	Avr	nd	Vir	Vir	nd	Avr
JN3 Reference	v23.1.3	Avr	Vir	Vir	Avr	Avr	Avr	Avr	Avr	Vir	nd	nd	nd

Avr means the isolate is avirulent for the corresponding resistance gene e.g. *AvrLm1* can be recognised by *Rlm1*

Vir means the isolate is virulent and cannot be recognised by the corresponding resistance gene, thus causing infection

Nd means there is no data for this resistance gene in this isolate

Hyola50 and ***Mustang*** are Australian canola varieties with to date unidentified resistance

5.2.5. PAV Prediction

The Present/Absent K-mer Analysis Pipeline (PAKAP) prediction method was developed by Dr. Kenneth Chan. It employs a K-mer based approach that compares each dataset to another dataset, by using small subsets of the reads as opposed to using whole reads. The sequence read containing the K-mer sequence is then catalogued in a new file as a PAV read, along with all the other PAV reads from that comparison (e.g. PAV reads present in D1 and not in D5 (D1vD5)). No reference genome is required to predict unique reads (PAVs).

5.2.6. Present/Absent K-mer Analysis Pipeline

The PAKAP method consists of three different steps:

1. Identify an optimal K-mer size.

2. Identify present/absent K-mers (PAKs).

3. Identify present/absent reads (PARs).

Jellyfish (Marcais and Kingsford, 2011) is an efficient K-mer counting program and is used to perform K-mer counting in PAKAP (from K=5 to K=22). In the first step, PAKAP runs Jellyfish for a user specified range of K-mers to obtain their K-mer occurrences. A K-mer uniqueness graph is plotted using the K-mer occurrence and an optimal K-mer size is determined by finding the “knee point” in the K-mer uniqueness graph (Kurtz et al., 2008). The "knee-point" of the plot, which balances the specificity and sensitivity of the information content, is identified as the optimal K-mer size. Each K-mer in the optimal K-mer size is tested for PAK with the following criterion:

K-mer=PAKs1 if (Os2 = 0 and Os1 ≥ min_occ)

K-mer=PAKs2 if (Os1 = 0 and Os2 ≥ min_occ)

Where PAKs1 and PAKs2 are the K-mers only presented in sample 1 (s1) and sample 2 (s2) respectively. Os1 and Os2 are the number of K-mer occurrences in s1 and s2 respectively. The min_occ is the user defined minimum K-mer occurrence.

Finally all the raw reads in each sample are tested for present/absent reads (PARs). A PAR is defined as when all the constituent K-mer in a read are PAKs in the corresponding sample. Two sets of PARs are outputted corresponding to reads only presented in sample 1 and sample 2 respectively.

5.2.7. Primer Design

Primer design was performed in order to amplify the PAV regions using PCR for validation. The primers used are described in table 5.2. Seven primer pairs were developed for each data set (to amplify those present in D1 and not D5 and vice versa), to amplify products across the *L. maculans* genome.

Table 5.2: Primers used for PCR amplification of PAV regions.

D1 Primers	Forward Primer	Reverse Primer	Super Contig	Position	~ Product Size (bp)
1	GCCAACATATTCGACGAC	CGTTGGTCATTGCAGTG	1	837,969	200
2	CACTTATCGCCTATTGTCACC	TGGATTGCCGTGTGG	1	1,931,457	200
3	AAGGGCTTCGAGGGC	CACCGCTGTATGTGCG	2	2,370,017	400
4	CTCAAATGTGCCCAGTTG	GAGAGCATCAATTTCGGAAC	3	1,423,008	320
5	CTTCTTGCAGACCCGC	GGTCGCTATGCCGATC	4	1,429,026	480
6	CCTACACTTCAACCTCAGCG	ATTTCTCACATGGCTGCG	15	555,196	820
7	GCACTGGCAACTTGGTATC	GCAAGAGCAAGAACCTGG	17	974,232	400
D5 Primers	**Forward Primer**	**Reverse Primer**	**Super Contig**	**Position**	**~ Product Size (bp)**
1	CGTCTATGCGGAAAGGG	GACTCCCACCAGCCC	24	218,584	300
2	AGCGATCAGAGGCAGG	TGGATAGCGGAGATGAGC	1	839,252	300
3	GCTATTTAATTTAGGAGCTTCTTGC	AGACACACATTAAGGACACTTAAACTC	1	2,095,182	570
4	GTGGACCTCAAATGTGCC	GGGATGACATCAAGTCCG	3	1,423,002	300
5	TGTCGCAGTCTTGCTTATTG	CACGACTACAGCTCAGCTTC	5	1,060,990	250
6	ATTCTTGGCGAGCGAC	ACAACGAAACCGAGCG	14	154,665	450
7	CCTGCTTGTTTCTGCTACG	ACGATACGGCAGCGTACTAC	15	555,439	200

5.2.8. PCR validation of PAVs

A total of 14 PAV regions were randomly selected from the *L. maculans* genome, on different SuperContigs of the v23.1.3 reference genome (Rouxel et al., 2011). These had a range of redundancy and co-segregation scores. Primers for the amplification are described in table 5.2. Genomic DNA was isolated from the two isolates D1 and D5 as described in section 2.4.1.

PCR amplification of the 14 loci in both isolates was performed using primers designed to conserved sequence flanking the PAV regions (table 5.2) in a 25 µL reaction volume containing 1 x iTaq PCR buffer (containing 100 mM Tris–HCl and 500 mM KCl, pH 8.3) (Bio-Rad, Hercules, CA), 200 µM each dNTP (Bio-Rad), 0.5 µM each primer, 1.5 U iTaq DNA polymerase (Bio-Rad), RNase- and DNase-free water (Gibco; Life Technologies, Carlsbad, CA) and 5 ng DNA. Thermocycling conditions for the reaction were 94°C for 2 min, followed by 35 cycles of 94°C for 30 s, annealing for 40 seconds at 58°C and extension for 1 min at 72°C. Final extension was performed at 72°C for 10 min. Gel electrophoresis to resolve the PCR products was performed using a 1% (w⁄v) agarose gel in 1 x TAE buffer (Sambrook et al., 2006) containing ethidium bromide resolved products, as described in section 2.5.1. The Australian Genome Research Facility's (AGRF) PD+ service was used to purify and subsequently sequence the PCR products. The purified PCR products were Sanger-sequenced using Big-Dye 3.1 (PerkinElmer, Waltham, MA), using forward and reverse PCR primers, and analysed using an ABI3730xl (see section 2.6.3). The sequences for each locus and isolate were aligned and compared using Geneious Pro v6.1.7 (Kearse et al., 2012) with a cost matrix of 65%, a gap open penalty of 12 and a gap extension penalty of 3, and each of the PAVs was assessed.

5.2.9. Trimming of Sequence Reads

The D5 sequence reads described in 5.2.3.1. and used for PAV validation were trimmed from 150 bp to 100 bp to ensure they matched the D1 dataset which had a read length of 100 bp.

The sequence reads from differential isolate D10 were trimmed in order to improve PAV prediction as the D10 reads exhibited low quality scores at the beginning and ends of the reads, a common occurrence in longer reads such as 250 bp. D10 reads were trimmed to 100 bp in length using a simple perl script to remove the first 10 and the last 140 bases of

each read. The length of 100 bp was chosen in order to match with the other datasets used which were 100 bp.

5.2.10. Sequence Read Mapping

Sequence reads for each PAV comparison were mapped to the *L. maculans* ('brassicae') isolate v23.1.3 reference genome (Rouxel et al., 2011) using SOAPaligner 2 (Li et al., 2009a). The SOAP –r 2 parameter was used to allow a sequence read to align to the genome multiple times, as well as the –M 0 parameter, allowing no mismatches to the reference genome. The –l parameter was set as 100 which means the reads used can be no more than 100 bp long.

5.2.11. Gene Annotation Hits

The reads that mapped to annotated genes from the reference genome (Rouxel et al., 2011) were identified during the SOAP mapping (section 5.2.10) by using the annotated reference file i.e. the gff3 file. The details of the gene annotation hits were determined by compiling the hits outside genes, hits at the start or end of genes and genes within the gene sequence.

5.3. PAV Method Validation Results

5.3.1. Sequencing results

The D1 sequencing library yielded 59,493,424 PE reads of 100 bp length. The insert size was ~200 bp, with a coverage of 266.3x. The D5 sequencing library yielded 1,387,906 PE reads of 150 bp length, with an insert size ~250 bp and a coverage of 9.3x. The overall coverage for D1 was over 26 times greater than for isolate D5, and even greater considering the further trimming of D5 reads from 150 to 100 bp (see section 5.2.9).

5.3.1. Number of PAVs Predicted between D1 and D5

Using a k-mer size of 18, a total of 44,180 (0.00074% of the whole dataset) and 9,469 (0.0068% of the whole dataset) PARs (Presence/Absence Reads) were found in D1 and D5, respectively.

5.3.2. PCR Validation Results

The results of the PCR amplification of the PAV regions are presented in table 5.3. Products were amplified and subsequently sequenced using Sanger sequencing. The results revealed a large number of nucleotide changes between isolates D1 and D5 within 8 of the products, ranging from a single SNP (D1 product number 7) to 149 nucleotide changes in D1 product number 6.

The D5 primer product 7 resulted in amplification of a product for both D1 and D5, however subsequent sequencing revealed very poor sequence quality for D1 PCR products. It was later realised that the D5 primer product 7 amplifies the same product as D1, primer pair 6. The sequence for this region was highly divergent, based on read mapping information. As the same product (with different primers) was validated and contained a high number of polymorphisms, therefore this failed PCR product was not included in the final count.

The product from D5 primer pair 5 showed high quality amplification with no sequence divergence observed, however it seems to be false positive result based on a survey of D1 sequence read coverage (figures 5.5 and 5.6).

Table 5.3: Blackleg Presence/Absence Variation (PAV) PCR Validation Results.

D1 Primers	~ Product Size	D1 Product	D5 Product	Product Size Analysed	Total Nucleotide Changes	SNPs	2 Adjacent SNPs	>3 Adjacent SNPs	Indels 1-3 bp	Indels > 3 bp	Comment
1	250 bp	Present	Present	225 bp	29	23	3	-	-	-	D5 product highly diverged
2	200 bp	Present	Absent	-	-	-	-	-	-	-	Validated
3	400	Present	Absent	-	-	-	-	-	-	-	Validated
4	500	Present	Present	449 bp	88	43	11	6	1	-	D5 product highly diverged
5	480	Present	Absent	-	-	-	-	-	-	-	Validated
6	800	Present	Present	761 bp	149	107	16	3	-	-	D5 product highly diverged
7	400	Present	Present	357 bp	1	1	-	-	-	-	D5 product diverged

D5 Primers	~ Product Size	D1 Product	D5 Product	Product Size Analysed	Total Nucleotide Changes	SNPs	2 Adjacent SNPs	>3 Adjacent SNPs	Indels 1-3 bp	Indels > 3 bp	Comment
1	270	Present	Present	240 bp	28	20	4	-	1	-	D1 product highly diverged
2	300	Present	Present	282 bp	66	31	4	5	6	2	D1 product highly diverged
3	570	Present	Present	531 bp	18	15	3	-	-	-	D1 product highly diverged
4	490	Present	Present	452 bp	89	43	11	6	1	-	D1 product highly diverged
5	250	Present	Present	253 bp	0	-	-	-	-	-	False Positive
6	450	Absent	Present	-	-	-	-	-	-	-	Validated
7	200	Present	Present	196 bp	0	-	-	-	-	-	Bad sequence quality in D1 isolate

The pipeline succeeded in predicting the correct Present/Absent allele in 12 of the 13 selected *L. maculans* genomic regions, which results in an accuracy of 92.3% for prediction (table 5.3). Of the 13 primer pairs, all amplified the correct product, as alignment to the genome showed. Four of the products showed a simple Presence/Absence phenotype with the “present” product being amplified and the “absent” product not being amplified as predicted. Eight further products amplified in both D1 and D5 but showed significant sequence divergence between the isolates with an average of 58.5 nucleotide differences between the D1 and D5 PCR products. Alignment of the sequences of both isolates for each product also showed a high incidence of Multiple Nucleotide Polymorphisms (MNPs) with an average of 6.5 Double Nucleotide Polymorphisms (DNPs) per product and a total of 20 MNPs of 3 nucleotides or more.

5.3.3. Alignment of PAV regions

Sequenced PCR products were aligned in Geneious (see section 2.8.2.) and analysed for polymorphisms within regions of high sequence quality. Only if the polymorphism was present in both forward and reverse product in both isolates was it scored as a change. Figures 5.1 – 5.6 describe a few examples of sequence alignments from the PCR validation.

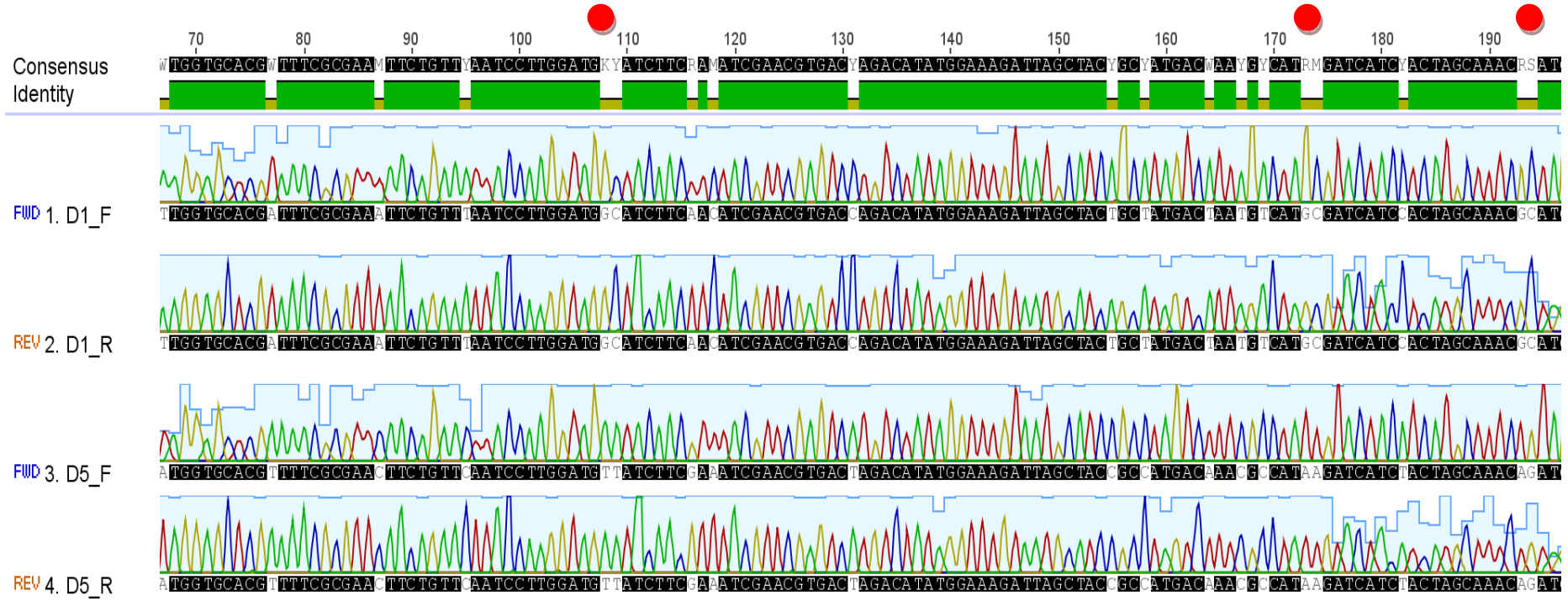

Figure 5.1: Alignment of PCR products for D1, primer set 1.

There were 29 nucleotide differences observed between the D1 products and D5 products (not all shown in figure 5.1) of which a number can be viewed here. Particularly interesting are the DNPs at positions 108-109, 173-174 and 193-194 (highlighted by the red circles). D1_F denotes the sequence product using the forward primer and D1_R denotes sequence product using the reverse primer (see table 5.2).

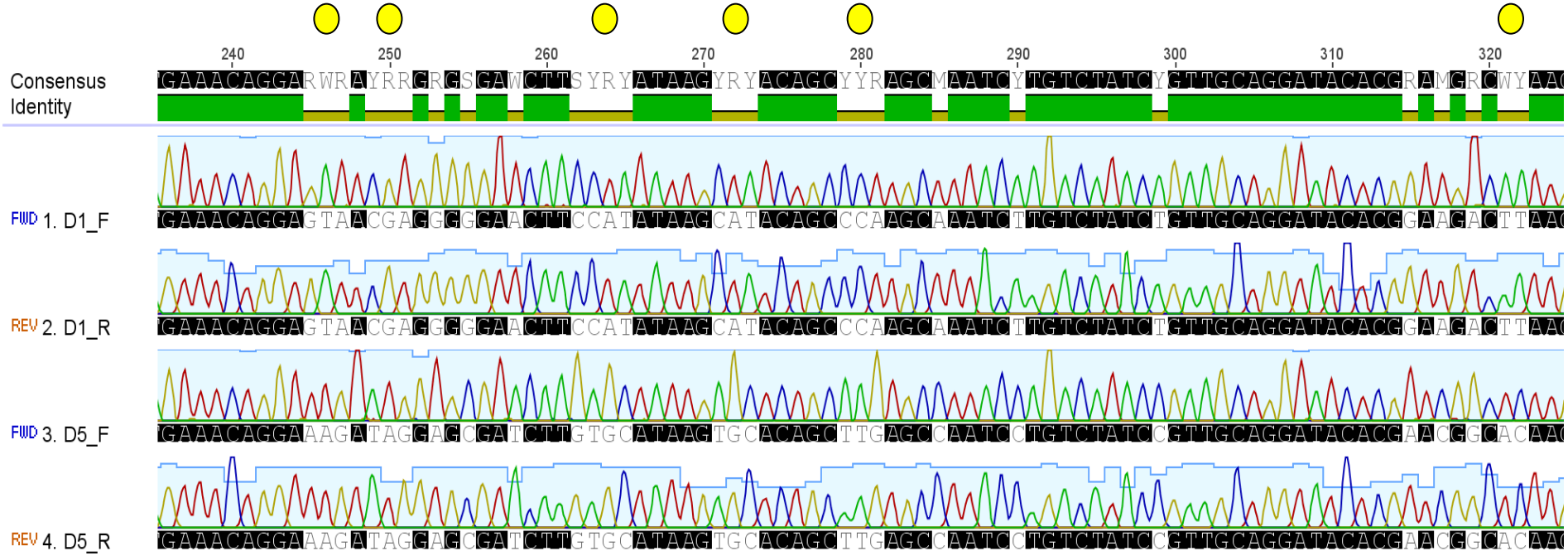

Figure 5.2: PCR product alignment for D1, primer set 4.

There were 88 nucleotide differences observed between the D1 products and D5 products (not all shown in figure 5.1) of which a number can be viewed here. This product had a high number of Multiple Nucleotide Polymorphisms (MNPs) (highlighted by the yellow circles). D1_F denotes the sequence product using the forward primer and D1_R denotes sequence product using the reverse primer (see table 5.2).

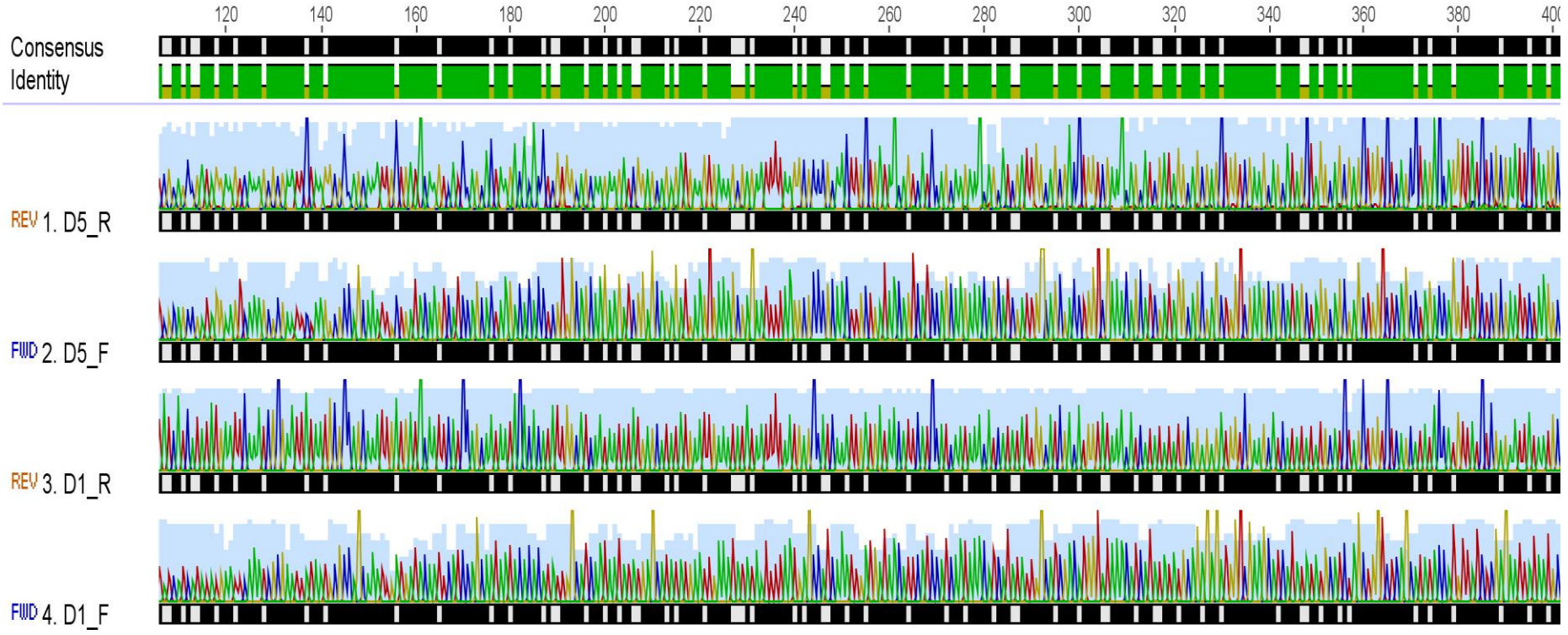

Figure 5.3: PCR product alignment for D1, primer set 6.

There were 149 nucleotide differences observed between the D1 products and D5 products (not all shown in figure 5.1) of which a number can be viewed here. This comparison had the highest number of polymorphisms. D1_F denotes the sequence product using the forward primer and D1_R denotes sequence product using the reverse primer (see table 5.2)

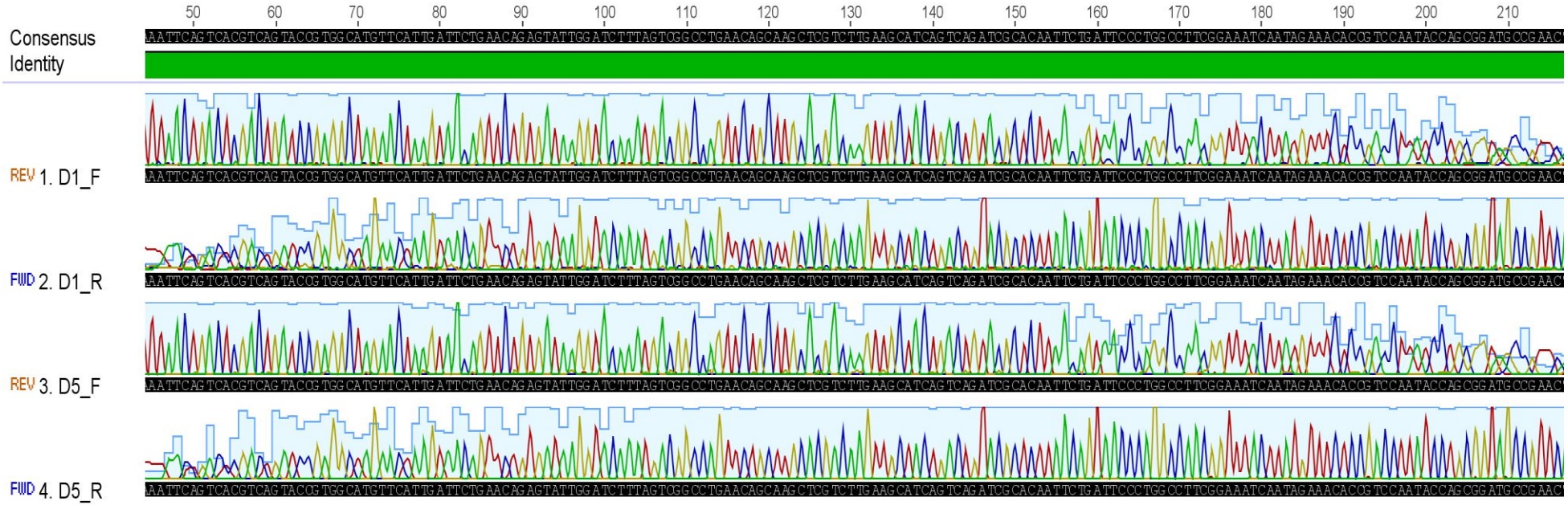

Figure 5.4: PCR product alignment for D5, primer 5.

No nucleotide changes were observed between D1 PCR products (D1_F and D1_R) and D5 products (D5_F and D5_R). The PCR primers amplified a portion of the LEMA_P053310.1 gene, from positions 1,060,990 to 1,061,243 on SC5.

Figure 5.5: D1 read coverage for LEMA_P053310.1 gene (highlighted by red arrow), product D5, primer 5. The gene is 9,671 bp long so only 3' and 5' ends of the alignment are shown.

The D5 primer 5 product amplifies a portion of this gene. The PCR amplified a product in isolate D1, however whole genome sequence read alignment, suggests that this gene is indeed lost in D1, as no read coverage can be observed, except two single reads in a 200 bp section of the gene (see figure 5.6; not in PCR area). Sequence coverage either side of this gene is above 60x coverage.

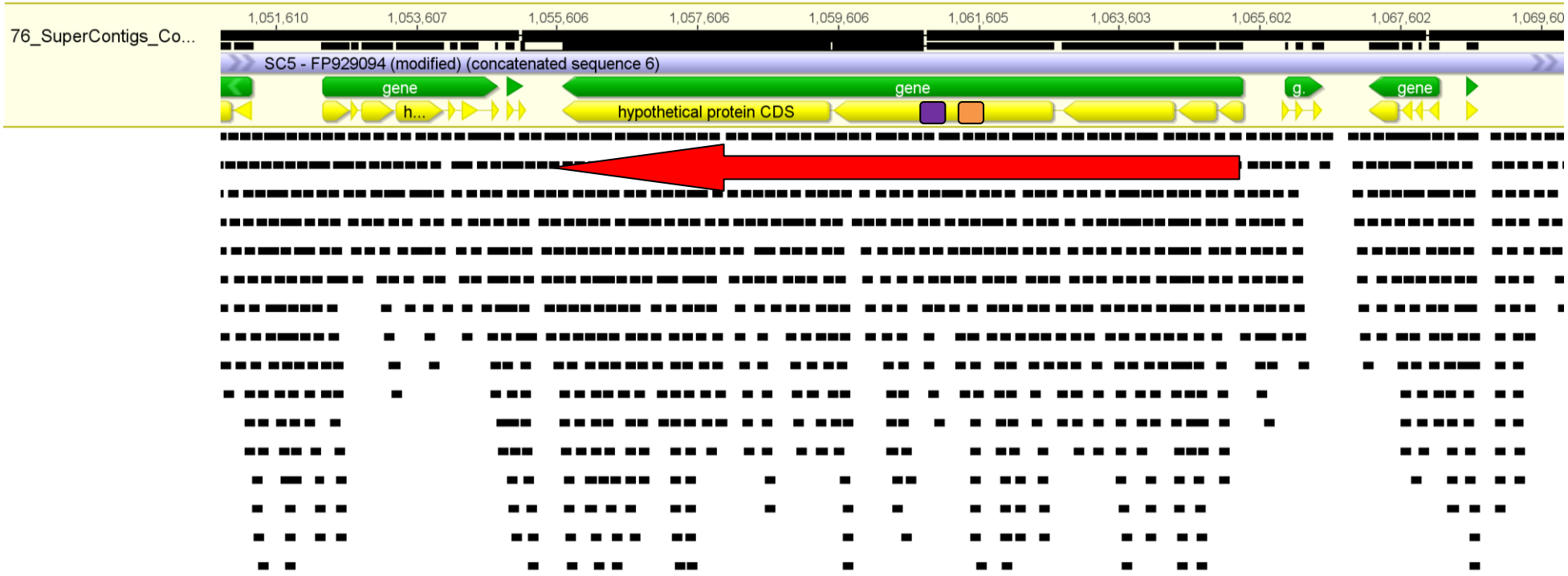

Figure 5.6: D5 read coverage for LEMA_P053310.1 gene (highlighted by red arrow), product D5, primer 5. The purple box indicates the PCR product region. The orange box indicates the area of the gene that two D1 reads mapped to (positions 1,061,362 – 1,061,565; not in PCR area). The primers were not within the region that the reads mapped to (primer and PCR region positions 1,060,990 – 1,061,243). The D5 isolate has over 10x coverage across the D5 primer 5 product, indicating it is indeed present in D5.

5.4. Differential Isolates Results

5.4.1. Sequencing

The Illumina whole genome sequencing results for all differential isolates can be seen in table 5.4. Considerable coverage was achieved with this method, ranging from 26.7x (isolate D10) to 266.3x (isolate D1), with an average coverage of 99.65x. The read length for the isolates was 100 bp, apart from D10 for which 250 bp read length was used, although these were trimmed to 100 bp for the analysis (resulting is 10.65x coverage). The same sample for the D5 isolate was re-sequenced in order to obtain higher coverage (101x coverage compared to a coverage of 9.3x in section 5.3.1).

Table 5.4: Whole genome sequencing output.

