Characterising populations of *Puccinia striiformis* f. sp. *tritici* (*Pst*) using molecular markers

By

Adeelah Alvi, BSc, MSc

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

September 2016

Abstract

As the world population is projected to increase from 6 to 9 billion by 2050, the demand for global food production is rising (United Nations, 2013). Protecting domesticated crops has become a challenge due to the emergence of pathogenic diseases, which are restricting environmentally sustainable food production. The plant pathogen *Puccinia striiformis* f. sp. *tritici* (*Pst*) most commonly known as yellow rust, is one of the devastating diseases of wheat (*Triticum* spp.), a major food source globally (Cantu *et al.*, 2011).

Rust fungi are an important group of pathogens that pose major concerns for agriculture globally as wheat is a host to three rust diseases, yellow, brown and stem rust. *Puccinia striiformis* f. sp. *tritici* (*Pst*) is a fungus producing dikaryotic urediniospores that spread sporadically causing global epidemics of wheat yellow rust. Serious implications of this rust pathogen have been reported as it continues to spread in various wheat growing regions of the world including China (Duan *et al.*, 2011), Pakistan (Bux *et al.*, 2011), Australia (Welling, 2007), the USA (Milus *et al.*, 2006), and the UK (Cantu *et al.*, 2013).

Losses are particularly high in cases when the epidemic is associated with an exotic strain that has not been considered in breeding for resistance locally (Walter *et al.*, 2016). Recently, the evolution of yellow rust pathotypes has been documented by Hovmoller *et al.* (2016); this study details the migration of rust pathotypes worldwide. They illustrated the emergence of new exotic races displaying a significantly higher genetic diversity than a previous clonal population. The Himalayan and near-Himalayan regions have been documented as the centre of diversity, as this is the place where most sexual reproduction appears to occur (Ali *et al.*, 2010; Ali *et al.*, 2014), whilst in other parts of the world the evidence is that there are clonal introductions that then evolve by stepwise mutation (Wellings *et al.*, 2007; Ali *et al.*, 2014, Hubbard *et al.*, 2015; Hovmoller *et al.*, 2015). This study confirmed these findings with data from Illumina sequencing in chapter 3 and AFLP results from chapter 2 that display an overall distinction between new and old isolates.

Developing a better understanding of pathogenicity and virulence factors is essential to reduce the severity of this obligate biotrophic fungus. The aim of this research project was to identify molecular markers linked to specific pathotypes of yellow rust through a random molecular marker approach such as Amplified Fragment Length Polymorphisms (AFLP), and also identify molecular markers such as Single Nucleotide Polymorphisms (SNPs) in potential effector genes, using a targeted NGS approach (Illumina Miseq). Both approaches enabled the comparison of genomes of old and new pathotypes of *Pst*, and enabled us to explore the association of both types of markers with virulence/avirulence phenotypes of isolates, with the aim of developing markers that could be used to test collections of spore samples using rapid molecular techniques to identify their pathotypes.

Using a targeted approach and sequencing potential effector protein regions, identifying presence/absence of genes correlated with presence/absence of avirulence genes using data obtained from chapter 3 proved to be a successful method. This thesis found that many of the genes correlated with the older virulences, and only one gene (*Pst*130_4892) shared any correlation with a new virulence (*Avr*Ste). In summary, it can be concluded that genes *Pst*43_3149, *Pst*877_18061, *Pst*0821_12746, *Pst*877_3038, *Pst*43_20465, *Pst*43_2877, *Pst*21_14230, *Pst*21_18506, *Pst*877_4028, *Pst*130_4892 from the 147 studied potential effector proteins, could be known avirulence genes, since polymorphisms and/or presence/absence of them showed correlation with particular *Avrs*.

Anlaysing AFLP data made it possible to correlate specific AFLP markers to virulence/avirulence traits. The sufficiency of markers to distinguish between the two distinct populations was reflected through the correlation with virulences. Virulence for the resistant genes *Yr1*, *Yr2, Yr3*, *Yr4, Yr9*, *and Yr17* was identified in most isolates suggesting these virulences are now fixed in the population.

Overall, this thesis has identified some markers potentially linked to specific virulence phenotypes. Future studies, using more isolates and gene expression studies should focus on confirming which markers to develop further so that a toolkit of primers can be developed for rapid identification of pathotypes in the field.

Acknowledgments

First and foremost, I would like to express my gratitude to my primary supervisor Professor Matthew Dickinson for your guidance, encouragement and patience throughout the course of my PhD. Your unwavering support and knowledge has been invaluable in completing this thesis and has enabled me to grow as a research scientist. Thank you.

I would also like to thank my supervisors in FERA, Dr Judith Turner and Dr Jeff Peters. Although Dr Jeff Peters was not around for the last three years of the completion of this thesis, I still greatly appreciate his help and guidance during my time in FERA. I would also like to take this opportunity to show my appreciation towards Dr Steve Rossall from the University of Nottingham, for the great advice and support you have always offered me.

The financial support from the Biotechnology and Biological Sciences Research Council (BBSRC) for carrying out this research is greatly acknowledged.