Isolate Name	Total PE reads	Read Length (in bp)	~ Insert size (in bp)	Total # bp	~ Coverage
D1	59,493,424	100	200	12,017,671,648	266.3x
D2	11,492,603	100	250	2,320,000,000	51.4x
D3	17,778,999	100	580	3,555,799,800	78.8x
D4	22,401,329	100	500	4,480,265,800	99.2x
D5	22,935,182	100	500	4,587,036,400	101x
D6	19,967,228	100	500	3,993,445,600	88.5x
D8	28,463,112	100	500	5,692,622,400	126x
D9	16,293,997	100	530	3,258,799,400	72.2x
D10	2,402,994	250	550	1,206,302,988	26.7x
D12	19,507,705	100	520	3,901,541,000	86.4x

PE reads are 'Paired-End' sequence reads, the output from the Illumina sequencing platforms

Insert size is the size of the fragment used for sequencing

bp stands for 'Base-pair'

Total # bp is the total number of reads (double the PE number) multiplied by the read length

Coverage; the 'x' refers to the 'times' coverage, which is how many times the total bp output can cover the 45.12 Mbp *L. maculans* genome

5.4.2. Optimal K-mer Size

The optimal K-mer size for each dataset comparison ranged from 16-17 bp in all results, which can be seen in table 5.5.

Table 5.5: Optimal K-mer size for each dataset comparison.

Comparison	Optimal k-mer Size (in bp)	Comparison	Optimal k-mer Size (in bp)
D1-D2	17	D3-D12	16
D1-D3	17	D4-D5	16
D1-D4	17	D4-D6	17
D1-D5	17	D4-D8	16
D1-D6	17	D4-D9	16
D1-D8	17	D4-D10	16
D1-D9	17	D4-D12	16
D1-D10	17	D5-D6	17
D1-D12	17	D5-D8	16
D2-D3	16	D5-D9	16
D2-D4	16	D5-D10	16
D2-D5	16	D5-D12	16
D2-D6	16	D6-D8	17
D2-D8	16	D6-D9	16
D2-D9	16	D6-D10	16
D2-D10	16	D6-D12	17
D2-D12	16	D8-D9	16
D3-D4	16	D8-D10	16
D3-D5	16	D8-D12	16
D3-D6	16	D9-D10	16
D3-D8	16	D9-D12	16
D3-D9	16	D10-D12	16
D3-D10	16		

The optimal K-mer size in this analysis was either 16 bp or 17 bp, depending on the dataset comparison. Thirteen comparison displayed 17 bp as the ideal K-mer size, while the 32 other comparisons had an optimal K-mer size of 16 bp. The only consistent trend was that D1 comparisons all had the optimal K-mer size of 17 bp.

5.4.3. PAV Numbers

The PAV prediction resulted in a large number of presence/absence reads being identified between the isolates, which can be seen in table 5.6. These numbers give an indication of the variability between these datasets. Each cell denotes only the unique reads of each comparison (e.g. D1 and D2). The reference genome is not used in the prediction.

Table 5.6: Number of single reads in each pair-wise comparison between isolates. The number above the diagonal denotes reads unique to D1 compared to D2 (23,611), later described as D1vD2. The number is different between D1vD2 (23,611) and D2vD1 (1,772) because they can contain a different number of unique reads compared to each other. Coverage also play a role in this, as multiple reads mapping to the same unique region are counted individually, thus making the total number in the same region high.

	D1	**D2**	**D3**	**D4**	**D5**	**D6**	**D8**	**D9**	**D10**	**D12**
D1		23,611	19,951	16,978	32,298	28,964	16,075	20,262	59,037	33,083
D2	1,772		381	143	446	662	246	1,197	10,760	700
D3	794	471		83	247	274	66	212	4,979	412
D4	2,632	1,089	371		677	12,209	170	658	12,322	773
D5	1,653	189	302	241		6,001	30	467	3,956	537
D6	145,703	107,859	36,098	805,788	1,248,464		918,074	44,051	992,390	1,228,075
D8	26,197	1,142	423	276	274	219,597		1,058	144,345	383
D9	716	562	110	90	274	484	78		3,231	249
D10	152	54	20	20	23	30	0	49		36
D12	2,855	376	370	185	318	3,850	48	819	12,772	

5.4.4. PAV Read Mapping

The predicted PAV reads were mapped to the *L. maculans* genome using SOAPaligner (section 6.2.10). Of the 6,266,356 PAV reads identified in all comparisons, 129,742 reads mapped to the reference genome (~2% of PAV reads). The number of reads from each comparison that mapped to the reference genome can be seen in table 5.7.

Table 5.7: Number of PAV reads that mapped to the reference genome.

	D1	D2	D3	D4	D5	D6	D8	D9	D10	D12
D1		6,347	8,660	8,589	5,945	8,166	3,597	7,113	5,721	11,669
D2	1,264		182	96	130	26	16	384	2,537	133
D3	404	68		56	37	185	5	47	705	171
D4	1,429	134	248		2,555	27	369	961	4,062	250
D5	992	113	142	192		3,954	2	420	1,474	339
D6	1,533	140	207	4,127	4,386		1,596	568	3,412	2,427
D8	1,750	213	302	167	136	3,354		589	2,948	12
D9	280	102	88	48	64	162	10		335	249
D10	120	13	16	15	13	11	0	47		22
D12	1,511	139	252	136	193	1,535	34	532	3,084	

The isolate D1 showed a higher number of PAV predictions against all the other isolates (table 5.6) with an average of 27,807 PAVs predicted. D10 had by far the lowest number of predicted PAVs at just 42.7 PAVs on average. D10 also had zero PAVs compared to isolate D8. D6 seems to have the highest number of unique PAV reads with a staggering average number of PAVs of ~322,143. Because of the high number of PAVs from isolate D6, the average number of predicted PAVs per comparison is 69,626. Excluding D6 comparisons, the number is more than ten times lower at 6,497.

5.4.5. PAV Mapping Results

The mapped PAV read statistics can be viewed in table 5.8. This includes the total number of predicted PAV reads for each comparison, the number of reads that mapped to the reference genome and the percentage of total reads that mapped to the reference.

The overall percentage of PAV reads mapping to the reference is ~2%. The D10vD9 comparison PAV reads showed the highest percentage of mapped reads (47/49 reads, 95.9%) and also had 45 hits at the start or inside genic regions. The D6vD2 comparison had the lowest percentage of mapped reads with 140/107,859 reads mapping (0.1%). Indeed the nine D6 comparisons had the lowest mapping percentage of all datasets, ranging from 1.3 to 0.1%. Of the ten comparisons with the most gene hits, five of these involved the isolate D10, namely D2vD10, D6vD10, D12vD10, D4vD10 and D8vD10. Thirty-three of the PAV dataset comparisons showed mapping statistics higher than 50% (highlighted in green in table 5.8) to the *L. maculans* reference genome. Thirty-one PAV comparisons had at least 10 gene hits within the genome.

The PAKAP method did not identify any PAVs in the D10vD8 comparison. This is likely because D10 had very low coverage after trimming, so overall the prediction could not find any unique reads in D10 that were absent in D8.

Table 5.8: PAV Read Mapping Statistics.

Comparisons can be described as D1vD2, meaning PAV reads present in D1 and absent in D2 (or unique to D1 versus D2). D2vD1 would signify reads present in D2 and absent in D1. Cells highlighted in green represent comparisons with over 50% of PAV reads mapped to the *L. maculans* reference genome.

Comparison	Total reads	Mapped reads	% of Reads mapped	No. Genes Hit	Hit Within Gene	Hit at Start	Hit at End	Inter-genic Hits	Duplicate hits
D10vD1	152	120	78.9	5	93	0	5	22	1
D10vD12	36	22	61.1	2	15	2	0	5	0
D10vD2	54	13	24.1	3	11	0	0	2	0
D10vD3	20	16	80.0	2	4	0	0	12	0
D10vD4	20	15	75.0	3	5	0	0	10	0
D10vD5	23	13	56.5	1	3	0	0	10	1
D10vD6	30	11	36.7	0	0	0	0	11	0
D10vD8	-	-	-	-	-	-	-	-	-
D10vD9	49	47	95.9	6	44	1	0	2	0
D12vD1	2,855	1,511	52.9	19	1,090	30	23	368	28
D12vD10	12,772	3,084	24.1	27	1,475	237	112	1,260	150
D12vD2	376	139	37.0	3	73	0	0	66	0
D12vD3	370	252	68.1	8	103	9	0	140	0
D12vD4	185	136	73.5	6	72	0	0	64	1
D12vD5	318	193	60.7	9	54	11	0	128	1
D12vD6	3,850	1,535	39.9	8	249	146	122	1,018	75
D12vD8	48	34	70.8	4	18	9	0	7	0
D12vD9	819	532	65.0	10	325	51	0	156	1
D1vD10	59,037	5,721	9.7	16	2,355	481	294	2,591	889
D1vD12	33,083	11,669	35.3	20	6,638	300	686	4,045	1,530
D1vD2	23,611	6,347	26.9	9	3,461	38	649	2,199	990
D1vD3	19,951	8,660	43.4	20	4,396	443	451	3,370	888
D1vD4	16,978	8,589	50.6	19	4,306	117	547	3,619	1,003
D1vD5	32,298	5,945	18.4	11	2,045	249	511	3,140	541
D1vD6	28,964	8,166	28.2	11	4,197	400	250	3,319	958
D1vD8	16,075	3,597	22.4	8	2,080	167	316	1,034	387
D1vD9	20,262	7,113	35.1	14	3,299	305	172	3,337	830
D2vD1	1,772	1,264	71.3	24	873	53	48	290	64

D2vD10	10,760	2,537	23.6	31	1,381	135	152	869	225
D2vD12	700	133	19.0	6	63	5	8	57	2
D2vD3	381	182	47.8	9	92	9	0	81	4
D2vD4	143	96	67.1	6	42	0	0	54	0
D2vD5	446	130	29.1	8	34	6	8	82	1
D2vD6	662	26	3.9	1	4	0	0	22	0
D2vD8	246	16	6.5	3	10	5	0	1	0
D2vD9	1,197	384	32.1	13	258	27	6	93	15
D3vD1	794	404	50.9	8	292	9	23	80	2
D3vD10	4,979	705	14.2	15	208	34	56	407	17
D3vD12	412	171	41.5	2	66	9	0	96	1
D3vD2	471	68	14.4	4	49	13	0	6	0
D3vD4	83	56	67.5	3	8	0	0	48	0
D3vD5	247	37	15.0	1	24	0	0	13	0
D3vD6	274	185	67.5	0	0	0	0	185	0
D3vD8	66	5	7.6	2	5	0	0	0	0
D3vD9	212	47	22.2	1	0	0	11	36	0
D4vD1	2,632	1,429	54.3	17	975	87	86	281	36
D4vD10	12,322	4,062	33.0	25	2,444	266	131	1,221	262
D4vD12	773	250	32.3	4	115	13	0	122	2
D4vD2	1,089	134	12.3	3	54	0	0	80	0
D4vD3	371	248	66.8	7	92	10	0	146	1
D4vD5	677	233	34.4	11	83	12	0	138	1
D4vD6	12,209	2,555	20.9	9	1,251	237	92	975	182
D4vD8	170	27	15.9	4	16	11	0	0	0
D4vD9	658	369	56.1	9	223	37	0	109	0
D5vD1	1,653	992	60.0	11	718	10	20	244	12
D5vD10	3,956	1,474	37.3	13	592	70	106	706	58
D5vD12	537	339	63.1	5	160	26	0	153	6
D5vD2	189	113	59.8	2	28	0	0	85	0
D5vD3	302	142	47.0	4	88	0	0	54	0
D5vD4	241	192	79.7	7	88	0	0	104	1
D5vD6	6,001	3,954	65.9	13	1,413	216	229	2,096	251
D5vD8	30	2	6.7	1	2	0	0	0	0
D5vD9	467	420	89.9	6	288	10	17	105	7
D6vD1	145,703	1,533	1.1	21	1,142	72	57	262	30
D6vD10	992,39	3,412	0.3	30	1,930	250	112	1,120	209

	0								
D6vD12	1,228,075	2,427	0.2	16	1,297	186	105	839	151
D6vD2	107,859	140	0.1	5	117	5	0	18	0
D6vD3	36,098	207	0.6	10	98	13	0	96	0
D6vD4	805,788	4,127	0.5	20	2,597	202	162	1,166	300
D6vD5	1,248,464	4,386	0.4	28	2,199	375	328	1,484	274
D6vD8	918,074	1,596	0.2	13	765	135	172	524	97
D6vD9	44,051	568	1.3	13	418	25	0	125	5
D8vD1	26,197	1,750	6.7	14	1,175	99	102	374	47
D8vD10	144,345	2,948	2.0	21	1,596	201	126	1,025	141
D8vD12	383	224	58.5	3	157	12	0	55	22
D8vD2	1,142	213	18.7	3	74	17	0	122	0
D8vD3	423	302	71.4	8	129	15	0	158	0
D8vD4	276	167	60.5	5	78	0	0	89	0
D8vD5	274	136	49.6	6	41	12	0	83	0
D8vD6	219,597	3,354	1.5	9	1,609	90	112	1,543	327
D8vD9	1,058	589	55.7	11	423	18	16	132	4
D9vD1	716	280	39.1	6	239	7	13	21	1
D9vD10	3,231	335	10.4	9	125	17	30	163	10
D9vD12	249	112	45.0	4	72	7	0	33	1
D9vD2	562	102	18.1	4	49	0	0	53	1
D9vD3	110	88	80.0	2	29	0	0	59	1
D9vD4	90	48	53.3	4	20	0	0	28	0
D9vD5	274	64	23.4	1	16	0	0	48	0
D9vD6	484	162	33.5	1	9	0	0	153	0
D9vD8	78	10	12.8	2	10	0	0	0	0

5.4.6. PAV Reads with Gene Hits

The PAV reads that were mapped to the reference genome showed a high number of gene hits, i.e. PAV reads that mapped to annotated genes on the reference genome (see section 5.4.7.). The distribution of these gene hits from all comparisons across the first 25 SuperContigs (SC0 – SC24) can be seen in figure 5.7. No gene hits were recorded on the other 51 SuperContigs.

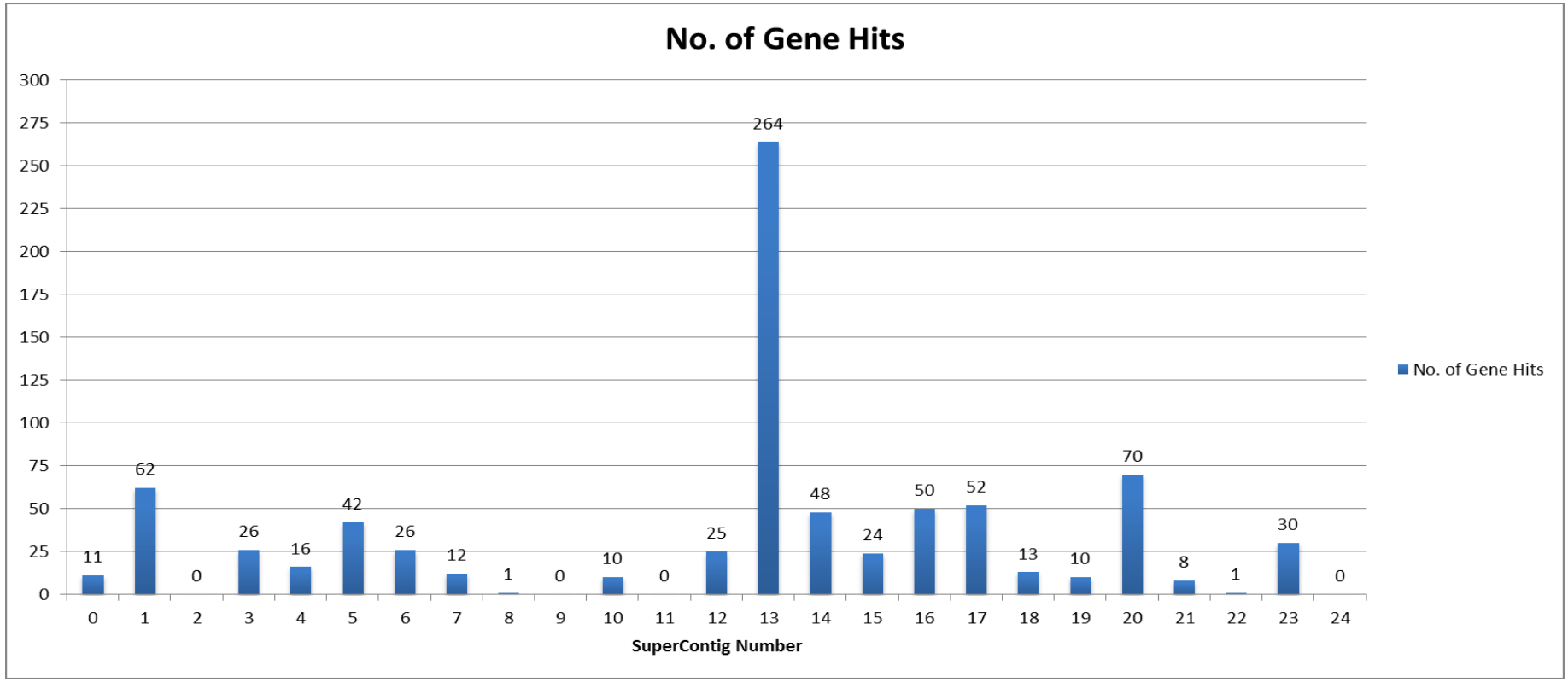

Figure 5.7: Number of gene hits per SuperContig (Values on the bars are the gene hit numbers for each SuperContig).

SuperContig 13 showed a significantly higher number of gene hits (nearly 10 times higher) than the other SCs with 264 gene hits, the next highest being SC20 with 70 gene hits. SCs 2, 9, 11 and 24 had no gene hits. Overall, the average number of gene hits per contig for SC1 – SC24 was 32.

5.4.7. PAV reads mapping to Gene Annotations

This table lists every gene annotation on the reference genome that PAV reads mapped to. This was conducted separately for every comparison e.g. D1vD2 and D2vD1 (table 5.9).

Start and **End** refer to the beginning and end of the annotation on the JN3 reference genome.

Green cells were highlighted to indicate the high incidence of PAV reads mapping to SuperContig 13.

Yellow cells highlight the annotation for the avirulence gene, *AvrLm6*.

Red cells highlight the annotation for the avirulence gene, *AvrLm1*.

Table 5.9: PAV reads that map to reference, with positions and annotation added.

Comparison	SuperContig No.	Position on Genome		Annotation	Strand (forward or reverse)
		Start	End		
D1vD2					
	4	1,428,187	1,430,545	Lema_G043290.1	-
	23	116,196	117,176	Lema_G119660.1	+
	16	519,580	522,064	Lema_G104290.1	-
	1	3,356,807	3,357,149	Lema_G023430.1	-
	16	514,658	516,407	Lema_G104270.1	-
	13	319,292	320,025	Lema_G078330.1	+
	23	112,460	115,068	Lema_G119640.1	+
	23	115,658	115,852	Lema_G119650.1	-
	16	513,425	514,524	Lema_G104260.1	+
D2vD1					
	13	1,426,780	1,428,165	Lema_G081940.1	+
	13	1,430,732	1,431,163	Lema_G081960.1	+

	13	1,428,596	1,429,536	Lema_G081950.1	-
	18	905,419	907,083	Lema_G109260.1	-
	3	1,668,837	1,669,337	Lema_G037120.1	-
	13	1,465,862	1,467,212	Lema_G082100.1	+
	1	844,294	846,759	Lema_G015110.1	+
	13	1,440,856	1,441,392	Lema_G081990.1	+
	14	152,839	155,417	Lema_G094000.1	-
	13	1,444,412	1,445,637	Lema_G082010.1	-
	13	1,451,035	1,454,379	Lema_G082050.1	-
	13	1,352,946	1,353,440	Lema_G01640.1	+
	14	158,218	167,699	Lema_G094020.1	+
	15	555,197	558,578	Lema_G091180.1	-
	14	1,292,449	1,292,811	Lema_G098090.1	+
	19	956,634	958,686	Lema_G113050.1	-
	13	1,441,845	1,442,615	Lema_G082000.1	+
	14	1,291,448	1,292,362	Lema_G098080.1	+
	14	157,792	158,001	Lema_G094010.1	+
	19	959,197	960,108	Lema_G113060.1	-
	5	1,055,685	1,065,356	Lema_G053310.1	-
	13	1,448,479	1,449,713	Lema_G082030.1	-
	0	763,884	765,737	Lema_G002200.1	+
	13	1,446,238	1,447,610	Lema_G082020.1	+
D1vD3					
	6	1,837,314	1,837,498	Lema_G049920.1	-
	20	535,205	538,183	Lema_G114290.1	-
	20	554,610	555,882	Lema_G114390.1	+
	23	116,196	117,176	Lema_G119660.1	+
	12	110,629	113,389	Lema_G082380.1	+
	1	1,931,202	1,931,996	Lema_G018900.1	+
	20	534,387	534,664	Lema_G114280.1	+
	12	109,077	109,426	Lema_G082370.1	+
	12	1,400,949	1,401,849	Lema_G086300.1	+
	1	3,356,807	3,357,149	Lema_G023430.1	-
	4	1,428,187	1,430,545	Lema_G043290.1	-
	20	529,112	533,279	Lema_G114270.1	-
	20	549,910	552,133	Lema_G114370.1	-

	23	112,460	115,068	Lema_G119640.1	+
	1	229,948	233,218	Lema_G013150.1	-
	12	1,402,878	1,403,303	Lema_G086310.1	+
	6	1,839,097	1,839,677	Lema_G049940.1	-
	6	1,837,589	1,837,925	Lema_G049930.1	-
	20	552,415	553,355	Lema_G114380.1	+
	22	105,263	106,588	Lema_G119050.1	-
D3vD1					
	14	157,792	158,001	Lema_G094010.1	+
	14	158,218	167,699	Lema_G094020.1	+
	18	905,419	907,083	Lema_G109260.1	-
	1	844,294	846,759	Lema_G015110.1	+
	15	555,197	558,578	Lema_G091180.1	-
	15	585,830	586,321	Lema_G091280.1	+
	0	763,884	765,737	Lema_G002200.1	+
	14	152,839	155,417	Lema_G094000.1	-
D1vD4					
	4	1,445,551	1,447,257	Lema_G043360.1	+
	18	347,790	348,175	Lema_G107350.1	-
	4	1,428,187	1,430,545	Lema_G043290.1	-
	5	415,359	416,426	Lema_G051020.1	-
	13	1,417,543	1,418,021	Lema_G081910.1	-
	1	1,931,202	1,931,996	Lema_G018900.1	+
	16	1,341,710	1,342,574	Lema_G106570.1	-
	7	142,280	143,248	Lema_G067720.1	-
	20	552,415	553,355	Lema_G114380.1	+
	20	535,205	538,183	Lema_G114290.1	-
	13	1,418,093	1,424,481	Lema_G081920.1	-
	18	31,059	32,963	Lema_G106610.1	+
	20	549,910	552,133	Lema_G114370.1	-
	21	263,332	263,889	Lema_G116710.1	-
	13	1,413,647	1,413,827	Lema_G081890.1	+
	20	554,610	555,882	Lema_G114390.1	+
	13	1,415,843	1,416,552	Lema_G081900.1	-
	20	534,387	534,664	Lema_G114280.1	+
	13	1,411,646	1,412,969	Lema_G081880.1	+

D4vD1					
	13	1,426,780	1,428,165	Lema_G081940.1	+
	13	1,441,845	1,442,615	Lema_G082000.1	+
	13	1,444,412	1,445,637	Lema_G082010.1	-
	13	1,430,732	1,431,163	Lema_G081960.1	+
	14	157,792	158,001	Lema_G094010.1	+
	3	1,668,837	1,669,337	Lema_G037120.1	-
	14	158,218	167,699	Lema_G094020.1	+
	13	1,428,596	1,429,536	Lema_G081950.1	-
	13	1,465,862	1,467,212	Lema_G082100.1	+
	18	905,419	907,083	Lema_G109260.1	-
	13	1,451,035	1,454,379	Lema_G082050.1	-
	15	555,197	558,578	Lema_G091180.1	-
	13	1,440,856	1,441,392	Lema_G081990.1	+
	0	763,884	765,737	Lema_G002200.1	+
	14	152,839	155,417	Lema_G094000.1	-
	13	1,446,238	1,447,610	Lema_G082020.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-
D1vD5					
	4	1,428,187	1,430,545	Lema_G043290.1	-
	1	1,722,362	1,724,272	Lema_G018140.1	-
	3	1,421,067	1,423,417	Lema_G036130.1	-
	16	519,580	522,064	Lema_G104290.1	-
	1	3,356,807	3,357,149	Lema_G023430.1	-
	6	1,837,314	1,837,498	Lema_G049920.1	-
	16	514,658	516,407	Lema_G104270.1	-
	16	513,425	514,524	Lema_G104260.1	+
	7	331,214	333,121	Lema_G068460.1	+
	6	1,837,589	1,837,925	Lema_G049930.1	-
	0	1,574,046	1,574,450	Lema_G004720.1	+
D5vD1					
	15	555,197	558,578	Lema_G091180.1	-
	14	1,291,448	1,292,362	Lema_G098080.1	+
	14	157,792	158,001	Lema_G094010.1	+
	5	1,055,685	1,065,356	Lema_G053310.1	-
	14	158,218	167,699	Lema_G094020.1	+

	1	844,294	846,759	Lema_G015110.1	+
	6	1,607,018	1,607,681	Lema_G049660.1	+
	18	905,419	907,083	Lema_G109260.1	-
	14	1,292,449	1,292,811	Lema_G098090.1	+
	0	763,884	765,737	Lema_G002200.1	+
	14	152,839	155,417	Lema_G094000.1	-
D1vD6					
	4	1,428,187	1,430,545	Lema_G043290.1	-
	15	500,031	500,476	Lema_G090930.1	-
	12	110,629	113,389	Lema_G082380.1	+
	7	142,280	143,248	Lema_G067720.1	-
	3	1,703,798	1,704,842	Lema_G037220.1	-
	16	1,341,710	1,342,574	Lema_G106570.1	-
	23	116,196	117,176	Lema_G119660.1	+
	23	112,460	115,068	Lema_G119640.1	+
	20	201,217	201,836	Lema_G113190.1	+
	5	490,652	491,090	Lema_G051290.1	-
	12	109,077	109,426	Lema_G082370.1	+
D6vD1					
	13	1,426,780	1,428,165	Lema_G081940.1	+
	13	1,430,732	1,431,163	Lema_G081960.1	+
	14	1,291,448	1,292,362	Lema_G098080.1	+
	13	1,444,412	1,445,637	Lema_G082010.1	-
	13	1,441,845	1,442,615	Lema_G082000.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	15	585,830	586,321	Lema_G091280.1	+
	14	157,792	158,001	Lema_G094010.1	+
	13	1,446,238	1,447,610	Lema_G082020.1	+
	14	158,218	167,699	Lema_G094020.1	+
	13	1,428,596	1,429,536	Lema_G081950.1	-
	13	1,465,862	1,467,212	Lema_G082100.1	+
	13	1,352,946	1,353,440	Lema_G081640.1	+
	5	1,055,685	1,065,356	Lema_G053310.1	-
	13	1,448,479	1,449,713	Lema_G082030.1	-
	6	1,607,018	1,607,681	Lema_G049660.1	+
	13	1,440,856	1,441,392	Lema_G081990.1	+

	0	763,884	765,737	Lema_G002200.1	+
	14	1,292,449	1,292,811	Lema_G098090.1	+
	14	152,839	155,417	Lema_G094000.1	-
	15	555,197	558,578	Lema_G091180.1	-
D1vD8					
	12	110,629	113,389	Lema_G082380.1	+
	6	590,695	591,212	Lema_G046870.1	-
	16	519,580	522,064	Lema_G104290.1	-
	12	109,077	109,426	Lema_G082370.1	+
	10	1,575,621	1,576,371	Lema_G076680.1	+
	13	1,418,093	1,424,481	Lema_G081920.1	-
	13	1,424,869	1,426,310	Lema_G081930.1	+
	16	513,425	514,524	Lema_G104260.1	+
D8vD1					
	13	1,444,412	1,445,637	Lema_G082010.1	-
	14	157,792	158,001	Lema_G094010.1	+
	13	1,446,238	1,447,610	Lema_G082020.1	+
	5	1,055,685	1,065,356	Lema_G053310.1	-
	13	1,441,845	1,442,615	Lema_G082000.1	+
	13	1,465,862	1,467,212	Lema_G082100.1	+
	14	158,218	167,699	Lema_G094020.1	+
	14	152,839	155,417	Lema_G094000.1	-
	15	555,197	558,578	Lema_G091180.1	-
	13	1,440,856	1,441,392	Lema_G081990.1	+
	0	763,884	765,737	Lema_G002200.1	+
	15	585,830	586,321	Lema_G091280.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	13	1,448,479	1,449,713	Lema_G082030.1	-
D1vD9					
	4	1,428,187	1,430,545	Lema_G043290.1	-
	15	500,031	500,476	Lema_G090930.1	-
	20	534,387	534,664	Lema_G114280.1	+
	20	535,205	538,183	Lema_G114290.1	-
	16	1,341,710	1,342,574	Lema_G106570.1	-
	20	549,910	552,133	Lema_G114370.1	-
	1	3,356,807	3,357,149	Lema_G023430.1	-

	20	529,112	533,279	Lema_G114270.1	-
	1	1,931,202	1,931,996	Lema_G018900.1	+
	20	554,610	555,882	Lema_G114390.1	+
	10	1,575,621	1,576,371	Lema_G076680.1	+
	8	1,427,511	1,429,907	Lema_G061030.1	-
	7	142,280	143,248	Lema_G067720.1	-
	20	552,415	553,355	Lema_G114380.1	+
D9vD1					
	14	1,291,448	1,292,362	Lema_G098080.1	+
	14	157,792	158,001	Lema_G094010.1	+
	14	158,218	167,699	Lema_G094020.1	+
	15	555,197	558,578	Lema_G091180.1	-
	14	1,292,449	1,292,811	Lema_G098090.1	+
	14	152,839	155,417	Lema_G094000.1	-
D1vD12					
	19	516,753	518,986	Lema_G111470.1	+
	15	500,031	500,476	Lema_G090930.1	-
	4	1,428,187	1,430,545	Lema_G043290.1	-
	20	70,380	71,347	Lema_G113170.1	-
	20	72,098	73,500	Lema_G113180.1	-
	12	110,629	113,389	Lema_G082380.1	+
	1	1,931,202	1,931,996	Lema_G018900.1	+
	7	142,280	143,248	Lema_G067720.1	-
	6	590,695	591,212	Lema_G046870.1	-
	5	347,544	347,994	Lema_G050760.1	+
	14	85,642	86,377	Lema_G093960.1	+
	1	229,948	233,218	Lema_G013150.1	-
	23	116,196	117,176	Lema_G119660.1	+
	23	112,460	115,068	Lema_G119640.1	+
	5	348,537	356,093	Lema_G050770.1	-
	20	57,305	58,784	Lema_G113150.1	+
	20	69,195	70,206	Lema_G113160.1	+
	23	115,658	115,852	Lema_G119650.1	-
	12	109,077	109,426	Lema_G082370.1	+
	3	1,421,067	1,423,417	Lema_G036130.1	-
D12vD1					

	13	1,426,780	1,428,165	Lema_G081940.1	+
	3	1,668,837	1,669,337	Lema_G037120.1	-
	14	1,291,448	1,292,362	Lema_G098080.1	+
	13	1,444,412	1,445,637	Lema_G082010.1	-
	1	844,294	846,759	Lema_G015110.1	+
	13	1,441,845	1,442,615	Lema_G082000.1	+
	14	157,792	158,001	Lema_G094010.1	+
	13	1,446,238	1,447,610	Lema_G082020.1	+
	5	1,055,685	1,065,356	Lema_G053310.1	-
	13	1,428,596	1,429,536	Lema_G081950.1	-
	13	1,465,862	1,467,212	Lema_G082100.1	+
	14	158,218	167,699	Lema_G094020.1	+
	13	1,440,856	1,441,392	Lema_G081990.1	+
	15	555,197	558,578	Lema_G091180.1	-
	0	763,884	765,737	Lema_G002200.1	+
	14	152,839	155,417	Lema_G094000.1	-
	14	1,292,449	1,292,811	Lema_G098090.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	13	1,448,479	1,449,713	Lema_G082030.1	-
D2vD3					
	13	1,441,845	1,442,615	Lema_G082000.1	+
	12	110,629	113,389	Lema_G082380.1	+
	13	1,449,952	1,450,340	Lema_G082040.1	+
	19	959,197	960,108	Lema_G113060.1	-
	13	1,352,946	1,353,440	Lema_G081640.1	+
	5	1,055,685	1,065,356	Lema_G053310.1	-
	13	1,428,596	1,429,536	Lema_G081950.1	-
	13	1,451,035	1,454,379	Lema_G082050.1	-
	13	1,448,479	1,449,713	Lema_G082030.1	-
D3vD2					
	15	585,830	586,321	Lema_G091280.1	+
	4	1,428,187	1,430,545	Lema_G043290.1	-
	16	514,658	516,407	Lema_G104270.1	-
	16	513,425	514,524	Lema_G104260.1	+
D2vD4					
	20	535,205	538,183	Lema_G114290.1	-

	14	1,291,448	1,292,362	Lema_G098080.1	+
	5	1,055,685	1,065,356	Lema_G053310.1	-
	13	1,418,093	1,424,481	Lema_G081920.1	-
	21	263,332	263,889	Lema_G116710.1	-
	13	1,411,646	1,412,969	Lema_G081880.1	+
D4vD2					
	16	514,658	516,407	Lema_G104270.1	-
	23	112,460	115,068	Lema_G119640.1	+
	16	513,425	514,524	Lema_G104260.1	+
D2vD5					
	19	959,197	960,108	Lema_G113060.1	-
	13	1,451,035	1,454,379	Lema_G082050.1	-
	3	1,668,837	1,669,337	Lema_G037120.1	-
	13	1,441,845	1,442,615	Lema_G082000.1	+
	13	1,465,862	1,467,212	Lema_G082100.1	+
	13	1,430,732	1,431,163	Lema_G081960.1	+
	13	1,440,856	1,441,392	Lema_G081990.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-
D5vD2					
	23	112,460	115,068	Lema_G119640.1	+
	6	1,607,018	1,607,681	Lema_G049660.1	+
D2vD6					
	12	110,629	113,389	Lema_G082380.1	+
D6vD2					
	15	585,830	586,321	Lema_G091280.1	+
	16	514,658	516,407	Lema_G104270.1	-
	4	1,428,187	1,430,545	Lema_G043290.1	-
	6	1,607,018	1,607,681	Lema_G049660.1	+
	16	513,425	514,524	Lema_G104260.1	+
D2vD8					
	13	1,424,869	1,426,310	Lema_G081930.1	+
	13	1,426,780	1,428,165	Lema_G081940.1	+
	3	1,668,837	1,669,337	Lema_G037120.1	-
D8vD2					
	15	585,830	586,321	Lema_G091280.1	+
	4	1,428,187	1,430,545	Lema_G043290.1	-

	23	112,460	115,068	Lema_G119640.1	+
D2vD9					
	20	535,205	538,183	Lema_G114290.1	-
	1	1,931,202	1,931,996	Lema_G018900.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	7	142,280	143,248	Lema_G067720.1	-
	13	1,446,238	1,447,610	Lema_G082020.1	+
	13	1,428,596	1,429,536	Lema_G081950.1	-
	5	1,055,685	1,065,356	Lema_G053310.1	-
	3	1,668,837	1,669,337	Lema_G037120.1	-
	13	1,459,766	1,465,599	Lema_G082090.1	-
	13	1,438,204	1,439,836	Lema_G081980.1	+
	20	534,387	534,664	Lema_G114280.1	+
	20	552,415	553,355	Lema_G114380.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-
D9vD2					
	16	514,658	516,407	Lema_G104270.1	-
	23	112,460	115,068	Lema_G119640.1	+
	4	1,428,187	1,430,545	Lema_G043290.1	-
	16	513,425	514,524	Lema_G104260.1	+
D2vD12					
	19	959,197	960,108	Lema_G113060.1	-
	12	110,629	113,389	Lema_G082380.1	+
	20	57,305	58,784	Lema_G113150.1	+
	5	347,544	347,994	Lema_G050760.1	+
	19	956,634	958,686	Lema_G113050.1	-
	13	1,352,946	1,353,440	Lema_G081640.1	+
D12vD2					
	16	514,658	516,407	Lema_G104270.1	-
	4	1,428,187	1,430,545	Lema_G043290.1	-
	16	513,425	514,524	Lema_G104260.1	+
D3vD4					
	21	263,332	263,889	Lema_G116710.1	-
	13	1,418,093	1,424,481	Lema_G081920.1	-
	13	1,411,646	1,412,969	Lema_G081880.1	+
D4vD3					

	13	1,449,952	1,450,340	Lema_G082040.1	+
	12	110,629	113,389	Lema_G082380.1	+
	13	1,441,845	1,442,615	Lema_G082000.1	+
	13	1,428,596	1,429,536	Lema_G081950.1	-
	23	112,460	115,068	Lema_G119640.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-
	13	1,451,035	1,454,379	Lema_G082050.1	-
D3vD5					
	16	513,425	514,524	Lema_G104260.1	+
D5vD3					
	5	1,055,685	1,065,356	Lema_G053310.1	-
	20	529,112	533,279	Lema_G114270.1	-
	23	112,460	115,068	Lema_G119640.1	+
	12	110,629	113,389	Lema_G082380.1	+
D3vD6					
D6vD3					
	13	1,441,845	1,442,615	Lema_G082000.1	+
	13	1,449,952	1,450,340	Lema_G082040.1	+
	13	1,428,596	1,429,536	Lema_G081950.1	-
	5	1,055,685	1,065,356	Lema_G053310.1	-
	20	529,112	533,279	Lema_G114270.1	-
	13	1,451,035	1,454,379	Lema_G082050.1	-
	6	1,607,018	1,607,681	Lema_G049660.1	+
	6	1,839,097	1,839,677	Lema_G049940.1	-
	13	1,352,946	1,353,440	Lema_G081640.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-
D3vD8					
	13	1,424,869	1,426,310	Lema_G081930.1	+
	16	513,425	514,524	Lema_G104260.1	+
D8vD3					
	13	1,441,845	1,442,615	Lema_G082000.1	+
	13	1,449,952	1,450,340	Lema_G082040.1	+
	5	1,055,685	1,065,356	Lema_G053310.1	-
	20	529,112	533,279	Lema_G114270.1	-
	23	112,460	115,068	Lema_G119640.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-

	6	1,839,097	1,839,677	Lema_G049940.1	-
	13	1,451,035	1,454,379	Lema_G082050.1	-
D3vD9					
	7	142,280	143,248	Lema_G067720.1	-
D9vD3					
	23	112,460	115,068	Lema_G119640.1	+
	12	110,629	113,389	Lema_G082380.1	+
D3vD12					
	5	347,544	347,994	Lema_G050760.1	+
	20	57,305	58,784	Lema_G113150.1	+
D12vD3					
	13	1,441,845	1,442,615	Lema_G082000.1	+
	13	1,449,952	1,450,340	Lema_G082040.1	+
	13	1,428,596	1,429,536	Lema_G081950.1	-
	5	1,055,685	1,065,356	Lema_G053310.1	-
	20	529,112	533,279	Lema_G114270.1	-
	6	1,839,097	1,839,677	Lema_G049940.1	-
	13	1,451,035	1,454,379	Lema_G082050.1	-
	13	1,448,479	1,449,713	Lema_G082030.1	-
D4vD5					
	13	1,426,780	1,428,165	Lema_G081940.1	+
	13	1,441,845	1,442,615	Lema_G082000.1	+
	16	513,425	514,524	Lema_G104260.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	3	1,668,837	1,669,337	Lema_G037120.1	-
	13	1,430,732	1,431,163	Lema_G081960.1	+
	13	1,465,862	1,467,212	Lema_G082100.1	+
	13	1,440,856	1,441,392	Lema_G081990.1	+
	13	1,459,766	1,465,599	Lema_G082090.1	-
	13	1,446,238	1,447,610	Lema_G082020.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-
D5vD4					
	20	535,205	538,183	Lema_G114290.1	-
	14	1,291,448	1,292,362	Lema_G098080.1	+
	21	263,332	263,889	Lema_G116710.1	-
	13	1,418,093	1,424,481	Lema_G081920.1	-

	5	1,055,685	1,065,356	Lema_G053310.1	-
	6	1,607,018	1,607,681	Lema_G049660.1	+
	13	1,411,646	1,412,969	Lema_G081880.1	+
D4vD6					
	18	909,482	910,569	Lema_G109270.1	-
	23	116,196	117,176	Lema_G119660.1	+
	12	110,629	113,389	Lema_G082380.1	+
	12	109,077	109,426	Lema_G082370.1	+
	3	1,703,798	1,704,842	Lema_G037220.1	-
	3	1,668,837	1,669,337	Lema_G037120.1	-
	23	112,460	115,068	Lema_G119640.1	+
	20	201,217	201,836	Lema_G113190.1	+
	5	490,652	491,090	Lema_G051290.1	-
D6vD4					
	20	535,205	538,183	Lema_G114290.1	-
	20	554,610	555,882	Lema_G114390.1	+
	5	415,359	416,426	Lema_G051020.1	-
	14	1,291,448	1,292,362	Lema_G098080.1	+
	1	1,931,202	1,931,996	Lema_G018900.1	+
	13	1,417,543	1,418,021	Lema_G081910.1	-
	20	534,387	534,664	Lema_G114280.1	+
	20	552,415	553,355	Lema_G114380.1	+
	18	31,059	32,963	Lema_G106610.1	+
	13	1,418,093	1,424,481	Lema_G081920.1	-
	13	1,352,946	1,353,440	Lema_G081640.1	+
	20	549,910	552,133	Lema_G114370.1	-
	5	1,055,685	1,065,356	Lema_G053310.1	-
	13	1,413,647	1,413,827	Lema_G081890.1	+
	6	1,607,018	1,607,681	Lema_G049660.1	+
	21	263,332	263,889	Lema_G116710.1	-
	13	1,415,843	1,416,552	Lema_G081900.1	-
	13	1,353,777	1,356,233	Lema_G081650.1	+
	14	1,292,449	1,292,811	Lema_G098090.1	+
	13	1,411,646	1,412,969	Lema_G081880.1	+
D4vD8					
	13	1,426,780	1,428,165	Lema_G081940.1	+

	3	1,668,837	1,669,337	Lema_G037120.1	-
	13	1,424,869	1,426,310	Lema_G081930.1	+
	16	513,425	514,524	Lema_G104260.1	+
D8vD4					
	21	263,332	263,889	Lema_G116710.1	-
	5	1,055,685	1,065,356	Lema_G053310.1	-
	13	1,418,093	1,424,481	Lema_G081920.1	-
	20	535,205	538,183	Lema_G114290.1	-
	13	1,411,646	1,412,969	Lema_G081880.1	+
D4vD9					
	3	1,669,441	1,669,644	Lema_G037130.1	-
	3	1,668,837	1,669,337	Lema_G037120.1	-
	13	1,428,596	1,429,536	Lema_G081950.1	-
	13	1,465,862	1,467,212	Lema_G082100.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	13	1,459,766	1,465,599	Lema_G082090.1	-
	13	1,438,204	1,439,836	Lema_G081980.1	+
	13	1,446,238	1,447,610	Lema_G082020.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-
D9vD4					
	21	263,332	263,889	Lema_G116710.1	-
	13	1,418,093	1,424,481	Lema_G081920.1	-
	14	1,291,448	1,292,362	Lema_G098080.1	+
	13	1,411,646	1,412,969	Lema_G081880.1	+
D4vD12					
	12	110,629	113,389	Lema_G082380.1	+
	5	347,544	347,994	Lema_G050760.1	+
	23	112,460	115,068	Lema_G119640.1	+
	20	57,305	58,784	Lema_G113150.1	+
D12vD4					
	20	535,205	538,183	Lema_G114290.1	-
	14	1,291,448	1,292,362	Lema_G098080.1	+
	5	1,055,685	1,065,356	Lema_G053310.1	-
	13	1,418,093	1,424,481	Lema_G081920.1	-
	21	263,332	263,889	Lema_G116710.1	-
	13	1,411,646	1,412,969	Lema_G081880.1	+

D5vD6					
	23	116,196	117,176	Lema_G119660.1	+
	18	909,482	910,569	Lema_G109270.1	-
	12	110,629	113,389	Lema_G082380.1	+
	1	844,294	846,759	Lema_G015110.1	+
	1	838,734	839,305	Lema_G015080.1	-
	3	1,703,798	1,704,842	Lema_G037220.1	-
	12	109,077	109,426	Lema_G082370.1	+
	23	112,460	115,068	Lema_G119640.1	+
	16	1,341,710	1,342,574	Lema_G106570.1	-
	20	201,217	201,836	Lema_G113190.1	+
	5	490,652	491,090	Lema_G051290.1	-
	7	142,280	143,248	Lema_G067720.1	-
	1	839,417	839,638	Lema_G015090.1	+
D6vD5					
	13	1,426,780	1,428,165	Lema_G081940.1	+
	1	1,722,362	1,724,272	Lema_G018140.1	-
	1	3,356,807	3,357,149	Lema_G023430.1	-
	13	1,430,732	1,431,163	Lema_G081960.1	+
	13	1,428,596	1,429,536	Lema_G081950.1	-
	13	1,459,766	1,465,599	Lema_G082090.1	-
	15	585,830	586,321	Lema_G091280.1	+
	13	1,353,777	1,356,233	Lema_G081650.1	+
	3	1,421,067	1,423,417	Lema_G036130.1	-
	13	1,431,699	1,432,295	Lema_G081970.1	-
	7	331,214	333,121	Lema_G068460.1	+
	13	1,465,862	1,467,212	Lema_G082100.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	13	1,440,856	1,441,392	Lema_G081990.1	+
	16	513,425	514,524	Lema_G104260.1	+
	16	511,036	512,849	Lema_G104250.1	-
	13	1,444,412	1,445,637	Lema_G082010.1	-
	6	1,837,589	1,837,925	Lema_G049930.1	-
	6	1,837,314	1,837,498	Lema_G049920.1	-
	13	1,352,946	1,353,440	Lema_G081640.1	+
	17	408,495	409,409	Lema_G099980.1	+

	13	1,438,204	1,439,836	Lema_G081980.1	+
	0	1,574,046	1,574,450	Lema_G004720.1	+
	16	519,580	522,064	Lema_G104290.1	-
	13	1,441,845	1,442,615	Lema_G082000.1	+
	16	514,658	516,407	Lema_G104270.1	-
	13	1,448,479	1,449,713	Lema_G082030.1	-
	13	1,446,238	1,447,610	Lema_G082020.1	+
D5vD8					
	13	1,424,869	1,426,310	Lema_G081930.1	+
D8vD5					
	13	1,441,845	1,442,615	Lema_G082000.1	+
	13	1,465,862	1,467,212	Lema_G082100.1	+
	13	1,446,238	1,447,610	Lema_G082020.1	+
	13	1,440,856	1,441,392	Lema_G081990.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	13	1,448,479	1,449,713	Lema_G082030.1	-
D5vD9					
	20	535,205	538,183	Lema_G114290.1	-
	7	142,280	143,248	Lema_G067720.1	-
	5	1,055,685	1,065,356	Lema_G053310.1	-
	6	1,607,018	1,607,681	Lema_G049660.1	+
	20	534,387	534,664	Lema_G114280.1	+
	20	552,415	553,355	Lema_G114380.1	+
D9vD5					
	16	513,425	514,524	Lema_G104260.1	+
D5vD12					
	23	112,460	115,068	Lema_G119640.1	+
	12	110,629	113,389	Lema_G082380.1	+
	5	347,544	347,994	Lema_G050760.1	+
	6	1,607,018	1,607,681	Lema_G049660.1	+
	20	57,305	58,784	Lema_G113150.1	+
D12vD5					
	16	513,425	514,524	Lema_G104260.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	13	1,441,845	1,442,615	Lema_G082000.1	+
	13	1,465,862	1,467,212	Lema_G082100.1	+

	3	1,668,837	1,669,337	Lema_G037120.1	-
	13	1,459,766	1,465,599	Lema_G082090.1	-
	13	1,440,856	1,441,392	Lema_G081990.1	+
	13	1,446,238	1,447,610	Lema_G082020.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-
D6vD8					
	13	1,426,780	1,428,165	Lema_G081940.1	+
	13	1,418,093	1,424,481	Lema_G081920.1	-
	14	1,291,448	1,292,362	Lema_G098080.1	+
	13	1,424,869	1,426,310	Lema_G081930.1	+
	16	511,036	512,849	Lema_G104250.1	-
	13	1,428,596	1,429,536	Lema_G081950.1	-
	10	1,575,621	1,576,371	Lema_G076680.1	+
	13	1,430,732	1,431,163	Lema_G081960.1	+
	6	1,607,018	1,607,681	Lema_G049660.1	+
	14	1,292,449	1,292,811	Lema_G098090.1	+
	13	1,353,777	1,356,233	Lema_G081650.1	+
	16	519,580	522,064	Lema_G104290.1	-
	16	513,425	514,524	Lema_G104260.1	+
D8vD6					
	4	1,428,187	1,430,545	Lema_G043290.1	-
	15	500,031	500,476	Lema_G090930.1	-
	23	116,196	117,176	Lema_G119660.1	+
	7	142,280	143,248	Lema_G067720.1	-
	3	1,703,798	1,704,842	Lema_G037220.1	-
	16	1,341,710	1,342,574	Lema_G106570.1	-
	23	112,460	115,068	Lema_G119640.1	+
	20	201,217	201,836	Lema_G113190.1	+
	5	490,652	491,090	Lema_G051290.1	-
D6vD9					
	20	535,205	538,183	Lema_G114290.1	-
	6	1,607,018	1,607,681	Lema_G049660.1	+
	20	534,387	534,664	Lema_G114280.1	+
	20	552,415	553,355	Lema_G114380.1	+
	13	1,446,238	1,447,610	Lema_G082020.1	+
	13	1,428,596	1,429,536	Lema_G081950.1	-

	13	1,465,862	1,467,212	Lema_G082100.1	+
	5	1,055,685	1,065,356	Lema_G053310.1	-
	13	1,451,035	1,454,379	Lema_G082050.1	-
	13	1,459,766	1,465,599	Lema_G082090.1	-
	13	1,438,204	1,439,836	Lema_G081980.1	+
	1	1,931,202	1,931,996	Lema_G018900.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-
D9vD6					
	12	110,629	113,389	Lema_G082380.1	+
D8vD9					
	20	535,205	538,183	Lema_G114290.1	-
	1	1,931,202	1,931,996	Lema_G018900.1	+
	7	142,280	143,248	Lema_G067720.1	-
	13	1,446,238	1,447,610	Lema_G082020.1	+
	5	1,055,685	1,065,356	Lema_G053310.1	-
	20	552,415	553,355	Lema_G114380.1	+
	13	1,459,766	1,465,599	Lema_G082090.1	-
	13	1,438,204	1,439,836	Lema_G081980.1	+
	20	534,387	534,664	Lema_G114280.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	13	1,448,479	1,449,713	Lema_G082030.1	-
D9vD8					
	13	1,424,869	1,426,310	Lema_G081930.1	+
	16	513,425	514,524	Lema_G104260.1	+
D8vD12					
	5	347,544	347,994	Lema_G050760.1	+
	23	112,460	115,068	Lema_G119640.1	+
	20	57,305	58,784	Lema_G113150.1	+
D12vD8					
	13	1,426,780	1,428,165	Lema_G081940.1	+
	13	1,424,869	1,426,310	Lema_G081930.1	+
	3	1,668,837	1,669,337	Lema_G037120.1	-
	16	513,425	514,524	Lema_G104260.1	+
D6vD12					
	20	70,380	71,347	Lema_G113170.1	-
	6	1,607,018	1,607,681	Lema_G049660.1	+

	1	229,948	233,218	Lema_G013150.1	-
	13	1,350,995	1,351,795	Lema_G081630.1	+
	20	57,305	58,784	Lema_G113150.1	+
	13	1,352,946	1,353,440	Lema_G081640.1	+
	5	347,544	347,994	Lema_G050760.1	+
	5	348,537	356,093	Lema_G050770.1	-
	3	1,421,067	1,423,417	Lema_G036130.1	-
	20	69,195	70,206	Lema_G113160.1	+
	20	72,098	73,500	Lema_G113180.1	-
	14	85,642	86,377	Lema_G093960.1	+
	15	585,830	586,321	Lema_G091280.1	+
	13	1,353,777	1,356,233	Lema_G081650.1	+
	19	516,753	518,986	Lema_G111470.1	+
	1	1,931,202	1,931,996	Lema_G018900.1	+
D12vD6					
	1	844,294	846,759	Lema_G015110.1	+
	1	838,734	839,305	Lema_G015080.1	-
	3	1,668,837	1,669,337	Lema_G037120.1	-
	3	1,703,798	1,704,842	Lema_G037220.1	-
	16	1,341,710	1,342,574	Lema_G106570.1	-
	20	201,217	201,836	Lema_G113190.1	+
	5	490,652	491,090	Lema_G051290.1	-
	1	839,417	839,638	Lema_G015090.1	+
D9vD12					
	23	112,460	115,068	Lema_G119640.1	+
	5	347,544	347,994	Lema_G050760.1	+
	12	110,629	113,389	Lema_G082380.1	+
	20	57,305	58,784	Lema_G113150.1	+
D12vD9					
	20	535,205	538,183	Lema_G114290.1	-
	20	534,387	534,664	Lema_G114280.1	+
	3	1,668,837	1,669,337	Lema_G037120.1	-
	13	1,446,238	1,447,610	Lema_G082020.1	+
	13	1,428,596	1,429,536	Lema_G081950.1	-
	5	1,055,685	1,065,356	Lema_G053310.1	-
	13	1,448,479	1,449,713	Lema_G082030.1	-

	13	1,459,766	1,465,599	Lema_G082090.1	-
	3	1,669,441	1,669,644	Lema_G037130.1	-
	20	552,415	553,355	Lema_G114380.1	+
D10vD1					
	0	763,884	765,737	Lema_G002200.1	+
	14	158,218	167,699	Lema_G094020.1	+
	14	1,292,449	1,292,811	Lema_G098090.1	+
	5	1,055,685	1,065,356	Lema_G053310.1	-
	14	152,839	155,417	Lema_G094000.1	-
D1vD10					
	15	500,031	500,476	Lema_G090930.1	-
	1	3,356,807	3,357,149	Lema_G023430.1	-
	17	1,347,458	1,347,676	Lema_G102850.1	-
	17	1,350,057	1,350,260	Lema_G102860.1	-
	17	1,350,963	1,352,196	Lema_G102870.1	-
	16	1,038,989	1,040,299	Lema_G105350.1	-
	17	1,353,494	1,353,697	Lema_G102880.1	+
	17	1,355,208	1,355,803	Lema_G102890.1	-
	20	1,048,525	1,050,375	Lema_G115920.1	+
	10	1,575,621	1,576,371	Lema_G076680.1	+
	1	966,655	967,848	Lema_G015420.1	+
	17	1,345,833	1,346,186	Lema_G102840.1	+
	0	1,193,050	1,193,633	Lema_G003440.1	+
	4	1,466,935	1,468,128	Lema_G043470.1	-
	1	1,722,362	1,724,272	Lema_G018140.1	-
	12	1,437,979	1,439,270	Lema_G086350.1	+
D10vD12					
	5	347,544	347,994	Lema_G050760.1	+
	20	57,305	58,784	Lema_G113150.1	+
D12vD10					
	13	1,426,780	1,428,165	Lema_G081940.1	+
	17	1,347,458	1,347,676	Lema_G102850.1	-
	1	1,722,362	1,724,272	Lema_G018140.1	-
	16	1,038,989	1,040,299	Lema_G105350.1	-
	13	1,430,732	1,431,163	Lema_G081960.1	+
	13	1,428,596	1,429,536	Lema_G081950.1	-

	13	1,459,766	1,465,599	Lema_G082090.1	-
	13	1,431,699	1,432,295	Lema_G081970.1	-
	13	1,444,412	1,445,637	Lema_G082010.1	-
	17	1,345,833	1,346,186	Lema_G102840.1	+
	13	1,465,862	1,467,212	Lema_G082100.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	13	1,440,856	1,441,392	Lema_G081990.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-
	13	1,449,952	1,450,340	Lema_G082040.1	+
	17	1,350,057	1,350,260	Lema_G102860.1	-
	17	1,355,208	1,355,803	Lema_G102890.1	-
	1	966,655	967,848	Lema_G015420.1	+
	13	1,438,204	1,439,836	Lema_G081980.1	+
	1	844,294	846,759	Lema_G015110.1	+
	17	1,350,963	1,352,196	Lema_G102870.1	-
	1	838,734	839,305	Lema_G015080.1	-
	13	1,441,845	1,442,615	Lema_G082000.1	+
	10	1,575,621	1,576,371	Lema_G076680.1	+
	1	839,417	839,638	Lema_G015090.1	+
	17	1,353,494	1,353,697	Lema_G102880.1	+
	13	1,446,238	1,447,610	Lema_G082020.1	+
D10vD2					
	16	514,658	516,407	Lema_G104270.1	-
	4	1,428,187	1,430,545	Lema_G043290.1	-
	23	112,460	115,068	Lema_G119640.1	+
D2vD10					
	13	1,426,780	1,428,165	Lema_G081940.1	+
	17	1,347,458	1,347,676	Lema_G102850.1	-
	1	1,722,362	1,724,272	Lema_G018140.1	-
	16	1,038,989	1,040,299	Lema_G105350.1	-
	17	1,355,208	1,355,803	Lema_G102890.1	-
	19	959,197	960,108	Lema_G113060.1	-
	13	1,459,766	1,465,599	Lema_G082090.1	-
	13	1,353,777	1,356,233	Lema_G081650.1	+
	19	956,634	958,686	Lema_G113050.1	-
	13	1,431,699	1,432,295	Lema_G081970.1	-

	13	1,449,952	1,450,340	Lema_G082040.1	+
	13	1,465,862	1,467,212	Lema_G082100.1	+
	13	1,352,946	1,353,440	Lema_G081640.1	+
	13	1,440,856	1,441,392	Lema_G081990.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-
	18	909,482	910,569	Lema_G109270.1	-
	13	1,444,412	1,445,637	Lema_G082010.1	-
	17	1,350,057	1,350,260	Lema_G102860.1	-
	13	1,430,732	1,431,163	Lema_G081960.1	+
	1	966,655	967,848	Lema_G015420.1	+
	13	1,428,596	1,429,536	Lema_G081950.1	-
	13	1,438,204	1,439,836	Lema_G081980.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	17	1,350,963	1,352,196	Lema_G102870.1	-
	1	844,294	846,759	Lema_G015110.1	+
	1	838,734	839,305	Lema_G015080.1	-
	13	1,441,845	1,442,615	Lema_G082000.1	+
	10	1,575,621	1,576,371	Lema_G076680.1	+
	1	839,417	839,638	Lema_G015090.1	+
	17	1,353,494	1,353,697	Lema_G102880.1	+
	13	1,446,238	1,447,610	Lema_G082020.1	+
D10vD3					
	23	112,460	115,068	Lema_G119640.1	+
	5	1,055,685	1,065,356	Lema_G053310.1	-
D3vD10					
	18	909,482	910,569	Lema_G109270.1	-
	1	1,722,362	1,724,272	Lema_G018140.1	-
	17	1,347,458	1,347,676	Lema_G102850.1	-
	1	966,655	967,848	Lema_G015420.1	+
	17	1,350,963	1,352,196	Lema_G102870.1	-
	17	1,350,057	1,350,260	Lema_G102860.1	-
	16	1,038,989	1,040,299	Lema_G105350.1	-
	17	1,355,208	1,355,803	Lema_G102890.1	-
	17	1,345,833	1,346,186	Lema_G102840.1	+
	10	1,575,621	1,576,371	Lema_G076680.1	+
	1	839,417	839,638	Lema_G015090.1	+

	17	1,353,494	1,353,697	Lema_G102880.1	+
	15	585,830	586,321	Lema_G091280.1	+
	1	838,734	839,305	Lema_G015080.1	-
	1	844,294	846,759	Lema_G015110.1	+
D10vD4					
	20	535,205	538,183	Lema_G114290.1	-
	5	1,055,685	1,065,356	Lema_G053310.1	-
	14	1,291,448	1,292,362	Lema_G098080.1	+
D4vD10					
	13	1,426,780	1,428,165	Lema_G081940.1	+
	17	1,347,458	1,347,676	Lema_G102850.1	-
	1	1,722,362	1,724,272	Lema_G018140.1	-
	13	1,449,952	1,450,340	Lema_G082040.1	+
	16	1,038,989	1,040,299	Lema_G105350.1	-
	1	3,356,807	3,357,149	Lema_G023430.1	-
	17	1,355,208	1,355,803	Lema_G102890.1	-
	13	1,428,596	1,429,536	Lema_G081950.1	-
	13	1,459,766	1,465,599	Lema_G082090.1	-
	13	1,431,699	1,432,295	Lema_G081970.1	-
	13	1,444,412	1,445,637	Lema_G082010.1	-
	13	1,465,862	1,467,212	Lema_G082100.1	+
	13	1,440,856	1,441,392	Lema_G081990.1	+
	17	1,345,833	1,346,186	Lema_G102840.1	+
	18	909,482	910,569	Lema_G109270.1	-
	13	1,430,732	1,431,163	Lema_G081960.1	+
	1	966,655	967,848	Lema_G015420.1	+
	13	1,438,204	1,439,836	Lema_G081980.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	17	1,350,963	1,352,196	Lema_G102870.1	-
	13	1,441,845	1,442,615	Lema_G082000.1	+
	10	1,575,621	1,576,371	Lema_G076680.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-
	17	1,353,494	1,353,697	Lema_G102880.1	+
	13	1,446,238	1,447,610	Lema_G082020.1	+
D10vD5					
	16	513,425	514,524	Lema_G104260.1	+

D5vD10					
	1	839,417	839,638	Lema_G015090.1	+
	17	1,347,458	1,347,676	Lema_G102850.1	-
	18	909,482	910,569	Lema_G109270.1	-
	17	1,350,963	1,352,196	Lema_G102870.1	-
	16	1,038,989	1,040,299	Lema_G105350.1	-
	17	1,353,494	1,353,697	Lema_G102880.1	+
	17	1,355,208	1,355,803	Lema_G102890.1	-
	10	1,575,621	1,576,371	Lema_G076680.1	+
	1	966,655	967,848	Lema_G015420.1	+
	1	844,294	846,759	Lema_G015110.1	+
	6	1,607,018	1,607,681	Lema_G049660.1	+
	1	838,734	839,305	Lema_G015080.1	-
	17	1,345,833	1,346,186	Lema_G102840.1	+
D10vD6					
D6vD10					
	13	1,426,780	1,428,165	Lema_G081940.1	+
	1	966,655	967,848	Lema_G015420.1	+
	1	1,722,362	1,724,272	Lema_G018140.1	-
	16	1,038,989	1,040,299	Lema_G105350.1	-
	1	3,356,807	3,357,149	Lema_G023430.1	-
	13	1,430,732	1,431,163	Lema_G081960.1	+
	13	1,350,995	1,351,795	Lema_G081630.1	+
	13	1,441,845	1,442,615	Lema_G082000.1	+
	13	1,459,766	1,465,599	Lema_G082090.1	-
	15	585,830	586,321	Lema_G091280.1	+
	13	1,353,777	1,356,233	Lema_G081650.1	+
	13	1,431,699	1,432,295	Lema_G081970.1	-
	13	1,449,952	1,450,340	Lema_G082040.1	+
	13	1,465,862	1,467,212	Lema_G082100.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
	13	1,440,856	1,441,392	Lema_G081990.1	+
	13	1,444,412	1,445,637	Lema_G082010.1	-
	13	1,448,479	1,449,713	Lema_G082030.1	-
	17	1,350,057	1,350,260	Lema_G102860.1	-
	17	1,355,208	1,355,803	Lema_G102890.1	-

	6	1,607,018	1,607,681	Lema_G049660.1	+
	13	1,438,204	1,439,836	Lema_G081980.1	+
	13	1,352,946	1,353,440	Lema_G081640.1	+
	17	1,347,458	1,347,676	Lema_G102850.1	-
	17	1,350,963	1,352,196	Lema_G102870.1	-
	13	1,428,596	1,429,536	Lema_G081950.1	-
	10	1,575,621	1,576,371	Lema_G076680.1	+
	17	1,345,833	1,346,186	Lema_G102840.1	+
	17	1,353,494	1,353,697	Lema_G102880.1	+
	13	1,446,238	1,447,610	Lema_G082020.1	+
D8vD10					
	17	1,350,057	1,350,260	Lema_G102860.1	-
	1	1,722,362	1,724,272	Lema_G018140.1	-
	13	1,431,699	1,432,295	Lema_G081970.1	-
	17	1,347,458	1,347,676	Lema_G102850.1	-
	13	1,444,412	1,445,637	Lema_G082010.1	-
	17	1,350,963	1,352,196	Lema_G102870.1	-
	13	1,449,952	1,450,340	Lema_G082040.1	+
	17	1,355,208	1,355,803	Lema_G102890.1	-
	13	1,446,238	1,447,610	Lema_G082020.1	+
	17	1,353,494	1,353,697	Lema_G102880.1	+
	13	1,441,845	1,442,615	Lema_G082000.1	+
	17	1,345,833	1,346,186	Lema_G102840.1	+
	13	1,465,862	1,467,212	Lema_G082100.1	+
	1	966,655	967,848	Lema_G015420.1	+
	13	1,448,479	1,449,713	Lema_G082030.1	-
	13	1,459,766	1,465,599	Lema_G082090.1	-
	13	1,440,856	1,441,392	Lema_G081990.1	+
	13	1,438,204	1,439,836	Lema_G081980.1	+
	16	1,038,989	1,040,299	Lema_G105350.1	-
	15	585,830	586,321	Lema_G091280.1	+
	13	1,451,035	1,454,379	Lema_G082050.1	-
D10vD9					
	20	535,205	538,183	Lema_G114290.1	-
	20	534,387	534,664	Lema_G114280.1	+
	5	1,055,685	1,065,356	Lema_G053310.1	-

	1	1,931,202	1,931,996	Lema_G018900.1	+
	3	1,669,441	1,669,644	Lema_G037130.1	-
	20	552,415	553,355	Lema_G114380.1	+
D9vD10					
	17	1,347,458	1,347,676	Lema_G102850.1	-
	17	1,350,963	1,352,196	Lema_G102870.1	-
	1	1,722,362	1,724,272	Lema_G018140.1	-
	17	1,350,057	1,350,260	Lema_G102860.1	-
	16	1,038,989	1,040,299	Lema_G105350.1	-
	17	1,355,208	1,355,803	Lema_G102890.1	-
	1	966,655	967,848	Lema_G015420.1	+
	17	1,353,494	1,353,697	Lema_G102880.1	+
	17	1,345,833	1,346,186	Lema_G102840.1	+

The genes affected by predicted PAVs and seen in table 5.9 are summarised in table 5.10 as a full list.

Table 5.10: List of genes with predicted PAVs. Annotation highlighted in yellow is a gene affected by a high number of SNPs (section 4.3.4). Green annotations are to highlight high incidence of PAVs on SC13 (note the large cluster highlighted in red). Positions marked in blue highlight the clustering of PAV affected genes. Genes highlighted in pink were identified as lost by Golicz et al. (2015).

Super Contig No.	Start	End	Annotation	Predicted function	Number of PAV Hits
0	763,884	765,737	Lema_P002200.1	similar to MFS multidrug transporter	8
0	1,193,050	1,193,633	Lema_P003440.1	-	1
0	1,574,046	1,574,450	Lema_P004720.1	-	2
1	229,948	233,218	Lema_P013150.1	-	3
1	838,734	839,305	Lema_P015080.1	-	6
1	839,417	839,638	Lema_P015090.1	-	6
1	844,294	846,759	Lema_P015110.1	-	10
1	966,655	967,848	Lema_P015420.1	-	9
1	1,722,362	1,724,272	Lema_P018140.1	-	10
1	1,931,202	1,931,996	Lema_P018900.1	-	10
1	3,356,807	3,357,149	Lema_P023430.1	-	8

3	1,421,067	1,423,417	Lema_P036130.1	-	4
3	1,668,837	1,669,337	Lema_P037120.1	-	14
3	1,669,441	1,669,644	Lema_P037130.1	-	3
3	1,703,798	1,704,842	Lema_P037220.1	similar to Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase	5
4	1,428,187	1,430,545	Lema_P043290.1	Similar to dynamin family protein	14
4	1,445,551	1,447,257	Lema_P043360.1	-	1
4	1,466,935	1,468,128	Lema_P043470.1	-	1
5	347,544	347,994	Lema_P050760.1	-	9
5	348,537	356,093	Lema_P050770.1	-	2
5	415,359	416,426	Lema_P051020.1	-	2
5	490,652	491,090	Lema_P051290.1	-	5
5	1,055,685	1,065,356	Lema_P053310.1	-	24
6	590,695	591,212	Lema_P046870.1	-	2
6	1,607,018	1,607,681	Lema_P049660.1	*AvrLm1*	14
6	1,837,314	1,837,498	Lema_P049920.1	-	3
6	1,837,589	1,837,925	Lema_P049930.1	-	3
6	1,839,097	1,839,677	Lema_P049940.1	*AvrLm6*	4
7	142,280	143,248	Lema_P067720.1	-	10
7	331,214	333,121	Lema_P068460.1	similar to beta-glucosidase	2
8	1,427,511	1,429,907	Lema_P061030.1	-	1
10	1,575,621	1,576,371	Lema_P076680.1	-	10
12	109,077	109,426	Lema_P082370.1	-	6
12	110,629	113,389	Lema_P082380.1	-	16
12	1,400,949	1,401,849	Lema_P086300.1	-	1
12	1,402,878	1,403,303	Lema_P086310.1	-	1
12	1,437,979	1,439,270	Lema_P086350.1	-	1
13	319,292	320,025	Lema_P078330.1	-	1
13	1,350,995	1,351,795	Lema_P081630.1	-	2
13	1,352,946	1,353,440	Lema_P081640.1	-	10
13	1,353,777	1,356,233	Lema_P081650.1	-	6
13	1,411,646	1,412,969	Lema_P081880.1	*sirH*	8
13	1,413,647	1,413,827	Lema_P081890.1	-	2
13	1,415,843	1,416,552	Lema_P081900.1	*LmUVI-1h*	2
13	1,417,543	1,418,021	Lema_P081910.1	similar to polyketide synthase	2
13	1,418,093	1,424,481	Lema_P081920.1	*LmPKS1*	10
13	1,424,869	1,426,310	Lema_P081930.1	*LmHDX1*	8
13	1,426,780	1,428,165	Lema_P081940.1	similar to aspartate aminotransferase	14

13	1,428,596	1,429,536	Lema_P081950.1	aflatoxin biosynthesis ketoreductase nor-1	18
13	1,430,732	1,431,163	Lema_P081960.1	similar to stress responsive A/B barrel domain protein	11
13	1,431,699	1,432,295	Lema_P081970.1	-	6
13	1,438,204	1,439,836	Lema_P081980.1	-	10
13	1,440,856	1,441,392	Lema_P081990.1	-	15
13	1,441,845	1,442,615	Lema_P082000.1	similar to short-chain dehydrogenase/reductase SDR	20
13	1,444,412	1,445,637	Lema_P082010.1	similar to oxidoreductase	11
13	1,446,238	1,447,610	Lema_P082020.1	similar to o-methyltransferase	19
13	1,448,479	1,449,713	Lema_P082030.1	similar to putative 2,3-dihydroxybiphenyl-1,2-dioxygenase or glyoxalase/bleomycin resistance protein	25
13	1,449,952	1,450,340	Lema_P082040.1	-	10
13	1,451,035	1,454,379	Lema_P082050.1	-	24
13	1,459,766	1,465,599	Lema_P082090.1	similar to fatty acid synthase beta subunit	13
13	1,465,862	1,467,212	Lema_P082100.1	-	17
14	85,642	86,377	Lema_P093960.1	-	2
14	152,839	155,417	Lema_P094000.1	-	9
14	157,792	158,001	Lema_P094010.1	-	8
14	158,218	167,699	Lema_P094020.1	-	9
14	1,291,448	1,292,362	Lema_P098080.1	-	12
14	1,292,449	1,292,811	Lema_P098090.1	-	8
15	500,031	500,476	Lema_P090930.1	-	5
15	555,197	558,578	Lema_P091180.1	-	8
15	585,830	586,321	Lema_P091280.1	similar to ribonuclease H	11
16	511,036	512,849	Lema_P104250.1	-	2
16	513,425	514,524	Lema_P104260.1	-	19
16	514,658	516,407	Lema_P104270.1	-	9
16	519,580	522,064	Lema_P104290.1	-	5
16	1,038,989	1,040,299	Lema_P105350.1	similar to actin monomer binding protein	9
16	1,341,710	1,342,574	Lema_P106570.1	-	6
17	408,495	409,409	Lema_P099980.1	similar to glutathione S-transferase protein	1

17	1,345,833	1,346,186	Lema_P102840.1	-	8
17	1,347,458	1,347,676	Lema_P102850.1	-	9
17	1,350,057	1,350,260	Lema_P102860.1	-	7
17	1,350,963	1,352,196	Lema_P102870.1	-	9
17	1,353,494	1,353,697	Lema_P102880.1	-	9
17	1,355,208	1,355,803	Lema_P102890.1	-	9
18	31,059	32,963	Lema_P106610.1	similar to tetracycline-efflux transporter	2
18	347,790	348,175	Lema_P107350.1	-	1
18	905,419	907,083	Lema_P109260.1	-	4
18	909,482	910,569	Lema_P109270.1	-	6
19	516,753	518,986	Lema_P111470.1	-	2
19	956,634	958,686	Lema_P113050.1	-	3
19	959,197	960,108	Lema_P113060.1	-	5
20	57,305	58,784	Lema_P113150.1	-	9
20	69,195	70,206	Lema_P113160.1	-	2
20	70,380	71,347	Lema_P113170.1	-	2
20	72,098	73,500	Lema_P113180.1	similar to salicylate hydroxylase	2
20	201,217	201,836	Lema_P113190.1	-	5
20	529,112	533,279	Lema_P114270.1	-	6
20	534,387	534,664	Lema_P114280.1	-	10
20	535,205	538,183	Lema_P114290.1	-	15
20	549,910	552,133	Lema_P114370.1	similar to Rho GTPase activator (Rgd1)	4
20	552,415	553,355	Lema_P114380.1	similar to pyridoxamine phosphate oxidase	10
20	554,610	555,882	Lema_P114390.1	-	4
20	1,048,525	1,050,375	Lema_P115920.1	similar to carboxylesterase	1
21	263,332	263,889	Lema_P116710.1	-	8
22	105,263	106,588	Lema_P119050.1	-	1
23	112,460	115,068	Lema_P119640.1	-	21
23	115,658	115,852	Lema_P119650.1	-	2
23	116,196	117,176	Lema_P119660.1	-	7

SC13 had the highest number of genes affected by PAVs, with 24 (22.4%) of the total genes being located there. The average number of PAV hits (hits from individual comparisons e.g. 10 reads from D1vD2 counts as 1 hit) per gene was 7.48. The gene annotation with the highest number of PAV hits was Lema_P082030.1, which has the

putative function of bleomycin resistance. Twelve genes (11.2%) only showed a single PAV hit and 32 genes (29.9%) had ten PAV hits or more.

The gene hits were often located within close proximity to other PAV read hits on the same SuperContig. For example in SC16, four PAV hits (out of a total of 6 for SC16) were located within a region of 11,028 bp (0.8% of the SC length). Five of the genes (highlighted in pink in table 5.10) were previously found to be present/lost by Golicz et al. (2015).

Of the 107 gene annotations that show mapping of PAV reads, the majority (74.8%) have thus far not been provided with a predicted function. Of the 27 genes with predicted function, two are the avirulence genes *AvrLm1* and *AvrLm6*. A number of these genes can potentially affect infection processes; e.g. the MFS multidrug transporter gene is involved in resistance of fungi to antifungals (Costa et al., 2014), dynamin proteins can play a role in plant resistance to pathogens (Praefcke and McMahon, 2004) and the nor-1 gene is involved in aflatoxin (mycotoxin) biosynthesis in *Aspergillus parasiticus* (Zhou and Linz, 1999). Furthermore, two annotations show putative involvement in resistance to the antibiotics bleomycin and tetracycline. Carboxylases can also play a role in metabolising foreign toxins and drugs (Aranda et al., 2014).

A number of metabolic genes showed PAV read hits, with genes involved in amino acid metabolism (aspartate aminotransferase), cellulase activity (beta-glucosidase) and fatty acid metabolism (fatty acid synthase and Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase).

5.4.8. SuperContig 13 PAVs

SC13 exhibits a very high number of predicted PAV reads. The SC is 1,634,580 bp in length and contains 492 gene annotations. The 264 PAV comparison gene hits on SC13 (figure 5.7) hit just 24 of these genes, which can be seen in table 5.11.

Table 5.11: Gene Annotations for PAV comparison hits on SC13.

Position on SC13		Gene Annotation	No. of PAV comparison Hits	Functional Annotation
Start	End			
319,292	320,025	Lema_G078330.1	1	
1,350,995	1,351,795	Lema_G081630.1	2	
1,352,946	1,353,440	Lema_G081640.1	10	
1,353,777	1,356,233	Lema_G081650.1	6	
1,411,646	1,412,969	Lema_G081880.1	8	*sirH*
1,413,647	1,413,827	Lema_G081890.1	2	
1,415,843	1,416,552	Lema_G081900.1	2	*LmUVI-1h*
1,417,543	1,418,021	Lema_G081910.1	2	similar to polyketide synthase
1,418,093	1,424,481	Lema_G081920.1	10	*LmPKS1*
1,424,869	1,426,310	Lema_G081930.1	8	*LmHDX1*
1,426,780	1,428,165	Lema_G081940.1	14	similar to aspartate aminotransferase
1,428,596	1,429,536	Lema_G081950.1	18	aflatoxin biosynthesis ketoreductase nor-1
1,430,732	1,431,163	Lema_G081960.1	11	similar to stress responsive A/B barrel domain protein
1,431,699	1,432,295	Lema_G081970.1	6	
1,438,204	1,439,836	Lema_G081980.1	10	
1,440,856	1,441,392	Lema_G081990.1	15	

1,441,845	1,442,615	Lema_G082000.1	20	similar to short-chain dehydrogenase/reductase SDR
1,444,412	1,445,637	Lema_G082010.1	11	similar to oxidoreductase
1,446,238	1,447,610	Lema_G082020.1	19	similar to o-methyltransferase
1,448,479	1,449,713	Lema_G082030.1	25	similar to putative 2,3-dihydroxybiphenyl-1,2-dioxygenase or glyoxalase/bleomycin resistance protein
1,449,952	1,450,340	Lema_G082040.1	10	
1,451,035	1,454,379	Lema_G082050.1	24	
1,459,766	1,465,599	Lema_G082090.1	13	similar to fatty acid synthase beta subunit
1,465,862	1,467,212	Lema_G082100.1	17	

The majority of these genes are present in a region of ~116 Kbp from positions 1,350,995 to 1,467,212 on SC13 (7% of SuperContig length), the only one outside this is Lema_G078330.1, which is 1 Mbp upstream. This region contains 48 annotated genes, of which 24 have predicted PAV reads that map to them.

The 68 Kbp cluster of sirodesmin biosynthesis genes (Gardiner et al., 2004) is located within this region of high PAV read prediction. Of the 23 annotated genes in this cluster (GenBank accession AY553235), four coincide with PAV read hits. These are *sirH*, *LmUVI-1h*, *LmPKS1* and *LmHDX1*.

5.4.9. PAV Prediction in *AvrLm* Genes

The results show that as expected the avirulence genes *AvrLm1* and *AvrLm6* were predicted in the PAV read comparisons. *AvrLm1* (Lema_G049660.1) and *AvrLm6* (Lema_G049940.1) are located on SC6 ~ 200 Kbp apart. The prediction results can be seen in table 5.12.

Table 5.12: *AvrLm* Gene Prediction.

Comparison	*AvrLm* Gene	Correct
D5vD1	*AvrLm1*	Y
D6vD1	*AvrLm1*	Y
D5vD2	*AvrLm1*	Y
D6vD2	*AvrLm1*	Y
D5vD4	*AvrLm1*	Y
D6vD4	*AvrLm1*	Y
D5vD9	*AvrLm1*	Y
D5vD12	*AvrLm1*	Y
D6vD8	*AvrLm1*	Y
D6vD9	*AvrLm1*	Y
D6vD12	*AvrLm1*	Y
D5vD10	*AvrLm1*	Y
D6vD10	*AvrLm1*	Y
D6vD3	*AvrLm1*	Y
	AvrLm6	N
D1vD3	*AvrLm6*	N
D8vD3	*AvrLm6*	Y
D12vD3	*AvrLm6*	Y

Based on table 6.1, 16 of the 18 (89%) PAV predictions for these avirulence genes were shown to be correct, based on this data alone. The other characterised avirulence genes *AvrLm4-7, AvrLmJ1, AvrLm2* and *AvrLm11* were not predicted to be present/absent in this analysis. No PAVs were predicted for any of these avirulence genes. D6 and D3 isolates appear to differ in a short stretch of sequence within the third intron of the *AvrLm6* gene. D1 and D3 isolates also seem to have a small stretch of sequence difference within the first intron of *AvrLm6*.

5.5. Discussion

5.5.1. Validation

An analysis of PAVs within *L. maculans* isolates, D1 and D5, which are routinely used to determine blackleg resistance gene composition in canola in Australia, was conducted. Paired-end reads, which were generated on an Illumina GAIIx platform from isolated DNA from D1 and D5, were used. The analysis testing this PAV method, showed that the predicted variations could be validated by PCR ~92% of the time.

Using PAKAP in *L. maculans* shows that the majority of predicted PAVs are indeed present/absent or highly diverged. Interestingly, Multi-Nucleotide Polymorphisms (MNPs) constituted a part of this divergence, and their impact and presence requires further study in *L. maculans*, as well as in other species (Rosenfeld et al., 2010). The presence/absence could be seen by the fact that the PCR did not amplify the product for D1 and successfully amplified the product in D5 and vice versa. In the case where the products were amplified in both samples, the one predicted to be absent showed high sequence divergence compared to the other sample. This indicates that the PAKAP method can also identify regions in the genome that show great sequence variation and are therefore considered present/absent when the datasets are compared to each other. One of the validation primer sets (D5 primer product 7) was disregarded due to poor sanger sequence quality from PCR amplification (see section 5.3.2). One primer pair showed a false positive PCR result, however sequence read mapping suggests that this gene is indeed absent in D1 and present in D5 (see figures 5.5 and 5.6). This indicates that more work needs to be performed to refine the PCR validation of the PAKAP method.

This approach has the potential to identify genomic regions that are undergoing diversification and show high levels of polymorphism across members of the same species. Certain gene types undergoing selection, that are involved in gains in fitness, can be identified in this manner by comparing the present/absent regions between individuals.

5.5.2. Differential Isolate Analysis Discussion

The sequence coverage achieved for these ten isolates was considerably high (table 5.4), ranging from 10.65 times (x) coverage (isolate D10 after trimming), to 266.3x coverage

(isolate D1). The average genome coverage was 99.65x. This provides confidence that, excluding isolate D10, the polymorphic presence/absence predictions are in fact due to sequence divergence, not merely a lack of reads for a certain portion of the genome.

The PAV prediction results show that considerable sequence variability exists between individual genomes of *L. maculans* differential isolates (table 5.6).

The D1 isolate yielded more PAV reads than most other comparisons (an average of 27,807 PAVs predicted), which can be explained by the fact that it also has the highest sequence coverage of all datasets and is thus more likely to contain unique reads. Of all predicted D1 PAV reads ~26% mapped to the reference genome (table 6.7). The D6 isolate had the highest number of predicted PAV reads of all isolates (with a staggering average number of PAVs of ~614,000), yet only ~0.33% of these mapped to the genome. This suggests that these reads are highly diverged from the reference genome but may also contain more errors which can prevent SOAP mapping. Indeed, D6 PAV reads account for over 88% of the total number of predicted PAVs (5,526,502 out of 6,266,349). Excluding D6 PAV numbers, the overall percentage of PAV reads mapping to the reference is actually ~15% (not the average ~2%). Another consideration in this analysis is the SOAP mapping parameters. The –M parameter used here is 0, which allows no mismatches to the genome. Permitting one or two mismatches to the genome would result in higher read mapping results. The most likely explanation however, is that the D6 dataset was contaminated by other DNA such as plant or bacterial DNA (both of which are ubiquitous in our laboratory environment).

5.5.3. D10 Comparisons

The D10 isolate, while containing a significant amount of read coverage across the whole genome, had the lowest number of predicted PAV reads (unique reads), compared with all other isolates (tables 5.6 and 5.7) at just 42.7 PAVs on average. The other isolates also had some of the highest predicted PAV read numbers compared to D10. This is likely due to the fact that the sequence reads were trimmed from 250 bp to 100 bp, thus reducing the total pool of data significantly. Trimming was deemed necessary as the 250 bp reads resulted in poor PAV prediction, due to a drop off in quality at the 3’ end of the reads, a common problem in longer Illumina sequence reads. A further problem may have been the different read length, which may have caused problems in the PAV prediction.

Already having the lowest overall genome coverage of 26.7x (table 6.2), trimming effectively reduced the coverage to 10.65x genome coverage, resulting in less data with which PAV reads could be predicted. In future studies, more sequence data for isolate D10 of similar size and quality as the other isolates will be necessary in order to achieve better results.

5.5.4. PAV Read Hits to Gene Annotations

Among the 45 PAV comparisons (D1vD2, D2vD1 etc), the SOAP mapping pointed to a large number of PAV reads mapping to gene annotations, which were termed gene hits. The average number of gene hits per SC was ~32.

PAV reads mapping to annotated genes in *L. maculans* (gene hits) seem to be present/absent in clusters across SCs (table 5.10), which indicates that specific genomic areas are undergoing selection pressures and have larger variations of several hundred base pairs. This mirrors the results of Golicz et al. (2015), which found that neighbouring genes were often lost together in the same *L. maculans* isolate. Chromosomal rearrangements may also account for neighbouring genes being affected by gene loss. Five of the genes with PAV hits were also identified to be lost between isolates 21, 41 and the *L. maculans* reference genome by Golicz et al. (2015), indicating that the PAKAP method is identifying true lost genes.

Three quarters of the gene annotations with PAV mapping have no predicted function from which effects can be deduced. The avirulence genes *AvrLm1* and *AvrLm6* were predicted along with other infection related genes, such as genes related to antifungal and antibiotic resistance. A number of secondary metabolite genes were also predicted to be present/absent between differential isolates. This indicates that certain classes of genes, such as those involved in pathogenicity and secondary metabolite production, may be subject to increased selection pressures in *L. maculans*. Furthermore, the metabolic genes that showed PAV read hits, including aspartate aminotransferase, cellulase and fatty acid synthase, belong to multi-domain families which exhibit complex functional redundancy (Amore et al., 2013; Lal et al., 2014; Schweizer and Hofmann, 2004). This may allow for the function of these genes to be complemented by other genes within the pathway.

Interestingly, the Lema_G082090.1 gene on SC13 with similarity to the beta subunit of fatty acid synthase was also identified in chapter 4 (section 4.3.4) as being divergent in the

differential isolates, with a high number of SNPs predicted within this gene. The PAKAP method therefore shows some overlap with the SGSautoSNP prediction when a high number of polymorphisms are present at a particular locus.

Of particular interest was SC13, which revealed 264 gene hits, over 8 times the average number of gene hits compared to the other SCs. SC13 also had the highest number of genes affected by PAVs, with 24 (22.4%) of the total genes being located there.

The gene hits are localised in a cluster on SC13 which spans ~116 Kbp and contains 48 genes of which exactly half revealed gene hits (table 5.11). This suggests that this region is undergoing selective pressure which is significantly higher than other areas of the genome, particularly considering that the variations are not merely SNPs but larger variations of several hundred bases. A study conducted in *Parastagonospora nodurum* found a similar pattern, which was termed sectional gene absence (Syme et al., 2013).

The sirodesmin biosynthetic cluster, described by Gardiner et al. (2004), is situated the middle of the region of high PAV gene hits on SC13. There are 23 genes in this cluster, and of these, four were found to have PAV gene hits, namely *sirH*, *LmUVI-1h*, *LmPKS1* and *LmHDX* (table 5.11). Sirodesmin PL is a fungal phytotoxin produced by pathogens of both plants and animals and is not host specific (Gardiner et al., 2004). In the *L. maculans* – *Brassica* pathosystem, sirodesmin PL causes lesions on plant leaves and has antibacterial and antiviral properties (Rouxel et al., 1988). The *LmUVI-1h*, *LmPKS1* and *LmHDX* genes in this cluster, are not part of the sirodesmin biosynthesis pathway but have important molecular functions. *LmHDX* encodes for a hydroxylase and *LmPKS1* encodes a polyketide synthase (a large class of secondary metabolites). *LmUVI-1h* is expressed when exposed to UV light and during the formation of an appressorium, a specialised hyphal organ that fungal pathogens use to enter plants. The *sirH* gene is believed to encode an acetyl transferase, an essential step required in the biosynthesis of sirodesmin. A simple sequence search revealed that there are multiple copies of *LmPKS1* in the *L. maculans* genome, but only one copy of both *LmUVI-1h* and *LmHDX.* This may suggest that the presence/absence of the *LmPKS1* gene may not have a significant effect on *L. maculans* due to functional redundancy within the polyketide synthase family. It appears that both *LmUVI-1h* and *LmHDX* are single-copy genes in *L. maculans*, which could have a deleterious effect on growth and infection success if absent. Interestingly, a cluster of

genes, including a polyketide synthase, was also found to be lost in a comparison of isolates in *Parastagonospora nodurum* (Syme et al., 2013).

Similar to the findings of Tan et al. (2012) for disease related genes in plants, the high number of predicted PAVs on SC13 may indicate that genes vital for infection success of this pathogen are undergoing selective pressures that are higher than elsewhere in the genome.

5.5.5. PAV prediction in *AvrLm* genes

Both the avirulence genes *AvrLm1* and *AvrLm6* were shown to be present/absent in a number of isolate comparisons. This is in line with previous studies that have found these two genes to be lost under selective pressure (Raman et al., 2012; Van de Wouw et al., 2010; Zander et al., 2013). The avirulence genes *AvrLm4-7, AvrLmJ1, AvrLm2* and *AvrLm11* were not predicted to be present/absent in this analysis. *AvrLm11* resides on a disposable chromosome and shows a presence/absence genotype (Balesdent et al., 2013), however this analysis did not predict its presence or absence in any of the isolates tested. The mechanism of *AvrLm2* and *AvrLm4-7* avirulence/virulence has been shown to be SNPs (Ghanbarnia et al., 2014; Parlange et al., 2009; Van de Wouw and Howlett, 2012) and was thus not picked up in this analysis. The *Brassica juncea* specific *AvrLmJ1* becomes virulent through a premature stop codon in the protein (Van de Wouw et al., 2014).

Table 5.12 describes the PAV results for *AvrLm1* and *AvrLm6.* Isolates D5 and D6 have been previously shown to contain *AvrLm1* through infection studies with plants containing the *Rlm1* resistance genes (table 5.1). The results correctly predict that D5 and D6 have *AvrLm1* whereas D1, D2, D3, D4, D8, D9, D10 and D12 all do not (table 5.12). In fact, the only missing comparisons for *AvrLm1* presence/absence are D5vD8 and D5vD3, meaning the PAV prediction successfully predicted 14/16 PAVs (87.5%) for *AvrLm1*. The fact that the D6vD8 and D6vD3 comparisons yielded the correct prediction, allows confidence that the missing comparisons are the outlier in this analysis. A possible reason for the absence of PAV prediction in the D5vD8 and D5vD3 comparisons, is the SNP identified within the *AvrLm1* gene in D5 at position 1,607,437, described in chapter 4 (section 4.3.5). This may have interfered with the PAKAP method in identifying the PAVs for this gene.

The results for *AvrLm6* presence/absence are not as clear. The PAV prediction incorrectly predicts *AvrLm6* being absent in D3 compared to D1 and D6 (table 5.12). However D3 also contains *AvrLm6*, based on past studies of this isolate. Sequence reads for the D3 isolate seems to differ to D6 within the third intron of *AvrLm6*, which may have caused this false negative prediction. D1 and D3 isolates also seem to differ within the first intron of *AvrLm6*. SNP data from chapter 4, suggests that there are a seven SNPs within up- and downstream regions of the *AvrLm6* genes which may have also affected PAV prediction.

Interestingly, in this study D8 and D12 are predicted to contain *AvrLm6* (although it should be absent based on infection studies) compared to D3 (which does contain it). It may be possible that the *AvrLm6* gene sequence is highly diverged in D8 and D12 and not completely absent. This sequence divergence therefore leads to the PAKAP method predicting a presence/absence allele for this gene between D8, D12 and D3. It may be helpful in future, to refer to PAV reads as cultivar specific reads, as the sequence may be highly diverged and not in fact absent.

This indicates that the PAV prediction may have the potential to confirm the correct *AvrLm* gene content within isolates of *L. maculans*, based on Illumina sequence data.

5.6. Conclusions:

This Presence/Absence Variation prediction approach was used on *L. maculans* sequence data to identify PAV reads between the datasets. The results show that there are a large number of variable regions in *L. maculans*, spread across the genome. This suggests that the genome is under selective pressures in certain regions, particularly on SC13 where several metabolic genes are localised. Interestingly, the high number of variants also indicates that there may be considerable functional redundancy in certain metabolic pathways but also potentially deleterious variations, especially in single copy genes. A number of infection related genes also seem to be undergoing selection, as well as genes related to antifungal and antibiotic resistance. The PAVs involving avirulence genes highlight the utility of this approach in providing a novel tool to identify the presence/absence of avirulence genes known to be avirulent/virulent.

6. Discussion and Final Thoughts

The analyses in this thesis have allowed for several new insights into the structure and organisation of the *L. maculans* genome. It has been shown that NGS technologies have the ability to provide large amounts of genomic sequence data. This availability of large datasets provides us with a novel resource, from which analyses of genome structure and diversity can be undertaken. A high amount of sequence coverage was achieved within the differential isolates described in chapters 4 and 5, as well as the isolates described in chapter 3, which allowed for confidence in the prediction of polymorphisms between them. The specific objectives described in section 1.8 were achieved in these analyses and are discussed in detail below.

Chapter 3 demonstrated that the SGSautoSNP (Lorenc et al., 2012) method is useful and applicable for SNP prediction and validation within small fungal genomes such as *L. maculans.* These polymorphisms can be used as novel markers for molecular and phylogenetic analysis. The re-sequencing of two *L. maculans* isolates (isolates 21 and 41) resulted in the identification of 21,814 SNPs. These were used for genome analysis, in order to determine SNP properties and their density in relation to genomic position and predicted function. This method also correctly predicted polymorphisms within *AvrLm* genes. This shows that re-sequencing, together with SNP prediction, can be used to validate known mutations within genomes, as well as identify new polymorphisms that cause phenotypic change. Overall whole-genome trends such as transition/transversion ratios, polymorphism directionality and genome-wide SNP density can also be determined using this method. SGSautoSNP was demonstrated to be robust and sensitive in the identification of SNPs in *L. maculans*. The SNPs identified in this analysis were further utilised by Patel et al. (2015), where a subset of the total SNPs was amplified across 59 isolates from around Australia. The Illumina GoldenGate assay was used to provide genotypic information for 384 SNPs. The population of Australian *L. maculans* isolates was found to be panmictic, meaning high rates of sexual reproduction and evolutionary diversification exist within this pathogen. These features make it possible for *L. maculans* to overcome host resistance, as mechanisms exist to share genetic information that provide fitness advantages across large populations.

SNP prediction was conducted on a larger scale in chapter 4. Ten differential isolates were used, with known avirulence gene information from infection studies toward *Brassica* species. These isolates were re-sequenced and yielded (for the most part) very high levels of sequence coverage across the ~45.12 Mbp genome of *L. maculans*. The SGSautoSNP pipeline was again utilised and resulted in the prediction of 47,097 SNPs between the differential isolates. This is more than double the number of SNPs previously predicted in chapter 3. The SNP density was also doubled with an average of one SNP every 953 bp, compared to one SNP every 2,065 bp (chapter 3) across the first 40 SuperContigs (SCs). These SNPs provide a vast marker resource for genetic studies in *L. maculans*.

Importantly, the SGSautoSNP prediction method successfully predicted the correct *AvrLm* gene complements within differential isolates and identified previously validated causative virulent/avirulent SNPs. It also provides a method of testing and validating infection data by genotyping the differential isolates for known *AvrLm* genes. For example, the data from this analysis suggests that isolate D1, believed to be avirulent for the *AvrLm2* gene, in fact exhibits the virulent alleles for this gene. In this way, phenotypic and genotypic studies can be matched to increase confidence in plant-fungal resistance analyses.

A number of SNPs were identified that cause non-synonymous coding changes within *L. maculans* genes. These were shown to affect a range of genes with varying functions, including those involved in plant infection processes. Of particular interest are genes similar to those from *Aspergillus* fungal species, which produce and regulate toxic secondary metabolites related to virulence. The dicer-like protein also seems to be highly divergent in *L. maculans* isolates, which is interesting considering it regulates and initiates defence toward viruses and transposons (Kadotani et al., 2004). The latter are found in high abundance with *L. maculans*, indicating that perhaps the mechanisms for their defence are under ongoing selection.

A number of SSPs were selected and found to contain SNPs within coding regions. These were selected from a list of genes with putative avirulence gene characteristics (Rouxel et al., 2011). They are small secreted proteins (SSPs), with both transcriptomic support and overexpression upon infection of the host. These SSPs will be followed up as avirulence or disease-related gene candidates in the future.

The novel method of predicting presence/absence variations (PAVs), discussed in chapter 5, has been a further step forward in the understanding of the *L. maculans* genome. Larger scale sequence differences can be studied using this method, which will aid in determining genes that are being lost or are highly diverged across different *L. maculans* isolates. The PAKAP method was applied to the ten *L. maculans* isolates previously used for SNP analysis. The results indicate that there are a number of variable regions in *L. maculans* that seem to be quite diverse across the differential isolates. The clustering of PAV gene hits suggests that certain genomic regions are under particular selective pressures. This is especially evident on SC13, where a number of secondary metabolite genes, some involved in plant infection processes, are located. A number of other genes of interest were also shown to be affected by PAVs, with genes involved in fatty acid metabolism and antibiotic and antifungal resistance also showing presence/absence variation in this study. The PAV prediction method was also effective in identifying the presence/absence of avirulence genes, known to be present or lost from previous studies, such as *AvrLm1* and *AvrLm6*.

Future analyses should focus on PAV reads that didn't map to the reference genome, as they may be absent in the JN3 (also known as v23.1.3) reference isolate. This will provide greater insights into present/absent genes across the differential isolates and allow for the identification of novel genes that are present in other isolates other than the reference. These could include be a number of genes of interest, including metabolic genes and novel avirulence genes. The study by Golicz et al. (2015) used a novel gene-loss prediction method called SGSGeneloss to determine genes lost between isolates 21, 41 and the *L. maculans* reference genome. De novo assembly of isolates 21 and 41 was also performed to identify genes not present in the reference, and a number of these were identified. A similar approach could be used to increase the reach of the PAV analysis in *L. maculans*.

An improvement in the functional annotation of genes, should be of high priority for the ongoing development of the reference genome of *L. maculans*. This will lead to more meaningful analyses of genetic diversity within this fungus, as well as shedding light on the effect of polymorphisms on phenotype, particularly traits involved in the plant-pathogen interaction and ETI. A further consideration may be the development of a pan-genome or supra-genome in *L. maculans*. This means creating a genome that includes all genomic

sequences available within a species, including those that are dispensable or present/absent between different individuals of the species (Medini et al., 2005; Tettelin et al., 2005). This genome would be significantly larger than the core genome, however it can provide a vastly improved understanding of intra-species diversity.

A combination of the novel genomic SNP and PAV markers will provide a resource for understanding the inheritance of disease related genes, especially avirulence genes. In this thesis it was shown that SNPs affecting recognition of avirulence genes by the host, could be successfully predicted using the SGSautoSNP method. Furthermore, avirulence genes that are known to be present/absent could be validated using the novel PAV prediction method, PAKAP.

By combining these novel methods, new disease related genes, such as avirulence genes, can be predicted in *L. maculans*. Existing genes, such as those analysed in this thesis, can also be studied in order to develop a better understanding of the *Brassica*-blackleg interaction. This is turn has the capacity to improve the breeding efforts towards resistant plant cultivars that provide reduced need for disease reduction processes such as fungicide application. Less disease leads to enhanced plant growth, improved yields and ultimately the viability and sustainability of the canola industry in Australia and internationally.

Looking towards the future, these methods can be applied to other organisms and other pathosystems. With further development of the PAV pipeline, this type of analysis will be possible in species with no reference genome available. This has the potential to aid in the understanding of pathogens, and the protection of plant species, that are currently not well studied.

List of References

Abdallah, N.A., Shah, D., Abbas, D. and Madkour, M. (2010) Stable integration and expression of a plant defensin in tomato confers resistance to fusarium wilt. *GM Crops* **1**, 344-350.

Ajilogba, C.F. and Babalola, O.O. (2013) Integrated management strategies for tomato Fusarium wilt. *Biocontrol Sci* **18**, 117-127.

Amore, A., Giacobbe, S. and Faraco, V. (2013) Regulation of cellulase and hemicellulase gene expression in fungi. *Current genomics* **14**, 230-249.

Ansan-Melayah, D., Balesdent, M.H., Delourme, R., Pilet, M.L., Tanguy, X., Renard, M. and Rouxel, T. (1998) Genes for race-specific resistance against blackleg disease in Brassica napus L. *Plant Breeding* **117**, 373-378.

AOF (2015) Australian Oilseeds Federation. www.australianoilseeds.com.

Appleby, N., Edwards, D. and Batley, J. (2009) New technologies for ultra-high throughput genotyping in plants. In: *Plant Genomics Methods and Protocols* (Somers, D.J., Langridge, P. and Gustafson, J.P. eds), pp. 19-40. UK: Humana Press.

Araki, H., Tian, D., Goss, E.M., Jakob, K., Halldorsdottir, S.S., Kreitman, M. and Bergelson, J. (2006) Presence/absence polymorphism for alternative pathogenicity islands in Pseudomonas viridiflava, a pathogen of Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5887-5892.

Aranda, J., Cerqueira, N.M., Fernandes, P.A., Roca, M., Tunon, I. and Ramos, M.J. (2014) The catalytic mechanism of carboxylesterases: a computational study. *Biochemistry* **53**, 5820-5829.

Balesdent, M.H., Attard, A., Ansan-Melayah, D., Delourme, R., Renard, M. and Rouxel, T. (2001) Genetic control and host range of avirulence toward Brassica napus cultivars Quinta and Jet Neuf in Leptosphaeria maculans. *Phytopathology* **91**, 70-76.

Balesdent, M.H., Attard, A., Kuhn, M.L. and Rouxel, T. (2002) New Avirulence Genes in the Phytopathogenic Fungus Leptosphaeria maculans. *Phytopathology* **92**, 1122-1133.

Balesdent, M.H., Fudal, I., Ollivier, B., Bally, P., Grandaubert, J., Eber, F., Chevre, A.M., Leflon, M. and Rouxel, T. (2013) The dispensable chromosome of Leptosphaeria maculans shelters an effector gene conferring avirulence towards Brassica rapa. *New Phytol* **198**, 887-898.

Barrins, J.M., Ades, P.K., Salisbury, P.A. and Howlett, B.J. (2004) Genetic diversity of Australian isolates of Leptosphaeria maculans, the fungus that causes blackleg of canola (Brassica napus). *Australasian Plant Pathology* **33**, 529-536.

Batley, J. and Edwards, D. (2007) SNP applications in plants. In: *Association Mapping in Plants* (Oraguzie, N.C., Rikkerink, E.H.A., Gardiner, S.E. and De Silva, H.N. eds), pp. 95-102. New York: Springer

Batley, J. and Edwards, D. (2009) Genome sequence data: management, storage, and visualization. *Biotechniques* **46**, 333-+.

Batley, J., Jewell, E. and Edwards, D. (2007) Automated discovery of Single Nucleotide Polymorphism (SNP) and Simple Sequence Repeat (SSR) molecular genetic markers. In: *Plant bioinformatics* (Edwards, D. ed) pp. 473-494. USA: Methods in molecular biology, Humana Press.

Belo, A., Beatty, M.K., Hondred, D., Fengler, K.A., Li, B. and Rafalski, A. (2010) Allelic genome structural variations in maize detected by array comparative genome hybridization. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **120**, 355-367.

Bent, A.F. (1996) Plant disease resistance genes: Function meets structure. *Plant Cell* **8**, 1757-1771.

Blatch, G.L. and Lassle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays : news and reviews in molecular, cellular and developmental biology* **21**, 932-939.

Blondeau, K., Blaise, F., Graille, M., Kale, S.D., Linglin, J., Ollivier, B., Labarde, A., Lazar, N., Daverdin, G., Balesdent, M.H., Choi, D.H., Tyler, B.M., Rouxel, T., van Tilbeurgh, H. and Fudal, I. (2015) Crystal structure of the effector AvrLm4-7 of Leptosphaeria maculans reveals insights into its translocation into plant cell and recognition by resistance proteins. *The Plant journal : for cell and molecular biology*.

Bohnert, H.U., Fudal, I., Dioh, W., Tharreau, D., Notteghem, J.L. and Lebrun, M.H. (2004) A putative polyketide synthase/peptide synthetase from Magnaporthe grisea signals pathogen attack to resistant rice. *Plant Cell* **16**, 2499-2513.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) Construction of a Genetic-Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *American Journal of Human Genetics* **32**, 314-331.

Boyle, J.S. and Lew, A.M. (1995) An Inexpensive Alternative to Glassmilk for DNA Purification. *Trends in Genetics* **11**, 8-8.

Broders, K.D., Woeste, K.E., SanMiguel, P.J., Westerman, R.P. and Boland, G.J. (2011) Discovery of single-nucleotide polymorphisms (SNPs) in the uncharacterized genome of the ascomycete *Ophiognomonia clavigignenti-juglandacearum* from 454 sequence data. *Molecular Ecology Resources* **11**, 693-702.

Brown, D.W., Adams, T.H. and Keller, N.P. (1996) Aspergillus has distinct fatty acid synthases for primary and secondary metabolism. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 14873-14877.

Brun, H., Chevre, A.M., Fitt, B.D.L., Powers, S., Besnard, A.L., Ermel, M., Huteau, V., Marquer, B., Eber, F., Renard, M. and Andrivon, D. (2010) Quantitative resistance increases the durability of qualitative resistance to Leptosphaeria maculans in Brassica napus. *New Phytologist* **185**, 285-299.

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 Biology* **5**, 501-514.

Bush, S.J., Castillo-Morales, A., Tovar-Corona, J.M., Chen, L., Kover, P.X. and Urrutia, A.O. (2014) Presence-absence variation in A. thaliana is primarily associated with

genomic signatures consistent with relaxed selective constraints. *Mol Biol Evol* **31**, 59-69.

Casci, T. (2010) Population Genetics Snps That Come in Threes. *Nature Reviews Genetics* **11**.

Castle, J.C. (2011) SNPs Occur in Regions with Less Genomic Sequence Conservation. *PloS one* **6**.

Catanzariti, A.M., Dodds, P.N., Lawrence, G.J., Ayliffe, M.A. and Ellis, J.G. (2006) Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* **18**, 243-256.

CCC (2015) Canola Council of Canada. In: *www.canolacouncil.org*.

Chagné, D., Batley, J., Edwards, D. and Forster, J.W. (2007) Single Nucleotide Polymorphisms genotyping in plants. In: *Association mapping in plants* (Oraguzie, N.C., Rikkerink, E.H.A., Gardiner, S.E. and De Silva, H.N. eds), pp. 77-94. New York: Springer.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Correa, M., Da Silva, C., Just, J., Falentin, C., Koh, C.S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, A., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alamery, S., Jabbari, K., Zhao, M., Edger, P.P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M.C., Fan, G., Renault, V., Bayer, P.E., Golicz, A.A., Manoli, S., Lee, T.H., Thi, V.H., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom, C.H., Wang, X., Canaguier, A., Chauveau, A., Berard, A., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y.P., Lyons, E., Town, C.D., Bancroft, I., Wang, X., Meng, J., Ma, J., Pires, J.C., King, G.J., Brunel, D., Delourme, R., Renard, M., Aury, J.M., Adams, K.L., Batley, J., Snowdon, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H., Guan, C. and Wincker, P. (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science* **345**, 950-953.

Chen, X., Shang, J., Chen, D., Lei, C., Zou, Y., Zhai, W., Liu, G., Xu, J., Ling, Z., Cao, G., Ma, B., Wang, Y., Zhao, X., Li, S. and Zhu, L. (2006) A B-lectin receptor kinase gene conferring rice blast resistance. *The Plant journal : for cell and molecular biology* **46**, 794-804.

Chevre, A.M., Eber, F., This, P., Barret, P., Tanguy, X., Brun, H., Delseny, M. and Renard, M. (1996) Characterization of Brassica nigra chromosomes and of blackleg resistance in B. napus–B. nigra addition lines. *Plant Breeding* **115**, 113-118.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X.Y. 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 w(1118); iso-2; iso-3. *Fly* **6**, 80-92.

Costa, C., Dias, P.J., Sa-Correia, I. and Teixeira, M.C. (2014) MFS multidrug transporters in pathogenic fungi: do they have real clinical impact? *Frontiers in physiology* **5**, 197.

Coulondre, C., Miller, J.H., Farabaugh, P.J. and Gilbert, W. (1978) Molecular basis of base substitution hot spots in Escherichia coli. *Nature* **274**, 775-780.

Cuomo, C.A., Gueldener, U., Xu, J.R., Trail, F., Turgeon, B.G., Di Pietro, A., Walton, J.D., Ma, L.J., Baker, S.E., Rep, M., Adam, G., Antoniw, J., Baldwin, T., Calvo, S., Chang, Y.L., DeCaprio, D., Gale, L.R., Gnerre, S., Goswami, R.S., Hammond-Kosack, K., Harris, L.J., Hilburn, K., Kennell, J.C., Kroken, S., Magnuson, J.K., Mannhaupt, G., Mauceli, E., Mewes, H.W., Mitterbauer, R., Muehlbauer, G., Munsterkotter, M., Nelson, D., O'Donnell, K., Ouellet, T., Qi, W.H., Quesneville, H., Roncero, M.I.G., Seong, K.Y., Tetko, I.V., Urban, M., Waalwijk, C., Ward, T.J., Yao, J.Q., Birren, B.W. and Kistler, H.C. (2007) The Fusarium graminearum genome reveals a link between localized polymorphism and pathogen specialization. *Science* **317**, 1400-1402.

Dangl, J.L. and Jones, J.D.G. (2001) Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826-833.

Daverdin, G., Rouxel, T., Gout, L., Aubertot, J.N., Fudal, I., Meyer, M., Parlange, F., Carpezat, J. and Balesdent, M.H. (2012) Genome Structure and Reproductive Behaviour Influence the Evolutionary Potential of a Fungal Phytopathogen. *Plos Pathogens* **8**.

Delourme, R., Chèvre, A.M., Brun, H., Rouxel, T., Balesdent, M.H., Dias, J.S., Salisbury, P., Renard, M. and Rimmer, S.R. (2006) Major Gene and Polygenic Resistance to Leptosphaeria maculans in Oilseed Rape (Brassica napus). *European Journal of Plant Pathology* **114**, 41-52.

Delourme, R., Pilet-Nayel, M.L., Archipiano, M., Horvais, R., Tanguy, X., Rouxel, T., Brun, H., Renard, M. and Balesdent, M.H. (2004) A Cluster of Major Specific Resistance Genes to Leptosphaeria maculans in Brassica napus. *Phytopathology* **94**, 578-583.

Dion, Y., Gugel, R.K., Rakow, G.F., Seguin-Swartz, G. and Landry, B.S. (1995) RFLP mapping of resistance to the blackleg disease [causal agent, Leptosphaeria maculans (Desm.) Ces. et de Not.] in canola (Brassica napus L.). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **91**, 1190-1194.

Dodds, P. and Thrall, P. (2009) Recognition events and host-pathogen co-evolution in gene-for-gene resistance to flax rust. *Funct Plant Biol* **36**, 395-408.

Dodds, P.N., Lawrence, G.J., Catanzariti, A.M., Ayliffe, M.A. and Ellis, J.G. (2004) The Melampsora lini AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell* **16**, 755-768.

Dodds, P.N., Lawrence, G.J. and Ellis, J.G. (2001) Six amino acid changes confined to the leucine-rich repeat beta-strand/beta-turn motif determine the difference between the P and P2 rust resistance specificities in flax. *Plant Cell* **13**, 163-178.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T. and Wilson, A. (2012) Geneious Pro v5.6. http://www.geneious.com.

Duran, C., Appleby, N., Clark, T., Wood, D., Imelfort, M., Batley, J. and Edwards, D. (2009a) AutoSNPdb: an annotated Single Nucleotide Polymorphism database for crop plants. *Nucleic Acids Research* **37**, 951-953.

Duran, C., Appleby, N., Edwards, D. and Batley, J. (2009b) Molecular genetic markers: discovery, applications, data storage and visualisation. *Current Bioinformatics* **4**, 16-27.

Duran, C., Boskovic, Z., Imelfort, M., Batley, J., Hamilton, N.A. and Edwards, D. (2010a) CMap3D: A 3D visualisation tool for comparative genetic maps. *Bioinformatics* **26**, 273-274.

Duran, C., Eales, D., Marshall, D., Imelfort, M., Stiller, J., Berkman, P.J., Clark, T., McKenzie, M., Appleby, N., Batley, J., Basford, K. and Edwards, D. (2010b) Future tools for association mapping in crop plants. *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada* **53**, 1017-1023.

Edwards, D. (2007) Bioinformatics and plant genomics for staple crops improvement. In: *Breeding major food staples* (Kang, M.S. and Priyadarshan, P.M. eds), pp. 93-106. Blackwell.

Edwards, D. and Batley, J. (2004) Plant bioinformatics: from genome to phenome. *Trends in Biotechnology* **22**, 232-237.

Edwards, D., Batley, J., Cogan, N.O.I., Forster, J.W. and Chagné, D. (2007a) Single Nucleotide Polymorphism discovery. In: *Association mapping in plants* (Oraguzie, N.C., Rikkerink, E.H.A., Gardiner, S.E. and De Silva, H.N. eds). New York: Springer.

Edwards, D., Batley, J. and Snowdon, R.J. (2013) Accessing complex crop genomes with next-generation sequencing. *Theoretical and Applied Genetics* **126**, 1-11.

Edwards, D., Burton, W.A., Hopkins, C.J. and Batley, J. (2007b) Indian mustard. In: *Genome mapping and molecular breeding in plants* (Kole, C. ed) pp. 179-210. New York: Springer.

Edwards, D., Forster, J.W., Chagné, D. and Batley, J. (2007c) What are SNPs? In: *Association mapping in plants* (Oraguzie, N.C., Rikkerink, E.H.A., Gardiner, S.E. and Silva, H.N.d. eds), pp. 41-52. New York: Springer.

Edwards, D., Zander, M., Dalton-Morgan, J. and Batley, J. (2012) New technologies for ultra-high throughput genotyping in plant taxonomy. In: *Molecular Plant Taxonomy* (Besse, P. ed).

Ellis, J.G., Lawrence, G.J., Luck, J.E. and Dodds, P.N. (1999) Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. *Plant Cell* **11**, 495-506.

Ellis, J.G., Rafiqi, M., Gan, P., Chakrabarti, A. and Dodds, P.N. (2009) Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. *Current opinion in plant biology* **12**, 399-405.

Erwin, T., Jewell, E., Love, C., Lim, G., Li, X., Chapman, R., Batley, J., Stajich, J., Mongin, E., Stupka, E., Ross, B., Spangenberg, G.C. and Edwards, D. (2007) BASC: an integrated bioinformatics system for *Brassica* research. *Nucleic Acids Research* **35**, D870-D873.

Ferreira, M.E., Satagopan, J., Yandell, B.S., Williams, P.H. and Osborn, T.C. (1995) Mapping loci controlling vernalization requirement and flowering time in Brassica napus. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **90**, 727-732.

Feuk, L., Carson, A.R. and Scherer, S.W. (2006) Structural variation in the human genome. *Nature reviews. Genetics* **7**, 85-97.

Fitt, B.D.L., Brun, H., Barbetti, M.J. and Rimmer, S.R. (2006) World-wide importance of phoma stem canker (Leptosphaeria maculans and L. biglobosa) on oilseed rape (Brassica napus). In: *Sustainable strategies for managing Brassica napus (oilseed rape) resistance to Leptosphaeria maculans (phoma stem canker)* (Fitt, B.D.L., Evans, N., Howlett, B.J. and Cooke, B.M. eds), pp. 3-15. Springer Netherlands.

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

Forche, A., Magee, P.T., Magee, B.B. and May, G. (2004) Genome-wide single-nucleotide polymorphism map for Candida albicans. *Eukaryotic Cell* **3**, 705-714.

Francis, A.R. and Tanaka, M.M. (2012) Evolution of variation in presence and absence of genes in bacterial pathways. *BMC evolutionary biology* **12**, 55.

Friedt, W. and Snowdon, R. (2009) Oilseed rape. In: *Oil crops* (Vollmann, J. and Rajcan, I. eds). Springer Science & Business Media.

Fudal, I., Ross, S., Brun, H., Besnard, A.-L., Ermel, M., Kuhn, M.-L., Balesdent, M.-H. and Rouxel, T. (2009) Repeat-Induced Point Mutation (RIP) as an Alternative Mechanism of Evolution Toward Virulence in Leptosphaeria maculans. *Molecular Plant-Microbe Interactions* **22**, 932-941.

Fudal, I., Ross, S., Gout, L., Blaise, F., Kuhn, M.L., Eckert, M.R., Cattolico, L., Bernard-Samain, S., Balesdent, M.H. and Rouxel, T. (2007) Heterochromatin-like regions as ecological niches for avirulence genes in the Leptosphaeria maculans genome: Map-based cloning of AvrLm6. *Molecular Plant-Microbe Interactions* **20**, 459-470.

Fulton, T.M., Chunwongse, J. and Tanksley, S.D. (1995) Microprep Protocol for Extraction of DNA from Tomato and Other Herbaceous Plants. *Plant Molecular Biology Reporter* **13**, 207-209.

Gardiner, D.M., Cozijnsen, A.J., Wilson, L.M., Pedras, M.S. and Howlett, B.J. (2004) The sirodesmin biosynthetic gene cluster of the plant pathogenic fungus Leptosphaeria maculans. *Molecular microbiology* **53**, 1307-1318.

Gassmann, W. and Bhattacharjee, S. (2012) Effector-Triggered Immunity Signaling: From Gene-for-Gene Pathways to Protein-Protein Interaction Networks. *Molecular Plant-Microbe Interactions* **25**, 862-868.

Gawehns, F., Houterman, P.M., Ichou, F.A., Michielse, C.B., Hijdra, M., Cornelissen, B.J., Rep, M. and Takken, F.L. (2014) The Fusarium oxysporum effector Six6 contributes to virulence and suppresses I-2-mediated cell death. *Mol Plant Microbe Interact* **27**, 336-348.

Ghanbarnia, K., Fudal, I., Larkan, N.J., Links, M.G., Balesdent, M.H., Profotova, B., Fernando, W.G., Rouxel, T. and Borhan, M.H. (2014) Rapid identification of the Leptosphaeria maculans avirulence gene AvrLm2 using an intraspecific comparative genomics approach. *Molecular plant pathology*.

Golicz, A.A., Martinez, P.A., Zander, M., Patel, D.A., Van De Wouw, A.P., Visendi, P., Fitzgerald, T.L., Edwards, D. and Batley, J. (2015) Gene loss in the fungal canola pathogen Leptosphaeria maculans. *Functional & integrative genomics* **15**, 189-196.

Gonzalez, V.M., Aventin, N., Centeno, E. and Puigdomenech, P. (2013) High presence/absence gene variability in defense-related gene clusters of Cucumis melo. *BMC genomics* **14**, 782.

Gout, L., Fudal, I., Kuhn, M.L., Blaise, F., Eckert, M., Cattolico, L., Balesdent, M.H. and Rouxel, T. (2006) Lost in the middle of nowhere: the AvrLm1 avirulence gene of the Dothideomycete Leptosphaeria maculans. *Molecular microbiology* **60**, 67-80.

Grandaubert, J., Lowe, R.G., Soyer, J.L., Schoch, C.L., Van de Wouw, A.P., Fudal, I., Robbertse, B., Lapalu, N., Links, M.G., Ollivier, B., Linglin, J., Barbe, V., Mangenot, S., Cruaud, C., Borhan, H., Howlett, B.J., Balesdent, M.H. and Rouxel, T. (2014) Transposable element-assisted evolution and adaptation to host plant within the Leptosphaeria maculans-Leptosphaeria biglobosa species complex of fungal pathogens. *BMC genomics* **15**, 891.

Hammond, K.E. and Lewis, B.G. (1987) The establishment of systemic infection in leaves of oilseed rape by Leptosphaeria maculans. *Plant Pathol* **36**, 135-147.

Hane, J.K., Lowe, R.G.T., Solomon, P.S., Tan, K.C., Schoch, C.L., Spatafora, J.W., Crous, P.W., Kodira, C., Birren, B.W., Galagan, J.E., Torriani, S.F.F., McDonald, B.A. and Oliver, R.P. (2007) Dothideomycete-plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen Stagonospora nodorum. *Plant Cell* **19**, 3347-3368.

Hayden, H.L., Cozijnsen, A.J. and Howlett, B.J. (2007) Microsatellite and minisatellite analysis of Leptosphaeria maculans in Australia reveals regional genetic differentiation. *Phytopathology* **97**, 879-887.

Hayward, A., Dalton-Morgan, J., Mason, A., Zander, M., Edwards, D. and Batley, J. (2012a) SNP discovery and applications in Brassica napus. *Journal of Plant Biotechnology* **39**, 1-12.

Hayward, A., McLanders, J., Campbell, E., Edwards, D. and Batley, J. (2012b) Genomic advances will herald new insights into the Brassica: Leptosphaeria maculans pathosystem. *Plant Biology* **14**, 1-10.

Hayward, A., Vighnesh, G., Delay, C., Samian, M.R., Manoli, S., Stiller, J., McKenzie, M., Edwards, D. and Batley, J. (2012c) Second-generation sequencing for gene discovery in the Brassicaceae. *Plant biotechnology journal* **10**, 750-759.

Heath, M.C. (2000) Hypersensitive response-related death. *Plant Molecular Biology* **44**, 321-334.

HGP (2013) Human Genome Project

Hogenhout, S.A., Van der Hoorn, R.A.L., Terauchi, R. and Kamoun, S. (2009) Emerging Concepts in Effector Biology of Plant-Associated Organisms. *Molecular Plant-Microbe Interactions* **22**, 115-122.

Houterman, P.M., Cornelissen, B.J. and Rep, M. (2008) Suppression of plant resistance gene-based immunity by a fungal effector. *PLoS Pathog* **4**, e1000061.

Houterman, P.M., Ma, L., van Ooijen, G., de Vroomen, M.J., Cornelissen, B.J., Takken, F.L. and Rep, M. (2009) The effector protein Avr2 of the xylem-colonizing fungus Fusarium oxysporum activates the tomato resistance protein I-2 intracellularly. *The Plant journal : for cell and molecular biology* **58**, 970-978.

Houterman, P.M., Speijer, D., Dekker, H.L., CG, D.E.K., Cornelissen, B.J. and Rep, M. (2007) The mixed xylem sap proteome of Fusarium oxysporum-infected tomato plants. *Molecular plant pathology* **8**, 215-221.

Howlett, B.J. (2015) Lifecycle of the blackleg fungus on canola in Australia. In: *http://www.botany.unimelb.edu.au/blackleg/*.

Howlett, B.J., Idnurm, A. and Pedras, M.S.C. (2001) Leptosphaeria maculans, the causal agent of blackleg disease of Brassicas. *Fungal Genetics and Biology* **33**, 1-14.

Huang, J., Si, W.N., Deng, Q.M., Li, P. and Yang, S.H. (2014) Rapid evolution of avirulence genes in rice blast fungus Magnaporthe oryzae. *Bmc Genetics* **15**.

Huebner, C., Petermann, I., Browning, B.L., Shelling, A.N. and Ferguson, L.R. (2007) Triallelic single nucleotide polymorphisms and genotyping error in genetic epidemiology studies: MDR1 (ABCB1) G2677/T/A as an example. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **16**, 1185-1192.

Ichinose, Y., Taguchi, F. and Mukaihara, T. (2013) Pathogenicity and virulence factors of Pseudomonas syringae. *Journal of General Plant Pathology* **79**, 285-296.

Idnurm, A. and Howlett, B.J. (2003) Analysis of loss of pathogenicity mutants reveals that repeat-induced point mutations can occur in the Dothideomycete Leptosphaeria maculans. *Fungal Genetics and Biology* **39**, 31-37.

Imelfort, M., Duran, C., Batley, J. and Edwards, D. (2009) Discovering genetic polymorphisms in next-generation sequencing data. *Plant biotechnology journal* **7**, 312-317.

Indexmundi (2015) Rapeseed Oil Monthly Price. www.indexmundi.com.

Islam, M.R. and Mayo, G.M.E. (1990) A Compendium on Host Genes in Flax Conferring Resistance to Flax Rust. *Plant Breeding* **104**, 89-100.

Jenni, S., Leibundgut, M., Boehringer, D., Frick, C., Mikolasek, B. and Ban, N. (2007) Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316**, 254-261.

Jia, Y., McAdams, S.A., Bryan, G.T., Hershey, H.P. and Valent, B. (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *Embo Journal* **19**, 4004-4014.

Kadotani, N., Nakayashiki, H., Tosa, Y. and Mayama, S. (2004) One of the two Dicer-like proteins in the filamentous fungi Magnaporthe oryzae genome is responsible for hairpin RNA-triggered RNA silencing and related small interfering RNA accumulation. *The Journal of biological chemistry* **279**, 44467-44474.

Kaur, S., Cogan, N.O.I., Ye, G., Baillie, R.C., Hand, M.L., Ling, A.E., Mcgearey, A.K., Kaur, J., Hopkins, C.J., Todorovic, M., Mountford, H., Edwards, D., Batley, J., Burton, W., Salisbury, P., Gororo, N., Marcroft, S., Kearney, G., Smith, K.F., Forster, J.W. and Spangenberg, G.C. (2009) Genetic map construction and QTL mapping of resistance to blackleg (Leptosphaeria maculans) disease in Australian canola (Brassica napus L.) cultivars. *Theoretical and Applied Genetics* **120**, 71-83.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

Kmiec, B., Woloszynska, M. and Janska, H. (2006) Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Current Genetics* **50**, 149-159.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X. and Wilson, A.C. (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences* **86**, 6196-6200.

Koressaar, T. and Remm, M. (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289-1291.

Kumar, R., Qiu, J., Joshi, T., Valliyodan, B., Xu, D. and Nguyen, H.T. (2007) Single feature polymorphism discovery in rice. *PloS one* **2**, e284.

Kurtz, S., Narechania, A., Stein, J.C. and Ware, D. (2008) A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes. *BMC genomics* **9**, 517.

Lai, K., Lorenc, M.T., Lee, H.C., Berkman, P.J., Bayer, P.E., Visendi, P., Ruperao, P., Fitzgerald, T.L., Zander, M., Chan, C.-K.K., Manoli, S., Stiller, J., Batley, J. and Edwards, D. (2015) Identification and characterization of more than 4 million intervarietal SNPs across the group 7 chromosomes of bread wheat. *Plant biotechnology journal* **13**, 97-104.

Lai, K.T., Duran, C., Berkman, P.J., Lorenc, M.T., Stiller, J., Manoli, S., Hayden, M.J., Forrest, K.L., Fleury, D., Baumann, U., Zander, M., Mason, A.S., Batley, J. and Edwards, D. (2012) Single nucleotide polymorphism discovery from wheat next-generation sequence data. *Plant biotechnology journal* **10**, 743-749.

Lal, P.B., Schneider, B.L., Vu, K. and Reitzer, L. (2014) The redundant aminotransferases in lysine and arginine synthesis and the extent of aminotransferase redundancy in Escherichia coli. *Molecular microbiology* **94**, 843-856.

Lam, H.M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.L., Li, M.W., He, W., Qin, N., Wang, B., Li, J., Jian, M., Wang, J., Shao, G., Wang, J., Sun, S.S. and Zhang, G. (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature genetics* **42**, 1053-1059.

Lapin, D. and Van den Ackerveken, G. (2013) Susceptibility to plant disease: more than a failure of host immunity. *Trends Plant Sci* **18**, 546-554.

Larkan, N.J., Lydiate, D.J., Parkin, I.A., Nelson, M.N., Epp, D.J., Cowling, W.A., Rimmer, S.R. and Borhan, M.H. (2013) The Brassica napus blackleg resistance gene LepR3 encodes a receptor-like protein triggered by the Leptosphaeria maculans effector AVRLM1. *New Phytol* **197**, 595-605.

Larkan, N.J., Ma, L. and Borhan, M.H. (2015) The Brassica napus receptor-like protein RLM2 is encoded by a second allele of the LepR3/Rlm2 blackleg resistance locus. *Plant biotechnology journal*.

Lawrence, G.J., Anderson, P.A., Dodds, P.N. and Ellis, J.G. (2010) Relationships between rust resistance genes at the M locus in flax. *Molecular plant pathology* **11**, 19-32.

Lee, C.Z., Liou, G.Y. and Yuan, G.F. (2006) Comparison of the aflR gene sequences of strains in Aspergillus section Flavi. *Microbiology* **152**, 161-170.

Lee, H.C., Lai, K.T., Lorenc, M.T., Imelfort, M., Duran, C. and Edwards, D. (2012) Bioinformatics tools and databases for analysis of next-generation sequence data. *Briefings in Functional Genomics* **11**, 12-24.

Lesemann, S.S., Schimpke, S., Dunemann, F. and Deising, H.B. (2006) Mitochondrial heteroplasmy for the cytochrome b gene controls the level of strobilurin resistance in the apple powdery mildew fungus Podosphaera leucotricha (Ell. & Ev.) ES Salmon. *Journal of Plant Diseases and Protection* **113**, 259-266.

Li, J., Lupat, R., Amarasinghe, K.C., Thompson, E.R., Doyle, M.A., Ryland, G.L., Tothill, R.W., Halgamuge, S.K., Campbell, I.G. and Gorringe, K.L. (2012) CONTRA: copy number analysis for targeted resequencing. *Bioinformatics* **28**, 1307-1313.

Li, R.Q., Yu, C., Li, Y.R., Lam, T.W., Yiu, S.M., Kristiansen, K. and Wang, J. (2009a) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966-1967.

Li, W., Wang, B.H., Wu, J., Lu, G.D., Hu, Y.J., Zhang, X., Zhang, Z.G., Zhao, Q., Feng, Q.Y., Zhang, H.Y., Wang, Z.Y., Wang, G.L., Han, B., Wang, Z.H. and Zhou, B. (2009b) The Magnaporthe oryzae Avirulence Gene AvrPiz-t Encodes a Predicted Secreted Protein That Triggers the Immunity in Rice Mediated by the Blast Resistance Gene Piz-t. *Molecular Plant-Microbe Interactions* **22**, 411-420.

Lievens, B., Houterman, P.M. and Rep, M. (2009) Effector gene screening allows unambiguous identification of Fusarium oxysporum f. sp. lycopersici races and discrimination from other formae speciales. *FEMS Microbiol Lett* **300**, 201-215.

Liu, B.H. and Chu, F.S. (1998) Regulation of aflR and its product, AflR, associated with aflatoxin biosynthesis. *Applied and environmental microbiology* **64**, 3718-3723.

Liu, J., Wang, X., Mitchell, T., Hu, Y., Liu, X., Dai, L. and Wang, G.L. (2010) Recent progress and understanding of the molecular mechanisms of the rice-Magnaporthe oryzae interaction. *Molecular plant pathology* **11**, 419-427.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. and Law, M. (2012) Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology* **2012**, 251364.

Lorenc, M.T., Hayashi, S., Stiller, J., Hong, L., Manoli, S., Ruperao, P., Visendi, P., Berkman, P.J., Lai, K., Batley, J. and Edwards, D. (2012) Discovery of Single Nucleotide Polymorphisms in Complex Genomes Using SGSautoSNP. *Biology* **1**, 370-382.

Love, C., Robinson, A., Lim, G., Hopkins, C., Batley, J., Barker, G., Spangenberg, G.C. and Edwards, D. (2005) *Brassica* ASTRA: an integrated database for *Brassica* genomic research. *Nucleic Acids Research* **33**, W493-W495.

Love, C.G., Batley, J., Lim, G., Robinson, A.J., Savage, D., Singh, D., Spangenberg, G.C. and Edwards, D. (2004) New computational tools for *Brassica* genome research. *Comparative and Functional Genomics* **5**, 276-280.

Love, C.G. and Edwards, D. (2007) Accessing integrated *Brassica* genetic and genomic data using the BASC server. In: *Plant bioinformatics* (Edwards, D. ed) pp. 229-244. USA: Methods in molecular biology, Humana Press.

Lowe, R.G., Cassin, A., Grandaubert, J., Clark, B.L., Van de Wouw, A.P., Rouxel, T. and Howlett, B.J. (2014) Genomes and transcriptomes of partners in plant-fungal-interactions between canola (Brassica napus) and two Leptosphaeria species. *PloS one* **9**, e103098.

Luck, J.E., Lawrence, G.J., Dodds, P.N., Shepherd, K.W. and Ellis, J.G. (2000) Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. *Plant Cell* **12**, 1367-1377.

Ma, L.J., van der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P.M., Kang, S., Shim, W.B., Woloshuk, C., Xie, X., Xu, J.R., Antoniw, J., Baker, S.E., Bluhm, B.H., Breakspear, A., Brown, D.W., Butchko, R.A., Chapman, S., Coulson, R., Coutinho, P.M., Danchin, E.G., Diener, A., Gale, L.R., Gardiner, D.M., Goff, S., Hammond-Kosack, K.E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C.D., Koehrsen, M., Kumar, L., Lee, Y.H., Li, L., Manners, J.M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S.Y., Proctor, R.H., Regev, A., Ruiz-Roldan, M.C., Sain, D., Sakthikumar, S., Sykes, S., Schwartz, D.C., Turgeon, B.G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C.A., Kistler, H.C. and Rep, M. (2010) Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. *Nature* **464**, 367-373.

Marcais, G. and Kingsford, C. (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764-770.

Marcroft, S. and Bluett, C. (2008) Blackleg of canola. In: *Agriculture notes, Department of Primary Industries*. State of Victoria.

Marcroft, S.J., Elliott, V.L., Cozijnsen, A.J., Salisbury, P.A., Howlett, B.J. and Van de Wouw, A.P. (2012) Identifying resistance genes to Leptosphaeria maculans in Australian Brassica napus cultivars based on reactions to isolates with known avirulence genotypes. *Crop and Pasture Science* **63**, 338-350.

Marshall, D.J., Hayward, A., Eales, D., Imelfort, M., Stiller, J., Berkman, P.J., Clark, T., McKenzie, M., Lai, K.T., Duran, C., Batley, J. and Edwards, D. (2010) Targeted identification of genomic regions using TAGdb. *Plant Methods* **6**.

Mayerhofer, R., Good, A.G., Bansal, V.K., Thiagarajah, M.R. and Stringam, G.R. (1997) Molecular mapping of resistance to Leptosphaeria maculans in Australian cultivars of Brassica napus. *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada* **40**, 294-301.

McDonald, B.A. and Linde, C. (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology* **40**, 349-+.

McGee, D.C. and Petrie, G.A. (1979) Seasonal patterns of ascospore discharge by *Leptosphaeria maculans* in relation to blackleg of oilseed rape. *Phytopathology* **69**, 586-589.

McHale, L.K., Haun, W.J., Xu, W.W., Bhaskar, P.B., Anderson, J.E., Hyten, D.L., Gerhardt, D.J., Jeddeloh, J.A. and Stupar, R.M. (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant physiology* **159**, 1295-1308.

Medini, D., Donati, C., Tettelin, H., Masignani, V. and Rappuoli, R. (2005) The microbial pan-genome. *Current opinion in genetics & development* **15**, 589-594.

Mes, J.J., Weststeijn, E.A., Herlaar, F., Lambalk, J.J., Wijbrandi, J., Haring, M.A. and Cornelissen, B.J. (1999) Biological and Molecular Characterization of Fusarium oxysporum f. sp. lycopersici Divides Race 1 Isolates into Separate Virulence Groups. *Phytopathology* **89**, 156-160.

Michielse, C.B. and Rep, M. (2009) Pathogen profile update: Fusarium oxysporum. *Molecular plant pathology* **10**, 311-324.

Morinaga, T. (1934) Interspecific hybridisation in *Brassica* VI. The cytology of F_1 hybrids of *B. juncea* and *B. nigra*. *Cytologia* **6**, 62-67.

Oku, H. (1994) *Plant pathogenesis and disease control*. Boca Raton, FL:Lewis Publishers.

Park, C.H., Chen, S., Shirsekar, G., Zhou, B., Khang, C.H., Songkumarn, P., Afzal, A.J., Ning, Y., Wang, R., Bellizzi, M., Valent, B. and Wang, G.L. (2012) The Magnaporthe oryzae effector AvrPiz-t targets the RING E3 ubiquitin ligase APIP6 to suppress pathogen-associated molecular pattern-triggered immunity in rice. *Plant Cell* **24**, 4748-4762.

Parlange, F., Daverdin, G., Fudal, I., Kuhn, M.L., Balesdent, M.H., Blaise, F., Grezes-Besset, B. and Rouxel, T. (2009) Leptosphaeria maculans avirulence gene AvrLm4-7 confers a dual recognition specificity by the Rlm4 and Rlm7 resistance genes of oilseed rape, and circumvents Rlm4-mediated recognition through a single amino acid change. *Molecular microbiology* **71**, 851-863.

Patel, D.A., Zander, M., Van de Wouw, A.P., Mason, A.S., Edwards, D. and Batley, J. (2015) Population Diversity of Leptosphaeria maculans in Australia. *International Journal of Biology* **7**, 18-36.

Payne, B.A.I., Wilson, I.J., Yu-Wai-Man, P., Coxhead, J., Deehan, D., Horvath, R., Taylor, R.W., Samuels, D.C., Santibanez-Koref, M. and Chinnery, P.F. (2012) Universal heteroplasmy of human mitochondrial DNA. *Human molecular genetics* **22**, 384-390.

Pongam, P., Osborn, T.C. and Williams, P.H. (1998) Genetic Analysis and Identification of Amplified Fragment Length Polymorphism Markers Linked to the alm1 Avirulence Gene of Leptosphaeria maculans. *Phytopathology* **88**, 1068-1072.

Praefcke, G.J. and McMahon, H.T. (2004) The dynamin superfamily: universal membrane tubulation and fission molecules? *Nature reviews. Molecular cell biology* **5**, 133-147.

Purwantara, A., Barrins, J.M., Cozijnsen, A.J., Ades, P.K. and Howle, B.J. (2000) Genetic diversity of isolates of the Leptosphaeria maculans species complex from Australia,

Europe and North America using amplified fragment length polymorphism analysis. *Mycological Research* **104**, 772-781.

Purwantara, A., Salisbury, P.A., Burton, W.A. and Howlett, B.J. (1998) Reaction of Brassica juncea (Indian mustard) lines to Australian isolates of Leptosphaeria maculans under glasshouse and field conditions. *European Journal of Plant Pathology* **104**, 895-902.

Rafalski, A. (2002) Applications of single nucleotide polymorphisms in crop genetics. *Current opinion in plant biology* **5**, 94-100.

Rafiqi, M., Gan, P.H.P., Ravensdale, M., Lawrence, G.J., Ellis, J.G., Jones, D.A., Hardham, A.R. and Dodds, P.N. (2010) Internalization of Flax Rust Avirulence Proteins into Flax and Tobacco Cells Can Occur in the Absence of the Pathogen. *Plant Cell* **22**, 2017-2032.

Raman, R., Taylor, B., Marcroft, S., Stiller, J., Eckermann, P., Coombes, N., Rehman, A., Lindbeck, K., Luckett, D., Wratten, N., Batley, J., Edwards, D., Wang, X.W. and Raman, H. (2012) Molecular mapping of qualitative and quantitative loci for resistance to Leptosphaeria maculans causing blackleg disease in canola (Brassica napus L.). *Theoretical and Applied Genetics* **125**, 405-418.

Ravensdale, M., Nemri, A., Thrall, P.H., Ellis, J.G. and Dodds, P.N. (2011) Co-evolutionary interactions between host resistance and pathogen effector genes in flax rust disease. *Molecular plant pathology* **12**, 93-102.

Raymer, P.L. (2002) Canola: an emerging oilseed crop. In: *Trends in new crops and new uses* (Janick, J. and Whipkey, A. eds). Alexandria: ASHS Press.

Rep, M. (2005) Small proteins of plant-pathogenic fungi secreted during host colonization. *FEMS Microbiol Lett* **253**, 19-27.

Rep, M., Meijer, M., Houterman, P.M., van der Does, H.C. and Cornelissen, B.J. (2005) Fusarium oxysporum evades I-3-mediated resistance without altering the matching avirulence gene. *Mol Plant Microbe Interact* **18**, 15-23.

Rimmer, S.R. (2006) Resistance genes to Leptosphaeria maculans in Brassica napus. *Canadian Journal of Plant Pathology* **28**, S288-S297.

Rosenfeld, J.A., Malhotra, A.K. and Lencz, T. (2010) Novel multi-nucleotide polymorphisms in the human genome characterized by whole genome and exome sequencing. *Nucleic acids research* **38**, 6102-6111.

Rouxel, T. and Balesdent, M.H. (2005) The stem canker (blackleg) fungus, Leptosphaeria maculans, enters the genomic era. *Molecular plant pathology* **6**, 225-241.

Rouxel, T., Chupeau, Y., Fritz, R., Kollmann, A. and Bousquet, J.-F. (1988) Biological effects of sirodesmin PL, a phytotoxin produced by Leptosphaeria maculans. *Plant Science* **57**, 45-53.

Rouxel, T., Grandaubert, J., Hane, J.K., Hoede, C., van de Wouw, A.P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., Cozijnsen, A.J., Ciuffetti, L.M., Degrave, A., Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S.B., Gout, L., Glaser, N., Linglin, J., Kema, G.H.J., Lapalu, N., Lawrence, C.B., May, K., Meyer, M., Ollivier, B., Poulain, J., Schoch, C.L., Simon, A., Spatafora, J.W., Stachowiak, A., Turgeon, B.G., Tyler, B.M., Vincent, D., Weissenbach, J., Amselem, J., Quesneville,

H., Oliver, R.P., Wincker, P., Balesdent, M.-H. and Howlett, B.J. (2011) Effector diversification within compartments of the Leptosphaeria maculans genome affected by Repeat-Induced Point mutations. *Nature Communications* **2**.

Rouxel, T., Willner, E., Coudard, L. and Balesdent, M.-H. (2003) Screening and identification of resistance to Leptosphaeria maculans (stem canker) in Brassica napus accessions. *Euphytica* **133**, 219-231.

Salisbury, P.A., Ballinger, D.J., Wratten, N., Plummer, K.M. and Howlett, B.J. (1995) Blackleg Disease on Oilseed Brassica in Australia - a Review. *Australian Journal of Experimental Agriculture* **35**, 665-672.

Sambrook, J., Russell, D.W. and Sambrook, J. (2006) *The condensed protocols from Molecular cloning : a laboratory manual*. Cold Spring Harbor, N.Y.:Cold Spring Harbor Laboratory Press.

Schacherer, J., Shapiro, J.A., Ruderfer, D.M. and Kruglyak, L. (2009) Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae. *Nature* **458**, 342-345.

Schweizer, E. and Hofmann, J. (2004) Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiology and molecular biology reviews : MMBR* **68**, 501-517, table of contents.

Seberry, D.E., Mailer, R.J. and Parker, P.A. (2008) Quality of Australian Canola. *NSW Department of Primary Industries & Australian Oilseeds Federation* **15**.

She, X.W., Rohl, C.A., Castle, J.C., Kulkarni, A.V., Johnson, J.M. and Chen, R.H. (2009) Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. *BMC genomics* **10**.

Shen, J., Araki, H., Chen, L., Chen, J.Q. and Tian, D. (2006) Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in Arabidopsis thaliana. *Genetics* **172**, 1243-1250.

Simons, G., Groenendijk, J., Wijbrandi, J., Reijans, M., Groenen, J., Diergaarde, P., Van der Lee, T., Bleeker, M., Onstenk, J., de Both, M., Haring, M., Mes, J., Cornelissen, B., Zabeau, M. and Vos, P. (1998) Dissection of the fusarium I2 gene cluster in tomato reveals six homologs and one active gene copy. *Plant Cell* **10**, 1055-1068.

Sparkes, I.A., Runions, J., Kearns, A. and Hawes, C. (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nature protocols* **1**, 2019-2025.

Statista (2015) Leading producing countries of rapeseed in 2013. http://www.statista.com/.

Stergiopoulos, I. and de Wit, P.J. (2009) Fungal effector proteins. *Annu Rev Phytopathol* **47**, 233-263.

Stotz, H.U., Thomson, J.G. and Wang, Y. (2009) Plant defensins: defense, development and application. *Plant Signal Behav* **4**, 1010-1012.

Syme, R.A., Hane, J.K., Friesen, T.L. and Oliver, R.P. (2013) Resequencing and comparative genomics of Stagonospora nodorum: sectional gene absence and effector discovery. *G3* **3**, 959-969.

Taiyan, Z., Lianli, L., Guang, Y. and Al-Shehbaz, I.A. (2001) Brassicaceae. In: *Flora of China* (Wu, Z. and Raven, P.H. eds). Beijing: Science Press.

Takken, F. and Rep, M. (2010) The arms race between tomato and Fusarium oxysporum. *Molecular plant pathology* **11**, 309-314.

Tan, S., Zhong, Y., Hou, H., Yang, S. and Tian, D. (2012) Variation of presence/absence genes among Arabidopsis populations. *BMC evolutionary biology* **12**, 86.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., Deboy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K.J., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappuoli, R. and Fraser, C.M. (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13950-13955.

Thatcher, L.F., Gardiner, D.M., Kazan, K. and Manners, J.M. (2012) A highly conserved effector in Fusarium oxysporum is required for full virulence on Arabidopsis. *Mol Plant Microbe Interact* **25**, 180-190.

Tian, X.J., Zheng, J., Hu, S.N. and Yu, J. (2006) The rice mitochondrial genomes and their variations. *Plant physiology* **140**, 401-410.

Tollenaere, R., Hayward, A., Dalton-Morgan, J., Campbell, E., Lee, J.R.M., Lorenc, M.T., Manoli, S., Stiller, J., Raman, R., Raman, H., Edwards, D. and Batley, J. (2012) Identification and characterization of candidate Rlm4 blackleg resistance genes in Brassica napus using next-generation sequencing. *Plant biotechnology journal* **10**, 709-715.

U, N. (1935) Genome-analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* **7**, 389-452.

UNFAO (2015) World Crop Production Aggregates. Food and Agriculture Organization of the United Nations.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. (2012) Primer3--new capabilities and interfaces. *Nucleic acids research* **40**, e115.

USDA (2015) United States Department of Agriculture Economic Research Service.

Van de Wouw, A.P., Cozijnsen, A.J., Hane, J.K., Brunner, P.C., McDonald, B.A., Oliver, R.P. and Howlett, B.J. (2010) Evolution of Linked Avirulence Effectors in Leptosphaeria maculans Is Affected by Genomic Environment and Exposure to Resistance Genes in Host Plants. *Plos Pathogens* **6**.

Van de Wouw, A.P. and Howlett, B.J. (2012) Estimating frequencies of virulent isolates in field populations of a plant pathogenic fungus, Leptosphaeria maculans, using high-throughput pyrosequencing. *Journal of Applied Microbiology* **113**, 1145-1153.

Van de Wouw, A.P., Lowe, R.G., Elliott, C.E., Dubois, D.J. and Howlett, B.J. (2014) An avirulence gene, AvrLmJ1, from the blackleg fungus, Leptosphaeria maculans, confers avirulence to Brassica juncea cultivars. *Molecular plant pathology* **15**, 523-530.

van der Does, H.C., Duyvesteijn, R.G., Goltstein, P.M., van Schie, C.C., Manders, E.M., Cornelissen, B.J. and Rep, M. (2008a) Expression of effector gene SIX1 of Fusarium oxysporum requires living plant cells. *Fungal Genet Biol* **45**, 1257-1264.

van der Does, H.C., Lievens, B., Claes, L., Houterman, P.M., Cornelissen, B.J. and Rep, M. (2008b) The presence of a virulence locus discriminates Fusarium oxysporum isolates causing tomato wilt from other isolates. *Environ Microbiol* **10**, 1475-1485.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I., Cheng, F., Huang, S., Li, X., Hua, W., Wang, J., Wang, X., Freeling, M., Pires, J.C., Paterson, A.H., Chalhoub, B., Wang, B., Hayward, A., Sharpe, A.G., Park, B.S., Weisshaar, B., Liu, B., Li, B., Liu, B., Tong, C., Song, C., Duran, C., Peng, C., Geng, C., Koh, C., Lin, C., Edwards, D., Mu, D., Shen, D., Soumpourou, E., Li, F., Fraser, F., Conant, G., Lassalle, G., King, G.J., Bonnema, G., Tang, H., Wang, H., Belcram, H., Zhou, H., Hirakawa, H., Abe, H., Guo, H., Wang, H., Jin, H., Parkin, I.A., Batley, J., Kim, J.S., Just, J., Li, J., Xu, J., Deng, J., Kim, J.A., Li, J., Yu, J., Meng, J., Wang, J., Min, J., Poulain, J., Wang, J., Hatakeyama, K., Wu, K., Wang, L., Fang, L., Trick, M., Links, M.G., Zhao, M., Jin, M., Ramchiary, N., Drou, N., Berkman, P.J., Cai, Q., Huang, Q., Li, R., Tabata, S., Cheng, S., Zhang, S., Zhang, S., Huang, S., Sato, S., Sun, S., Kwon, S.J., Choi, S.R., Lee, T.H., Fan, W., Zhao, X., Tan, X., Xu, X., Wang, Y., Qiu, Y., Yin, Y., Li, Y., Du, Y., Liao, Y., Lim, Y., Narusaka, Y., Wang, Y., Wang, Z., Li, Z., Wang, Z., Xiong, Z., Zhang, Z. and Brassica rapa Genome Sequencing Project, C. (2011) The genome of the mesopolyploid crop species Brassica rapa. *Nature genetics* **43**, 1035-1039.

Wang, Y., Lu, J., Chen, S., Shu, L., Palmer, R.G., Xing, G., Li, Y., Yang, S., Yu, D., Zhao, T. and Gai, J. (2014) Exploration of presence/absence variation and corresponding polymorphic markers in soybean genome. *Journal of integrative plant biology* **56**, 1009-1019.

White, F.F., Yang, B. and Johnson, L.B. (2000) Prospects for understanding avirulence gene function. *Current opinion in plant biology* **3**, 291-298.

Xu, J., Guo, H. and Yang, Z.L. (2007) Single nucleotide polymorphisms in the ectomycorrhizal mushroom Tricholoma matsutake. *Microbiology* **153**, 2002-2012.

Yu, F., Gugel, R.K., Kutcher, H.R., Peng, G. and Rimmer, S.R. (2013) Identification and mapping of a novel blackleg resistance locus LepR4 in the progenies from Brassica napus x B. rapa subsp. sylvestris. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **126**, 307-315.

Yu, F., Lydiate, D.J. and Rimmer, S.R. (2005) Identification of two novel genes for blackleg resistance in Brassica napus. *Theoretical and Applied Genetics* **110**, 969-979.

Zander, M., Patel, D.A., Van de Wouw, A., Lai, K.T., Lorenc, M.T., Campbell, E., Hayward, A., Edwards, D., Raman, H. and Batley, J. (2013) Identifying genetic diversity of avirulence genes in Leptosphaeria maculans using whole genome sequencing. *Functional & integrative genomics* **13**, 294-308.

Zhai, C., Lin, F., Dong, Z., He, X., Yuan, B., Zeng, X., Wang, L. and Pan, Q. (2011) The isolation and characterization of Pik, a rice blast resistance gene which emerged after rice domestication. *New Phytol* **189**, 321-334.

Zhou, R. and Linz, J.E. (1999) Enzymatic function of the nor-1 protein in aflatoxin biosynthesis in Aspergillus parasiticus. *Applied and environmental microbiology* **65**, 5639-5641.

Zhu, B. and Rimmer, S.R. (2003) Inheritance of resistance to Leptosphaeria maculans in two accessions of Brassica napus. *Canadian Journal of Plant Pathology* **25**, 98-103.

Appendices

Appendix 1: Zander et al. 2015 submitted publication.

Plant Biotechnology Journal Proof

Understanding Effector-Triggered Immunity of Plants to Pathogens: A Genomics Approach

Journal:	*Plant Biotechnology Journal*
Manuscript ID:	PBI-00388-2015
Manuscript Type:	SI Genomics
Date Submitted by the Author:	26-Jul-2015
Complete List of Authors:	Zander, Manuel; University of Queensland, Alamery, Salman; University of Queensland, Lorberg, Sarah; University of Queensland, Patel, Dhwani; University of Queensland, Tollenaere, Reece; University of Queensland, Edwards, David; University of Western Australia, School of Plant Biology Batley, Jacqueline; University of Western Australia, School of Plant Biology
Keywords:	Effector Triggered Immunity, Plant pathogen interactions, avirulence genes, resistance genes

Understanding Effector-Triggered Immunity of Plants to Pathogens: A Genomics Approach

Manuel Zander[1], Salman Alamery[1], Sarah Kay Lorberg[1,2], Dhwani Apurva Patel[1], Reece Tollenaere[1], David Edwards[1,2], Jacqueline Batley[1,2,@]

[1]School of Agriculture and Food Sciences, University of Queensland, St Lucia, Queensland, 4072, Australia

[2]School of Plant Biology, University of Western Australia, Crawley, Western Australia, 6009, Australia

[@]Corresponding Author: Prof Jacqueline Batley. jacqueline.batley@uwa.edu.au

Word count: 5484

Abstract:

Plant defence responses to pathogens are an important area of study to help contribute to food security worldwide. Of particular interest is the specific interaction between plant resistance genes and their respective pathogenic effector genes. This mechanism of pathogen detection and plant immune response are widely considered the best strategy for stable and ongoing resistance. In this review we briefly discuss a number of pathosystems and the genetic characteristics of both resistance and pathogenic effector genes and their interaction. We highlight the key aspects of resistance mechanisms in these pathosystems and explore the similarities and differences between these. The uptake of next-generation sequencing technology has allowed for improved identification and validation of genes involved in effector-triggered immunity such as the resistance and avirulence genes in host-pathogen interactions.

Introduction to Plant Defence Responses

Being sessile organisms with no means of physically escaping a disease threat, plants defend themselves against pathogens in their environment through a complex array of structural, chemical and protein-based pathogen defence mechanisms that could be seen as an "immune system", recognising and mounting an appropriate immune response.

The primary level of plant defence is represented by the passive defences, which consist of physical barriers (wax, cuticle, cell wall, stomata etc.) and chemical barriers (nutrient deprivation, pH, plant defensins, phytoanticipins etc) (Dangl and Jones, 2001; Guest and Brown, 1997). Plants also have a secondary, active level of defence, which is characterised by being pathogen triggered and therefore more specific in its response (Guest and Brown, 1997). This results in pathogen-associated molecular pattern (PAMP) triggered immunity (PTI). Examples of molecular patterns recognised by plants are bacterial flagellin and fungal chitin molecules (Dodds and Thrall, 2009; Zipfel, 2009). These primary active defence reactions can often be sufficient to halt pathogen growth and spread within the plant and include rapid defences such as membrane function, oxidative burst, cell wall reinforcement, hypersensitive cell death as well as the delayed defences such as pathogenesis-related proteins, systemic acquired resistance and pathogen containment (Zipfel, 2009).

The active defence provided by resistance genes is of particular interest and has been extensively characterised. Plant disease resistance genes (*R* genes) are important components of the genetic resistance defence mechanisms in plants (Dangl and Jones, 2001). Many *R* genes conferring resistance to a wide spectrum of plant pathogens, including bacteria, fungi, oomycetes, viruses and nematodes, have been cloned from different plant species (Wan et al., 2012). The plant pathogens in turn secrete proteins and other molecules in order to overcome plant defence responses and colonise the host plant (Hogenhout et al., 2009). In addition, *R* genes are involved in direct or indirect interaction with pathogen effector genes, also known as avirulence (*Avr*) genes, in order to trigger a defence response (Lorang et al., 2012). It is important to note that quantitative or partial resistance, mediated by multiple genes, also plays a role in minimising disease spread and complementing plant *R* genes in providing stable resistance in the field (Brun et al., 2010). Thus, understanding the molecular structure and function of *R* genes has been crucial for plant resistance research.

With the increasing implementation of next generation sequencing technologies in the study of plant-pathogen interactions, novel opportunities have arisen that aid the discovery and understanding of plant resistance genes and their corresponding pathogen avirulence genes.

Genome sequences are now available for many pathogens and their corresponding host species, allowing detailed studies of the co-evolution of these important plants and pathogens. For example, these reference genome sequences can be used as a basis for whole genome resistance gene identification, as a reference for re-sequencing projects to identify polymorphisms associated with resistance or virulence. As genome re-sequencing projects expand rapidly, pan genomes can be established and utilised for the identification of resistance and avirulence genes not present in the reference genome (Golicz et al., 2015)

Effector-Triggered Immunity

Effector-triggered immunity (ETI), also known as the gene-for-gene hypothesis, is a pathogen race-specific immune response used to explain genetic interactions between host plants and their pathogens. This concept proposes that for each resistance gene in the host there is a specific effector gene corresponding to avirulence in the pathogen (Dangl and Jones, 2001). This was first proposed by Harold Flor in 1942 in the flax-flax rust pathosystem, which has become an important model system for ETI (Flor, 1942). The gene for gene interaction suggests direct or indirect recognition of pathogen avirulence (*Avr)*-encoded effectors by the protein product encoded by the corresponding resistance gene (*R* gene) in the host plant. In response to these effectors, resistance proteins in plants activate defence mechanisms such as the hypersensitive response (Heath, 2000). Often *R* genes encode nucleotide binding site - Leucine rich repeat (NBS-LRR) proteins leading to the induction of a signalling cascade and subsequent downstream defence responses (Bent, 1996). However, recent studies have concluded that the interaction between *R* and *Avr* genes is very complex and may involve multiple effectors to trigger resistance in the plant (Gassmann and Bhattacharjee, 2012).

Resistance Gene Structure and Function

A large number of *R* genes have been cloned from a wide range of plant species, including Arabidopsis (Yi and Richards, 2007), Flax (Lawrence et al., 2010), Rice (Jia et al., 2000), Tomato (Takken and Rep, 2010), Wheat (Periyannan et al., 2013; Saintenac et al., 2013) and Brassicas (Larkan et al., 2015). This identification has accelerated along with the advances in genomic technologies. These *R* genes are diverse in terms of their structure, function and evolution, however they have been grouped into several different classes, based on structural similarities of their predicted protein products (Liu et al., 2007; Staskawicz et al., 1995). The majority of proteins encoded by *R* genes in

plant genomes belong to the nucleotide binding site and leucine-rich repeat (NBS-LRR) domain-containing class of proteins (Baumgarten et al., 2003; Persson et al., 2009). NBS-LRR proteins have a variable N-terminus, which commonly contains a domain with similarity to the *Drosophila* Toll and mammalian Interleukin-1 receptor (TIR) or a coiled coil (CC) sequence (Persson et al., 2009). These form the two major subfamilies of plant NBS-LRR proteins; TIR-NBS-LRR proteins (TNLs) and CC-NBS-LRR proteins (CNLs). The TNL group of genes has been observed only in dicot plant species (Goff et al., 2002; Meyers et al., 1999) whereas CNLs are also found in monocot plant species. Both families are involved in defence mechanisms against pathogens in different signalling pathways. They are distinct in the sequence of the conserved NBS motif so that they are found in separate clusters in phylogenetic analyses (McHale et al., 2006).

NBS-LRR proteins are located intracellularly and may interact directly with pathogen *Avr* proteins, known as effectors, or with host factor proteins that enable the indirect identification of pathogen effectors via a signal transduction pathway (Ellis et al., 2000). The only known roles of plant NBS LRR proteins are to provide microbial recognition and activate defence responses (Eitas and Dangl, 2010). The NBS region is thought to be important for ATP binding activity. It is highly conserved, containing several motifs that are strictly ordered (Meyers et al., 1999). Sequence analyses have revealed that NBS domains share a high degree of homology and have a number of conserved motifs (Wan et al., 2012). To date, eight conserved motifs have been identified in the NBS domains of these two NBS LRR classes including P-loop, kinase-2, kinase-3a, GLPL, RNBS-A-TIR, RNBS-D-TIR, RNBS-A-non-TIR and RNBS-D-non-TIR; some of which are specific to the non-TIR and the TIR NBS-LRR families (Wan et al., 2013).

Pathogen Effector Genes

In 1994 Oku defined (1) entering a plant, (2) overcoming host resistance and (3) initiating disease as the three abilities for microorganisms to parasitically adapt into plant pathogens (Ichinose et al., 2013). These observations were primarily based on infections caused by fungal pathogens Understanding the role of secreted proteins or effectors that target plant defence responses is essential to gaining knowledge of how pathogens infect and colonise the host.

Effectors, or avirulence (*Avr*) genes, are phytopathogenic genes encoding a protein that are specifically recognised by host plants containing the 'matching' resistance (*R*) gene (Rouxel and Balesdent, 2005). The products of these genes are known as effector proteins, and these carry out functions such as plant defence suppression and host cell-wall modifications upon infection (Rouxel

et al., 2011). Identifying novel effectors is a challenge due to the large portion that have novel sequences with no similarity to other effectors, other proteins in known databases and no relatedness to other species in the same genus or same species in different host ranges (Ellis et al., 2009). Many effectors possess a dual function of avirulence and virulence (White et al., 2000). Several effectors are recognised by host resistance (R) proteins, triggering the hypersensitive reaction (HR) in host cells during expression or secretion in host cells. This is their avirulence function caused due to the pathogen being avirulent on hosts that express the corresponding R protein (Ellis et al., 2009).

Plant pathogens have evolved to employ a variety of mechanisms that deliver effectors directly into host cells. Bacteria have been shown to utilise special secretory systems, such as TS33, whilst fungi and oomycetes use structures known as haustoria (Hogenhout et al., 2009). Parasitic nematodes use stylets or special feeding tubes to deliver effectors into the plant vascular cell. There are selective fungal effector proteins, such as ToxA in *Pyrenophora tritici-repentis,* that translocate independently within the host cell by using a plant-surface receptor bound to specialised motifs (Hogenhout et al., 2009). Effectors are also known to affect plant apoplastic defences. Fungi such as *Cladosporium fulvum*, which is an extracellular parasite of tomato, grow only in the apoplast and do not form haustoria. Apoplastic effectors are commonly known to protect the pathogen against plant hydrolytic enzymes such as glucanases, chitnases and proteases. *Phytophthora infestans* secretes protease inhibitors like EPIC2B and EPI1 and EPI10, which are cysteine protease and serine protease inhibitors respectively, protecting the pathogen against plant defence activity. *P. infestans* is also known to secrete host translocated or cytoplasmic effectors (Hogenhout et al., 2009).

How Pathogen Effectors Affect the Host

Plant pathogens elicit a variety of responses in the host that can be beneficial for the pathogen, have neutral effects or have negative consequences (Hogenhout et al., 2009). Some effectors are known to supress plant immunity as their primary function. Examples of these are T3SS effectors of bacteria that act by suppressing basal plant defences like PAMP or by supressing hypersensitive cell death. These effectors target three main plant processes- Protein turnover, phosphorylation pathway and RNA homeostasis (Block et al., 2008). Other oomycetes and biotropic fungi also suppress host cell death. RLXR effectors of oomycetes such as Avr3a of *P. infestans* and Avr1b of *P. sojae* have been found to supress the hypersensitive cell death caused by the INFI elicitin protein (Bos et al., 2006) and mouse BAX protein (Dou et al., 2008) respectively.

There exists a concept that some fungal effectors evolve to evade host resistance conferred by *R* genes. A study conducted on *Fusarium oxysporum* f. sp. *Lycopersici* showed that its effector Avr1 supresses host resistance response conferred by the *R* genes *I-2* and *I-3*. Since there is no evidence to suggest virulent effects on plants not containing these *I* genes, it is suggested that this effector is only involved in pathogen recognition by host R proteins (Houterman et al., 2008).

Some effectors are known to affect the morphology of their host plants and change their behaviour. The phytotoxin coronatine has been shown to inhibit stomatal closure when PAMPs from *Pseudomonas syringae* infection are detected by *Arabidopsis*, to allow the entry of bacteria into the plant (Melotto et al., 2006). Another example is the plant fungus *Gibberella fujikuroi* or *Fusarium moniliforme*, which infects rice seedlings. The pathogen causes the growth of a seedling that is taller than the non-infected seedlings via production of gibberellin the growth hormone (Tudzynski, 1999). The added height of the infected seedling allows the pathogen to spread its airborne spores. This effector action is also a good example of how this pathogen effector mimics a plant hormone, displaying molecule mimicry. Other examples of molecule mimicry have also been seen in nematodes such as the soybean cyst nematode *Heterodera glycines*. This pathogen possesses the CLAVATA3-like 4G12 gene and is thought to negatively regulate the *Arabidopsis WUSCHEL* gene, possibly interfering with plant growth and development (Hogenhout et al., 2009).

Several studies have proven that genes directly involved in effector function will be preferentially selected and evolve at a higher rate as compared to the rest of the genome. In the same token, several host targets have also been found to evolve to evade effector manipulation. Understanding the interaction between effectors, effector targets and R proteins and the subsequent mechanistic actions will be valuable information to correctly employ *R* genes in disease management (Persson et al., 2009). Sequencing of pathogen populations will enable an understanding of this selection pressure, through the identification of allelic variants of known effector proteins, track changes in the effector protein composition within the population in relation to crops and cultivars grown and identification of novel allelic variants leading to virulence or loss of virulence.

Example of Plant-Pathogen systems:

Flax and Flax Rust

The interaction between flax (*Linum usitatissimum, L. marginale* and other *Linum* species*)* and the pathogen flax rust (*Melampsora lini*) provided the earliest evidence of the genetic principle for plant disease resistance, being defined in 1956 by Flor et al. and remains one of the most important models for plant-pathogen interactions. It has been used to understand the genetic and molecular characteristics of host-pathogen interactions and how they develop in natural disease systems (Ravensdale et al., 2011). As an obligate biotroph, *M. lini* depends on living plant tissue for survival and extracts nutrients through specialised feeding structures known as haustoria.

Flax resistance (*R*) genes have been mapped to five distinct loci across the genome, referred to as the *K, L, M, N*, and *P* groups (Islam and Mayo, 1990; Lawrence et al., 2010). Within these loci, 30 genes conferring resistance to flax rust have been mapped and subsequently shown to be closely linked genes or allelic variants of the same gene. Of the nineteen *R* genes cloned, 11 are from group *L*, three from group *N*, two from group *P* and three from group *M*. The loci are complex, with each consisting of varying avirulence gene specificities. For example, a single gene at the *L* locus possesses 13 allelic variants, reacting to rust strains with different *Avr* genes. The *N* locus contains four, the *M* locus up to 15, and the *P* locus six to eight *R* gene paralogues, arranged in tandem at these locations, and more complex than the other loci (Ravensdale et al., 2011).

So far all the *R* genes in flax have been shown to encode TIR-NBS-LRR class proteins, with much of the variation between these occurring in the LRR domain (Ravensdale et al., 2011), a horseshoe-shaped domain which seems to be critical in the recognition specificity between *R* and *Avr* proteins (Ellis et al., 1999). Polymorphisms of just 6 amino acids in the exposed residues within the LRR domains of *P* and *P2*, give these two *R* genes different avirulence gene specificities (Dodds et al., 2001). The role of the LRR domain in pathogen effector recognition has been observed, however evidence also suggests regions outside the LRR, such as the TIR domain, may further influence effector specificity (Luck et al., 2000), which has also been shown in the *N* resistance gene of the tobacco-Tobacco mosaic virus (TMV) pathosystem (Burch-Smith et al., 2007).

The effector genes of flax rust have been classified into four *Avr* families which contribute to approximately 30 different *Avr* specificities. All the identified *Avr* genes and their variants encode small secreted proteins, and the majority were identified by screening expressed rust genes during infection and colonisation of flax (Catanzariti et al., 2006; Dodds et al., 2004). These small secreted proteins appear to be translocated into plant cells and indicate that *Avr* protein recognition occurs

inside the plant, as flax R proteins are cytoplasmic (Catanzariti et al., 2006; Dodds et al., 2004). Immunolocalisation studies have identified the presence of Avr proteins inside host cells, as well as evidence to suggest that the pathogen exploits plant-derived transport mechanisms to enter the cells (Rafiqi et al., 2010).

Flax rust effectors exhibit a high level of sequence polymorphism that is associated with a difference in recognition specificity (Dodds and Thrall, 2009). The four *Avr* families show no significant sequence similarity with each other as well as to other known proteins, indicating that they are unique and specific to this pathosystem (Catanzariti et al., 2006).

Rice and Rice Blast

Rice blast (*Magnaporthe grisea*) pathosystem is one of the most significant diseases of cereal crops in the world and has proven to be a useful model for ETI in both monocots and plants in general. *M. grisea* is an ascomycete fungus that infects both aerial and root tissue in cereals, primarily rice species (*Oryza* species). The genomes of the *O. sativa* ssp. *japonica* and *indica* contain approximately 80 rice blast *R* genes (Zhai et al., 2011). Of these, 13 have been isolated and characterised to date (Liu et al., 2010) and their interaction with rice blast *Avr* genes is still being elucidated, with direct R-AVR protein interaction so far studied in Pita:AVR-Pita (Jia et al., 2000) and Pizt:AVR-Pizt (Li et al., 2009). Improvements have been achieved through the availability of whole genome sequence for both host and pathogen species and will continue to improve the knowledge of this patho-system. Host resistance in rice contains both R gene mediated resistance, conferring complete resistance, and quantitative resistance which provide partial resistance in the field.

The R genes of rice have been shown to be clustered throughout the genome, especially on chromosomes 6, 11 and 12 (Liu et al., 2010). Of the 13 cloned rice blast R genes, all except one (*Pi-d2*) encode NBS-LRR proteins, *Pi-d2* encoding a receptor kinase (Chen et al., 2006). As in other plant species, the LRR domain of rice R genes have been shown to form a direct interaction with the pathogen effector proteins, which has been identified between the rice *Pi-ta R* gene and the rice blast *Avr-Pita* (Jia et al., 2000).

The genome of *Magnaporthe* species has been shown to encode many secreted proteins which could be involved in the suppression of host defence responses. Indeed the The *AvrPiz-t* effector of *M. oryzae* induces suppression of PAMP-triggered immunity in rice, targeting the ubiquitin ligase APIP6 which in turn suppresses flg22 and chitin induction of reactive oxygen species (ROS) (Park et al., 2012). The *Avr* genes, which induce ETI in rice, vary greatly in structure and function. For

example, the *ACE1* gene, avirulent towards the rice *Pi33 R* gene, encodes a hybrid polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS) and has been postulated to drive the biosynthesis of a secondary metabolite that in turn is recognised by the Pi22 protein (Bohnert et al., 2004).

Tomato and Fusarium Wilt

Fusarium wilt disease of Tomato (*Solanum lycopersicum*) is caused by the fungus *Fusarium oxysporum f. sp. Lycopersici (Fol).* Resistance genes in Tomato against *Fol* are known as *Immunity (I)* genes. So far the genes *I, I-1, I-2* and *I-3* have been identified. These genes were originally sourced from wild species and were successfully introgressed to create resistant tomato cultivars (Mes et al., 1999; Takken and Rep, 2010). Resistance genes *I* and *I-2* are both effective against the fungal avirulence gene *Avr1/Six4. I-3* is thought to interact with the *Fol Avr3/*SIX1 protein (Rep et al., 2005). To date, only one Tomato-Fusarium R gene, *I-2*, has been successfully cloned (Simons et al., 1998) and tomato lines transformed with *Fol Avr2* exhibit endogenous *I*-2 expression in the leaves (Houterman et al., 2009; Takken and Rep, 2010). Unfortunately, proving a direct interaction between *Avr2* and *I-2* proteins through yeast two-hybrid methods has been unsuccessful, indicating that indirect signalling mechanism may be responsible for the interaction (Takken and Rep, 2010). Similar to other pathosystems, the *I-2* gene encodes an NBS-LRR protein. Mutation of the nucleotide binding domain suggests that the active form of the protein requires the specific binding of a nucleotide for pathogen recognition (Takken and Rep, 2010). Unlike other R gene mediated defence responses where hypersensitive response (HR) is involved, *I* gene resistance utilises other mechanisms, including callose deposition, phenolic production, and deposition of other protective barrier compounds (Michielse and Rep, 2009; Takken and Rep, 2010).

During pathogenesis Fol secretes a multitude of proteins into the host, including effector proteins. Some of the *Fol* avirulence genes belong to a family which encode 'Secreted in Xylem' (SIX) proteins. There are 11 of these proposed, however, not all are thought to be true effectors based on the lack of interaction with host R genes (Gawehns et al., 2014; Takken and Rep, 2010). Of these, *SIX1/Avr3*, the *Avr* gene recognised by *R* gene *I-3*, is present in all *Fol* tested isolates and appears to be unique to this forma specialis, suggesting a dual, tomato-fusarium specific role (Rep, 2005; Rep et al., 2005). Three confirmed effectors interacting with host *R* genes are *Avr1/Six4, Avr2/Six3* and *Avr3/Six1* (Houterman et al., 2009; Houterman et al., 2007; Lievens et al., 2009; Rep et al., 2005). Recently, homologs of SIX1, SIX4, SIX8 and SIX9 were identified in an Arabidopsis infecting strain of Fol

(Thatcher et al., 2012). Most SIX proteins appear to be unique to *Fusarium oxysporum*. However, recently, homologs of SIX6 were identified in several different microbes (Gawehns et al., 2014). A SIX6 knockout displayed inhibited virulence, with complementation restoring virulence, in two independent transformants suggesting an effector role for this protein (Gawehns et al., 2014). Furthermore, expression analysis show that SIX6 is not expressed on synthetic media but requires a live-host cell for expression as has been previously shown for *SIX1/Avr3* (Gawehns et al., 2014; van der Does et al., 2008a).

Although members of *Fusarium oxysporum f. sp. Lycopersici* have a polyphyletic origin all pathogenic strains require an 8 kb region for pathogenicity and specificity of this forma specialis (van der Does et al., 2008b). Within *Fol* mobile pathogenicity chromosomes have been described; when chromosome 14 was transferred into a non-pathogenic *Fol* this resulted in pathogenicity (Ma et al., 2010).

Due to the high reproduction rates and abundance of *Fusarium* in the soil there is a constant race between host and pathogen to overcome ETI. It was proposed that the breakdown of single gene resistance is not dependant on loss or mutation of avirulence genes rather dependant on gene suppression (Rep et al., 2005). However, further studies have revealed this is through *Avr1* triggered suppression of *I-2* and *I-3* (Houterman et al., 2008; Lapin and Van den Ackerveken, 2013). It is proposed that Avr1 was acquired by Fol and suppression mechanism exploited in order to maintain other Avr content (Ma et al., 2010; Thatcher et al., 2012).

Futhermore, plant resistance utilising specific aspects of PAMP-triggered immunity (PTI) has been shown. For example two reports, Abdallah et al. (2010) and Khan et al (2011) successfully created *Fusarium oxysporum f. sp. Lycopersici* resistant GM lines through over expressing defensin genes, small cysteine rich proteins (Stotz et al., 2009). Other non-ETI related research targets in other microbial species, such as non-pathogenic *Fusarium* strains and *Bacillus spp.* have been used as biological controls against *Fol* (Ajilogba and Babalola, 2013). Despite these findings, *R* genes are still considered the most promising mechanism of resistance in plants.

Brassica and Blackleg

The plant-pathogen interaction between *Brassica* species and Blackleg or Phoma disease, caused by *Leptosphaeria maculans*, has been shown to involve both qualitative and quantitative resistance (Balesdent et al., 2001; Balesdent et al., 2002). Qualitative resistance is race-specific and depends on the presence of a single resistance (*R*) gene in the plant corresponding to an avirulence (*Avr*) gene in

the pathogen (Ansan-Melayah et al., 1998). Qualitative resistance is expressed from seedlings to adults in cotyledons and leaves (Delourme et al., 2006). Quantitative resistance is generally thought to be race non-specific, mediated by many genes and expressed at the adult plant stage, conferring only partial resistance to the pathogen (Delourme et al., 2006; Rimmer, 2006).

The first genetic evidence for the gene-for-gene interaction between *Brassica* species and *L. maculans* was identified between the *BnRlm1* (*Resistance to Leptosphaeria maculans 1*) resistance gene in *B. napus* cv. Quinta and the corresponding avirulence gene *AvrLm1* (Ansan-Melayah et al., 1998). Fifteen specific resistance genes have been subsequently identified in the cultivated *Brassica* species and are often clustered together on the genome (Delourme et al., 2006; Delourme et al., 2004; Rouxel et al., 2003).

At present, only one of the genetically mapped *Rlm* genes have been sequenced and validated in *B. napus*. However, the advances in Brassica genome sequencing and development of molecular markers have provided a basis for the identification of candidate genes. By performing sequence comparison of the blackleg *R* genes, and combining this with infection assays, using a range of *L. maculans* isolates that carry single (or very few) avirulence genes, the number and uniqueness of blackleg *R* genes can be confirmed. Recently two candidates for Rlm4 were identified (Tollenaere et al., 2012) in *B. napus* using a molecular marker approach combined with phenotypic data. The first Blackleg resistance gene in a *Brassica* species, *LepR3*, has also been cloned in *B. napus* lines with introgressions from *B. rapa* ssp. *sylvestris* (Larkan et al., 2013), with an allelic variant identified as Rlm2 (Larkan et al., 2014).

In *L. maculans*, eleven *Avr* genes have been genetically mapped. Of these five *Avr* genes (*AvrLm1*, *AvrLm6, AvrLm4-7, AvrLm11 and AvrLmJ1*), have been characterised (Balesdent et al., 2013; Fudal et al., 2007; Gout et al., 2006; Parlange et al., 2009; Van de Wouw et al., 2014). These genes display features typical of effector genes in that they are expressed early in infection and secrete small proteins into the apoplast (Stergiopoulos and de Wit, 2009). Rouxel et al. (2011) confirmed and physically mapped *AvrLm1, AvrLm4-7* and *AvrLm6* to the *L. maculans* genome. Previously, Fudal et al. (2009) used map-based cloning to map *AvrLm1* and *AvrLm6.* These genes were positioned within a repeat-induced point mutation (RIP) affected *Avr1-2-6* region of the genome that contains two kinds of isochores. Isochores can be defined as long stretches of DNA homogenous in GC content, with abrupt changes in GC content from one segment to another (Parlange et al., 2009). RIP is a type of pre-meiotic repeat inactivation mechanism that causes substitutions from C to T and G to A and has played a role in *L. maculans* genome amassing a large proportion of repeats, comprising one third of the genome (Rouxel et al., 2011). This RIP mechanism is postulated to intermittently cause

SNP mutations in *Avr* genes, as these are often located in regions of RIP activity. These polymorphisms may then be responsible for altered *Avr* gene sequences that can evade recognition by *R* genes.

The interaction of *Brassica* resistance and blackleg avirulence gene proteins at the molecular level has not been studied so far. Determining the crystal protein structure of the resistance and avirulence genes will help in identifying novel *R* and *Avr* genes.

Similarities, Differences and Challenges in understanding ETI

The multitude of studies of ETI in these model systems has led to a greater understanding of plant defence responses to a variety of pathogens with many similarities between the plant-pathogen systems. A major find has been the importance of NBS-LRR genes in the plant resistance response, with these genes being the most common *R* gene type in Flax, Brassica, Rice and Tomato. This is also the case in Arabidopsis and many other plant species. There is a trend for dicots to possess a higher number of TIR-type NBS-LRRs (e.g. Arabidopsis, Brassica, Tomato) and monocots to contain only CC-type NBS-LRRs. This may be due to the different ancestral lineages from which these plants descended and their subsequent diversification and specificity to their respective pathogens and diseases.

A difference can be seen in the downstream effects of pathogen recognition by *R* genes. In flax, brassica and rice, hypersensitive response seems to be the major defence response in controlling and limiting infection. However in the Tomato-Fusarium system, other factors such as callose deposition, phenolic production, and deposition of other protective barrier compounds seem to be the main consequences of ETI.

Avirulence genes appear to be quite unique to each pathogen and largely host-specific. They have been shown to exhibit similar characteristics in that they are generally small proteins, secreted into the host upon infection and play a role in infection success. *Avr* genes elicit *R*-gene mediated immune responses in the host and have been implicated in the suppression of host immune responses.

Some of the challenges in the identification and validation of resistance and avirulence genes can be seen in the examples described in this review. Avirulence genes have proven elusive due to their specificity to the host and with each individual pathogen. This often prevents a homology based

identification strategy within phylogenetic lineages of plant fungi. They can often possess great sequence diversity even within organisms, compounding this issue.

Resistance genes show a higher level of conservation between plant species and within phylogenetic lineages, which can be used to identify candidates. However plant genomes can contain hundreds of *R* genes, making it difficult to narrow down the correct candidate gene responsible for the interaction with a given effector gene.

Improving Methods of Gene-for-Gene Detection/Validation

The advent of next generation sequencing (NGS) technologies has allowed for significant improvements in the detection of avirulence and resistance genes. By generating relatively lost cost whole genome sequence data, it has become possible to study plant and pathogen genomes in high degrees of detail. By combining previously characterised genetic maps with a physical map of the genome, candidate genes within regions of interest (such as within QTLs) can lead to a more rapid identification of the gene/s responsible for plant-pathogen interactions, particularly ETI. The development of high density SNP arrays and using genotyping by sequencing permit very high density genetic mapping in a very short space of time and at a relatively low cost. QTL regions can then be narrowed rapidly and candidate genes identified. Re-sequencing of isolates or lines segregating for resistance or susceptibility allows sequence based validation of the candidate genes of interest.

Marker-assisted identification of candidate resistance genes can be performed between plants with different resistance gene composition. This was recently carried out using an *Rlm4* containing resistant *B. napus* line and a susceptible *B. napus* line without the *Rlm4* gene using SNP markers (Tollenaere et al., 2012). This approach has also been applied to pathogen effector study, such as in the corresponding pathogen of *B. napus, L. maculans* (Zander et al., 2013).

Genome-wide association studies (GWAS) using SNPs are another tool that can be implemented in identifying genes involved in ETI. This method can highlight polymorphisms that show a strong correlation towards a trait of interest, such as disease resistance in plants or avirulence/virulence in pathogen species. This has shown to be an effective method in narrowing down disease-related loci in several species such as in maize for head smut resistance (Wang et al., 2012), in wheat for stem rust resistance (Yu et al., 2011) and in rice for rice blast resistance (Zhao et al., 2011).

However, of the resistance genes or avirulence genes may not be present in the reference genome species. To address this and develop a pan-genomics approach to identify candidate genes, further methods that are being developed include studying presence/absence variations (PAVs) and gene

loss between different individuals of a species. Bush et al. (2014) found that disease resistance genes in Arabidopsis were overrepresented in having partial or total exon loss, indicating that PAVs may be a mechanism of adaptation to pathogen population. Other structural variation in genomes such as copy-number variations (CNVs) have also been implicated in disease resistance. Springer et al. (2009) describes the role of PAVs and CNVs in providing novel mechanisms of resistance allele adaptation in maize. Recently Golicz et al. (2015) presented a gene loss method in blackleg in which candidate avirulence genes were identified and known avirulence gene composition of isolates were predicted. Through utilising this on the rapidly expanding number of isolates being sequenced, novel avirulence genes could be very quickly identified.

Conclusions

The study of plant-pathogen interactions is vital in the task of feeding a growing world population. Understanding ETI will be an important part of ensuring plant species can continue to provide good yield and reduce environmental impact that comes from pesticide and herbicide use. The improving technologies for genome study, such as next generation sequencing and high density SNP arrays, are proving to be successful in increasing the speed of scientific discoveries in this area and will no doubt further improve in the future. Utilising the information together to develop the pan genome and understand variation from all re-sequencing projects will accelerate knowledge of these processes at an unprecedented rate as the technologies continue to develop.

References

Abdallah, N.A., Shah, D., Abbas, D. and Madkour, M. (2010) Stable integration and expression of a plant defensin in tomato confers resistance to fusarium wilt. *GM crops* **1**, 344-350.

Ajilogba, C.F. and Babalola, O.O. (2013) Integrated management strategies for tomato Fusarium wilt. *Biocontrol science* **18**, 117-127.

Ansan-Melayah, D., Balesdent, M.H., Delourme, R., Pilet, M.L., Tanguy, X., Renard, M. and Rouxel, T. (1998) Genes for race-specific resistance against blackleg disease in Brassica napus L. *Plant Breeding* **117**, 373-378.

Balesdent, M.H., Attard, A., Ansan-Melayah, D., Delourme, R., Renard, M. and Rouxel, T. (2001) Genetic control and host range of avirulence toward Brassica napus cultivars Quinta and Jet Neuf in Leptosphaeria maculans. *Phytopathology* **91**, 70-76.

Balesdent, M.H., Attard, A., Kuhn, M.L. and Rouxel, T. (2002) New Avirulence Genes in the Phytopathogenic Fungus Leptosphaeria maculans. *Phytopathology* **92**, 1122-1133.

Balesdent, M.H., Fudal, I., Ollivier, B., Bally, P., Grandaubert, J., Eber, F., Chevre, A.M., Leflon, M. and Rouxel, T. (2013) The dispensable chromosome of Leptosphaeria maculans shelters an effector gene conferring avirulence towards Brassica rapa. *New Phytol* **198**, 887-898.

Baumgarten, A., Cannon, S., Spangler, R. and May, G. (2003) Genome-level evolution of resistance genes in Arabidopsis thaliana. *Genetics* **165**, 309-319.

Bent, A.F. (1996) Plant disease resistance genes: Function meets structure. *Plant Cell* **8**, 1757-1771.

Block, A., Li, G., Fu, Z.Q. and Alfano, J.R. (2008) Phytopathogen type III effector weaponry and their plant targets. *Curr Opin Plant Biol* **11**, 396-403.

Bohnert, H.U., Fudal, I., Dioh, W., Tharreau, D., Notteghem, J.L. and Lebrun, M.H. (2004) A putative polyketide synthase/peptide synthetase from Magnaporthe grisea signals pathogen attack to resistant rice. *Plant Cell* **16**, 2499-2513.

Bos, J.I., Kanneganti, T.D., Young, C., Cakir, C., Huitema, E., Win, J., Armstrong, M.R., Birch, P.R. and Kamoun, S. (2006) The C-terminal half of Phytophthora infestans RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in Nicotiana benthamiana. *The Plant journal : for cell and molecular biology* **48**, 165-176.

Brun, H., Chevre, A.M., Fitt, B.D.L., Powers, S., Besnard, A.L., Ermel, M., Huteau, V., Marquer, B., Eber, F., Renard, M. and Andrivon, D. (2010) Quantitative resistance increases the durability of qualitative resistance to Leptosphaeria maculans in Brassica napus. *New Phytol* **185**, 285-299.

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**, 501-514.

Bush, S.J., Castillo-Morales, A., Tovar-Corona, J.M., Chen, L., Kover, P.X. and Urrutia, A.O. (2014) Presence-absence variation in A. thaliana is primarily associated with genomic signatures consistent with relaxed selective constraints. *Molecular biology and evolution* **31**, 59-69.

Catanzariti, A.M., Dodds, P.N., Lawrence, G.J., Ayliffe, M.A. and Ellis, J.G. (2006) Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* **18**, 243-256.

Chen, X., Shang, J., Chen, D., Lei, C., Zou, Y., Zhai, W., Liu, G., Xu, J., Ling, Z., Cao, G., Ma, B., Wang, Y., Zhao, X., Li, S. and Zhu, L. (2006) A B-lectin receptor kinase gene conferring rice blast resistance. *The Plant journal : for cell and molecular biology* **46**, 794-804.

Dangl, J.L. and Jones, J.D.G. (2001) Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826-833.

Delourme, R., Chèvre, A.M., Brun, H., Rouxel, T., Balesdent, M.H., Dias, J.S., Salisbury, P., Renard, M. and Rimmer, S.R. (2006) Major Gene and Polygenic Resistance to Leptosphaeria maculans in Oilseed Rape (Brassica napus). *European Journal of Plant Pathology* **114**, 41-52.

Delourme, R., Pilet-Nayel, M.L., Archipiano, M., Horvais, R., Tanguy, X., Rouxel, T., Brun, H., Renard, M. and Balesdent, M.H. (2004) A Cluster of Major Specific Resistance Genes to Leptosphaeria maculans in Brassica napus. *Phytopathology* **94**, 578-583.

Dodds, P. and Thrall, P. (2009) Recognition events and host-pathogen co-evolution in gene-for-gene resistance to flax rust. *Functional plant biology : FPB* **36**, 395-408.

Dodds, P.N., Lawrence, G.J., Catanzariti, A.M., Ayliffe, M.A. and Ellis, J.G. (2004) The Melampsora lini AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell* **16**, 755-768.

Dodds, P.N., Lawrence, G.J. and Ellis, J.G. (2001) Six amino acid changes confined to the leucine-rich repeat beta-strand/beta-turn motif determine the difference between the P and P2 rust resistance specificities in flax. *Plant Cell* **13**, 163-178.

Dou, D., Kale, S.D., Wang, X., Chen, Y., Wang, Q., Wang, X., Jiang, R.H., Arredondo, F.D., Anderson, R.G., Thakur, P.B., McDowell, J.M., Wang, Y. and Tyler, B.M. (2008) Conserved C-terminal motifs required for avirulence and suppression of cell death by Phytophthora sojae effector Avr1b. *Plant Cell* **20**, 1118-1133.

Eitas, T.K. and Dangl, J.L. (2010) NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Curr Opin Plant Biol* **13**, 472-477.

Ellis, J., Dodds, P. and Pryor, T. (2000) Structure, function and evolution of plant disease resistance genes. *Curr Opin Plant Biol* **3**, 278-284.

Ellis, J.G., Lawrence, G.J., Luck, J.E. and Dodds, P.N. (1999) Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. *Plant Cell* **11**, 495-506.

Ellis, J.G., Rafiqi, M., Gan, P., Chakrabarti, A. and Dodds, P.N. (2009) Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. *Curr Opin Plant Biol* **12**, 399-405.

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

Fudal, I., Ross, S., Brun, H., Besnard, A.-L., Ermel, M., Kuhn, M.-L., Balesdent, M.-H. and Rouxel, T. (2009) Repeat-Induced Point Mutation (RIP) as an Alternative Mechanism of Evolution Toward Virulence in Leptosphaeria maculans. *Molecular Plant-Microbe Interactions* **22**, 932-941.

Fudal, I., Ross, S., Gout, L., Blaise, F., Kuhn, M.L., Eckert, M.R., Cattolico, L., Bernard-Samain, S., Balesdent, M.H. and Rouxel, T. (2007) Heterochromatin-like regions as ecological niches for avirulence genes in the Leptosphaeria maculans genome: Map-based cloning of AvrLm6. *Molecular Plant-Microbe Interactions* **20**, 459-470.

Gassmann, W. and Bhattacharjee, S. (2012) Effector-Triggered Immunity Signaling: From Gene-for-Gene Pathways to Protein-Protein Interaction Networks. *Molecular Plant-Microbe Interactions* **25**, 862-868.

Gawehns, F., Houterman, P.M., Ichou, F.A., Michielse, C.B., Hijdra, M., Cornelissen, B.J., Rep, M. and Takken, F.L. (2014) The Fusarium oxysporum effector Six6 contributes to virulence and suppresses I-2-mediated cell death. *Molecular plant-microbe interactions : MPMI* **27**, 336-348.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A. and Briggs, S. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science* **296**, 92-100.

Gout, L., Fudal, I., Kuhn, M.L., Blaise, F., Eckert, M., Cattolico, L., Balesdent, M.H. and Rouxel, T. (2006) Lost in the middle of nowhere: the AvrLm1 avirulence gene of the Dothideomycete Leptosphaeria maculans. *Mol Microbiol* **60**, 67-80.

Guest, D.I. and Brown, J.F. (1997) Plant defences against pathogens. In: *Plant Pathogens and Plant Diseases* (Brown, J.F. and Ogle, H.J. eds), pp. 263-286. Armidale, Australia: Rockvale Publications.

Heath, M.C. (2000) Hypersensitive response-related death. *Plant Mol Biol* **44**, 321-334.

Hogenhout, S.A., Van der Hoorn, R.A.L., Terauchi, R. and Kamoun, S. (2009) Emerging Concepts in Effector Biology of Plant-Associated Organisms. *Molecular Plant-Microbe Interactions* **22**, 115-122.

Houterman, P.M., Cornelissen, B.J. and Rep, M. (2008) Suppression of plant resistance gene-based immunity by a fungal effector. *Plos Pathog* **4**, e1000061.

Houterman, P.M., Ma, L., van Ooijen, G., de Vroomen, M.J., Cornelissen, B.J., Takken, F.L. and Rep, M. (2009) The effector protein Avr2 of the xylem-colonizing fungus Fusarium oxysporum activates the tomato resistance protein I-2 intracellularly. *The Plant journal : for cell and molecular biology* **58**, 970-978.

Houterman, P.M., Speijer, D., Dekker, H.L., CG, D.E.K., Cornelissen, B.J. and Rep, M. (2007) The mixed xylem sap proteome of Fusarium oxysporum-infected tomato plants. *Mol Plant Pathol* **8**, 215-221.

Ichinose, Y., Taguchi, F. and Mukaihara, T. (2013) Pathogenicity and virulence factors of Pseudomonas syringae. *Journal of General Plant Pathology* **79**, 285-296.

Islam, M.R. and Mayo, G.M.E. (1990) A Compendium on Host Genes in Flax Conferring Resistance to Flax Rust. *Plant Breeding* **104**, 89-100.

Jia, Y., McAdams, S.A., Bryan, G.T., Hershey, H.P. and Valent, B. (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *Embo J* **19**, 4004-4014.

Lapin, D. and Van den Ackerveken, G. (2013) Susceptibility to plant disease: more than a failure of host immunity. *Trends in plant science* **18**, 546-554.

Larkan, N.J., Lydiate, D.J., Parkin, I.A., Nelson, M.N., Epp, D.J., Cowling, W.A., Rimmer, S.R. and Borhan, M.H. (2013) The Brassica napus blackleg resistance gene LepR3 encodes a receptor-like protein triggered by the Leptosphaeria maculans effector AVRLM1. *New Phytol* **197**, 595-605.

Lawrence, G.J., Anderson, P.A., Dodds, P.N. and Ellis, J.G. (2010) Relationships between rust resistance genes at the M locus in flax. *Mol Plant Pathol* **11**, 19-32.

Li, W., Wang, B.H., Wu, J., Lu, G.D., Hu, Y.J., Zhang, X., Zhang, Z.G., Zhao, Q., Feng, Q.Y., Zhang, H.Y., Wang, Z.Y., Wang, G.L., Han, B., Wang, Z.H. and Zhou, B. (2009) The Magnaporthe oryzae Avirulence Gene AvrPiz-t Encodes a Predicted Secreted Protein That Triggers the Immunity in Rice Mediated by the Blast Resistance Gene Piz-t. *Molecular Plant-Microbe Interactions* **22**, 411-420.

Lievens, B., Houterman, P.M. and Rep, M. (2009) Effector gene screening allows unambiguous identification of Fusarium oxysporum f. sp. lycopersici races and discrimination from other formae speciales. *FEMS microbiology letters* **300**, 201-215.

Liu, J., Liu, X., Dai, L. and Wang, G. (2007) Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants. *Journal of genetics and genomics = Yi chuan xue bao* **34**, 765-776.

Liu, J., Wang, X., Mitchell, T., Hu, Y., Liu, X., Dai, L. and Wang, G.L. (2010) Recent progress and understanding of the molecular mechanisms of the rice-Magnaporthe oryzae interaction. *Mol Plant Pathol* **11**, 419-427.

Lorang, J., Kidarsa, T., Bradford, C.S., Gilbert, B., Curtis, M., Tzeng, S.C., Maier, C.S. and Wolpert, T.J. (2012) Tricking the guard: exploiting plant defense for disease susceptibility. *Science* **338**, 659-662.

Luck, J.E., Lawrence, G.J., Dodds, P.N., Shepherd, K.W. and Ellis, J.G. (2000) Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. *Plant Cell* **12**, 1367-1377.

Ma, L.J., van der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P.M., Kang, S., Shim, W.B., Woloshuk, C., Xie, X., Xu, J.R., Antoniw, J., Baker, S.E., Bluhm, B.H., Breakspear, A., Brown, D.W., Butchko, R.A., Chapman, S., Coulson, R., Coutinho, P.M., Danchin, E.G., Diener, A., Gale, L.R., Gardiner, D.M., Goff, S., Hammond-Kosack, K.E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C.D., Koehrsen, M., Kumar, L., Lee, Y.H., Li, L., Manners, J.M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S.Y., Proctor, R.H., Regev, A., Ruiz-Roldan, M.C., Sain, D., Sakthikumar, S., Sykes, S., Schwartz, D.C., Turgeon, B.G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C.A., Kistler, H.C. and Rep, M. (2010) Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. *Nature* **464**, 367-373.

McHale, L., Tan, X., Koehl, P. and Michelmore, R.W. (2006) Plant NBS-LRR proteins: adaptable guards. *Genome biology* **7**, 212.

Melotto, M., Underwood, W., Koczan, J., Nomura, K. and He, S.Y. (2006) Plant stomata function in innate immunity against bacterial invasion. *Cell* **126**, 969-980.

Mes, J.J., Weststeijn, E.A., Herlaar, F., Lambalk, J.J., Wijbrandi, J., Haring, M.A. and Cornelissen, B.J. (1999) Biological and Molecular Characterization of Fusarium oxysporum f. sp. lycopersici Divides Race 1 Isolates into Separate Virulence Groups. *Phytopathology* **89**, 156-160.

Meyers, B.C., Dickerman, A.W., Michelmore, R.W., Sivaramakrishnan, S., Sobral, B.W. and Young, N.D. (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *The Plant journal : for cell and molecular biology* **20**, 317-332.

Michielse, C.B. and Rep, M. (2009) Pathogen profile update: Fusarium oxysporum. *Mol Plant Pathol* **10**, 311-324.

Park, C.H., Chen, S., Shirsekar, G., Zhou, B., Khang, C.H., Songkumarn, P., Afzal, A.J., Ning, Y., Wang, R., Bellizzi, M., Valent, B. and Wang, G.L. (2012) The Magnaporthe oryzae effector AvrPiz-t targets the RING E3 ubiquitin ligase APIP6 to suppress pathogen-associated molecular pattern-triggered immunity in rice. *Plant Cell* **24**, 4748-4762.

Parlange, F., Daverdin, G., Fudal, I., Kuhn, M.L., Balesdent, M.H., Blaise, F., Grezes-Besset, B. and Rouxel, T. (2009) Leptosphaeria maculans avirulence gene AvrLm4-7 confers a dual recognition specificity by the Rlm4 and Rlm7 resistance genes of oilseed rape, and circumvents Rlm4-mediated recognition through a single amino acid change. *Mol Microbiol* **71**, 851-863.

Periyannan, S., Moore, J., Ayliffe, M., Bansal, U., Wang, X., Huang, L., Deal, K., Luo, M., Kong, X., Bariana, H., Mago, R., McIntosh, R., Dodds, P., Dvorak, J. and Lagudah, E. (2013) The gene Sr33, an ortholog of barley Mla genes, encodes resistance to wheat stem rust race Ug99. *Science* **341**, 786-788.

Persson, M., Staal, J., Oide, S. and Dixelius, C. (2009) Layers of defense responses to Leptosphaeria maculans below the RLM1- and camalexin-dependent resistances. *New Phytol* **182**, 470-482.

Rafiqi, M., Gan, P.H.P., Ravensdale, M., Lawrence, G.J., Ellis, J.G., Jones, D.A., Hardham, A.R. and Dodds, P.N. (2010) Internalization of Flax Rust Avirulence Proteins into Flax and Tobacco Cells Can Occur in the Absence of the Pathogen. *Plant Cell* **22**, 2017-2032.

Ravensdale, M., Nemri, A., Thrall, P.H., Ellis, J.G. and Dodds, P.N. (2011) Co-evolutionary interactions between host resistance and pathogen effector genes in flax rust disease. *Mol Plant Pathol* **12**, 93-102.

Rep, M. (2005) Small proteins of plant-pathogenic fungi secreted during host colonization. *FEMS microbiology letters* **253**, 19-27.

Rep, M., Meijer, M., Houterman, P.M., van der Does, H.C. and Cornelissen, B.J. (2005) Fusarium oxysporum evades I-3-mediated resistance without altering the matching avirulence gene. *Molecular plant-microbe interactions : MPMI* **18**, 15-23.

Rimmer, S.R. (2006) Resistance genes to Leptosphaeria maculans in Brassica napus. *Canadian Journal of Plant Pathology* **28**, S288-S297.

Rouxel, T. and Balesdent, M.H. (2005) The stem canker (blackleg) fungus, Leptosphaeria maculans, enters the genomic era. *Mol Plant Pathol* **6**, 225-241.

Rouxel, T., Grandaubert, J., Hane, J.K., Hoede, C., van de Wouw, A.P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., Cozijnsen, A.J., Ciuffetti, L.M., Degrave, A., Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S.B., Gout, L., Glaser, N., Linglin, J., Kema, G.H.J., Lapalu, N., Lawrence, C.B., May, K., Meyer, M., Ollivier, B., Poulain, J., Schoch, C.L., Simon, A., Spatafora, J.W., Stachowiak, A., Turgeon, B.G., Tyler, B.M., Vincent, D., Weissenbach, J., Amselem, J., Quesneville, H., Oliver, R.P., Wincker, P., Balesdent, M.-H. and Howlett, B.J. (2011) Effector diversification within compartments of the Leptosphaeria maculans genome affected by Repeat-Induced Point mutations. *Nature Communications* **2**.

Rouxel, T., Willner, E., Coudard, L. and Balesdent, M.-H. (2003) Screening and identification of resistance to Leptosphaeria maculans (stem canker) in Brassica napus accessions. *Euphytica* **133**, 219-231.

Saintenac, C., Zhang, W., Salcedo, A., Rouse, M.N., Trick, H.N., Akhunov, E. and Dubcovsky, J. (2013) Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. *Science* **341**, 783-786.

Simons, G., Groenendijk, J., Wijbrandi, J., Reijans, M., Groenen, J., Diergaarde, P., Van der Lee, T., Bleeker, M., Onstenk, J., de Both, M., Haring, M., Mes, J., Cornelissen, B., Zabeau, M. and Vos, P. (1998) Dissection of the fusarium I2 gene cluster in tomato reveals six homologs and one active gene copy. *Plant Cell* **10**, 1055-1068.

Springer, N.M., Ying, K., Fu, Y., Ji, T., Yeh, C.T., Jia, Y., Wu, W., Richmond, T., Kitzman, J., Rosenbaum, H., Iniguez, A.L., Barbazuk, W.B., Jeddeloh, J.A., Nettleton, D. and Schnable, P.S. (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS genetics* **5**, e1000734.

Staskawicz, B.J., Ausubel, F.M., Baker, B.J., Ellis, J.G. and Jones, J.D. (1995) Molecular genetics of plant disease resistance. *Science* **268**, 661-667.

Stergiopoulos, I. and de Wit, P.J. (2009) Fungal effector proteins. *Annu Rev Phytopathol* **47**, 233-263.

Stotz, H.U., Thomson, J.G. and Wang, Y. (2009) Plant defensins: defense, development and application. *Plant signaling & behavior* **4**, 1010-1012.

Takken, F. and Rep, M. (2010) The arms race between tomato and Fusarium oxysporum. *Mol Plant Pathol* **11**, 309-314.

Thatcher, L.F., Gardiner, D.M., Kazan, K. and Manners, J.M. (2012) A highly conserved effector in Fusarium oxysporum is required for full virulence on Arabidopsis. *Molecular plant-microbe interactions : MPMI* **25**, 180-190.

Tollenaere, R., Hayward, A., Dalton-Morgan, J., Campbell, E., Lee, J.R.M., Lorenc, M.T., Manoli, S., Stiller, J., Raman, R., Raman, H., Edwards, D. and Batley, J. (2012) Identification and characterization of candidate Rlm4 blackleg resistance genes in Brassica napus using next-generation sequencing. *Plant Biotechnol J* **10**, 709-715.

Tudzynski, B. (1999) Biosynthesis of gibberellins in Gibberella fujikuroi: biomolecular aspects. *Applied microbiology and biotechnology* **52**, 298-310.

Van de Wouw, A.P., Lowe, R.G., Elliott, C.E., Dubois, D.J. and Howlett, B.J. (2014) An avirulence gene, AvrLmJ1, from the blackleg fungus, Leptosphaeria maculans, confers avirulence to Brassica juncea cultivars. *Mol Plant Pathol* **15**, 523-530.

van der Does, H.C., Duyvesteijn, R.G., Goltstein, P.M., van Schie, C.C., Manders, E.M., Cornelissen, B.J. and Rep, M. (2008a) Expression of effector gene SIX1 of Fusarium oxysporum requires living plant cells. *Fungal Genet Biol* **45**, 1257-1264.

van der Does, H.C., Lievens, B., Claes, L., Houterman, P.M., Cornelissen, B.J. and Rep, M. (2008b) The presence of a virulence locus discriminates Fusarium oxysporum isolates causing tomato wilt from other isolates. *Environmental microbiology* **10**, 1475-1485.

Wan, H., Yuan, W., Bo, K., Shen, J., Pang, X. and Chen, J. (2013) Genome-wide analysis of NBS-encoding disease resistance genes in Cucumis sativus and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. *Bmc Genomics* **14**, 109.

Wan, H.J., Yuan, W., Ye, Q.J., Wang, R.Q., Ruan, M.Y., Li, Z.M., Zhou, G.Z., Yao, Z.P., Zhao, J., Liu, S.J. and Yang, Y.J. (2012) Analysis of TIR- and non-TIR-NBS-LRR disease resistance gene analogous in pepper: characterization, genetic variation, functional divergence and expression patterns. *Bmc Genomics* **13**.

Wang, M., Yan, J., Zhao, J., Song, W., Zhang, X., Xiao, Y. and Zheng, Y. (2012) Genome-wide association study (GWAS) of resistance to head smut in maize. *Plant science : an international journal of experimental plant biology* **196**, 125-131.

White, F.F., Yang, B. and Johnson, L.B. (2000) Prospects for understanding avirulence gene function. *Curr Opin Plant Biol* **3**, 291-298.

Yi, H. and Richards, E.J. (2007) A cluster of disease resistance genes in Arabidopsis is coordinately regulated by transcriptional activation and RNA silencing. *Plant Cell* **19**, 2929-2939.

Yu, L.-X., Lorenz, A., Rutkoski, J., Singh, R., Bhavani, S., Huerta-Espino, J. and Sorrells, M. (2011) Association mapping and gene–gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. *Theor Appl Genet* **123**, 1257-1268.

Zander, M., Patel, D.A., Van de Wouw, A., Lai, K.T., Lorenc, M.T., Campbell, E., Hayward, A., Edwards, D., Raman, H. and Batley, J. (2013) Identifying genetic diversity of avirulence genes in Leptosphaeria maculans using whole genome sequencing. *Funct Integr Genomic* **13**, 294-308.

Zhai, C., Lin, F., Dong, Z., He, X., Yuan, B., Zeng, X., Wang, L. and Pan, Q. (2011) The isolation and characterization of Pik, a rice blast resistance gene which emerged after rice domestication. *New Phytol* **189**, 321-334.

Zhao, K., Tung, C.-W., Eizenga, G.C., Wright, M.H., Ali, M.L., Price, A.H., Norton, G.J., Islam, M.R., Reynolds, A., Mezey, J., McClung, A.M., Bustamante, C.D. and McCouch, S.R. (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. *Nat Commun* **2**, 467.

Zipfel, C. (2009) Early molecular events in PAMP-triggered immunity. *Curr Opin Plant Biol* **12**, 414-420.