I would especially like to thank Dr Ian Adams for providing his expertise on DNA sequencing technologies, and thus supporting me in sequencing my own rust samples. I greatly appreciate Dr Toby Hodges help and support during the bioinformatics analyses. With no previous background in this area, Toby's guidance was extremely valuable. My thanks and gratitude also extends to the technical staff at the University of Nottingham that made this thesis possible, Linda Staniforth for your assistance in all matters relating to AFLPs and Mark Meacham for handling all glasshouse issues for my inoculation experiments.

To all my friends in labA58, you have all made my experience in the lab an enjoyable one, particularly Praphat, Mai, Bryar, Naofel and Kwasi- you guys have always been a great source of support, thanks for your ongoing encouragement and advice, I will miss my time in the lab with you all! Also, thank you to all my friends within and outside of academia, whose encouraging words kept me going when coffee had lost its stimulating effect.

Last but not least, I would like to thank my family who have supported me throughout my academic career, but also for everything I do in life. Dad, thanks for teaching us how important it is to develop into an individual who tries to constantly make positive change, I believe a career in research can be a worthy part of that pursuit. Mum, your blessings and encouragement to believe in myself cannot be thanked enough. You have both been my biggest support and without you this PhD would not be possible.

Baji and Sanna, I can’t thank you enough for always encouraging and reassuring me throughout this experience. Thank you for putting up with my emotional highs and lows. Saab and Areebah, despite being stressed or overwhelmed you both always found a way to make me smile. I love you all.

This thesis is dedicated to my parents without whom I would have not achieved so much.

List of abbreviations

AFLP	Amplified Fragment Length Polymorphisms
AN	Total number of alleles in called genotypes
APR	Adult plant resistance
ASR	All-stage resistance
Avr	Avirulence gene
BCF	Binary variant call format
Bgyr	Barley grass yellow rust
BWA	Burrows-Wheeler Alignment
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide
dsDNA	Double stranded DNA
EHM	Extrahaustorial matrix
ETI	Effector-triggered immunity
FAM	5’,6’-carboxyfluorescein
FERA	Food Environment Research Agency
HESPs	Haustorially expressed secreted proteins
HEX	5’,4’,7’,2’,4’,5’,7-hexachloro-6-carboxyfluorescein
HTAP	High temperature- adult plants resistance
HR	Hypersensitive response
MEGA	Molecular Evolutionary Genetics Analysis
M. lini	*M. larici-populina*

NGS	Next generation sequencing
NIAB	National Institute of Agricultural Botany
NLS	Nuclear Localisation Signals
MAPK	Mitogen Activated Protein Kinase
MDA	Multiple Displacement Amplification
PAMPs	pathogen associated molecular patterns
PCR	Polymerase Chain Reaction
Pga	P. graminis f. sp. avenae
Pgt	*P. graminis* f. sp. *tritici*
PRRs	pattern recognition receptors
Pst	*Puccinia striiformis* f. sp. *tritici*
PTI	PAMP triggered immunity
R	Resistance gene
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RNA baits	Biotinylated RNA Library
RTP1	Rust Transferred Protein 1
SAM	Selectively amplified microsatellites
SAM	Sequence Alignment Map
SB	Strand bias for a particular position
S-SAP	Sequence-specific amplification polymorphisms

SNP	Single Nucleotide Polymorphism
SOLiD	Sequencing by Oligo Ligation Detection
SSR	Simple sequence repeats
STRs	Short tandem repeats
UKCPVS	UK, Cereal Pathogen Virulence Survey
UV	Ultraviolet
VCF	Variant call format

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Chapter 1: Introduction

1.1 General Introduction

As the world population is projected to increase from 6 to 9 billion by 2050, the demand for global food production is rising (United Nations, 2013). Protecting domesticated crops has become a challenge due to the emergence of pathogenic diseases, which are restricting environmentally sustainable food production. Food Security (defined as "when all people, at all times, have physical and economic access to sufficient, safe, and nutritious food to meet their dietary needs and food preferences for an active and healthy life" (World Food Summit (FAO), 1974) is threatened by plant pathogens, causing significant losses globally to major cereal crops such as wheat, rice and maize.

For this reason, plant pathologists must confront the challenge of pathogens in crop production systems in order to enhance food security (Gregory *et al.*, 2009). In 2016, wheat (*Triticum aestivum*) was a major contributor towards improving global food security as one of the most produced crops (alongside rice and maize) in the world at 723 million tons (FAO, 2016). Additionally, current and future concerns for sustaining the quality and production of wheat in the United Kingdom (UK) have been prominent. Demands for food production is forecasted to increase by 70% by 2050 (Food and Agriculture Organization, 2009), placing pressures on research groups to focus on protection against pathogenic organisms, and developing reliable detection methods for these causal agents.

Rust fungi are an important group of pathogens that pose major concerns for agriculture globally as wheat is a host to three rust diseases, yellow, brown and stem rust. *Puccinia striiformis* f. sp. *tritici* (*Pst*) is a fungus producing dikaryotic urediniospores that spread sporadically causing global epidemics of wheat yellow rust. Serious implications of this rust pathogen have been reported as it continues to spread in various wheat growing regions of the world including China (Duan *et al.*, 2010), Pakistan (Bux *et al.*, 2011), Australia (Wellings, 2007), the USA (Milus et al., 2006), and the UK (Cantu *et al., 2013)*. These studies support that the fungus has a debilitating effect on young plants and on adult plants because yield losses caused by wheat yellow rust will affect grain quality and germination due to the production of low vigour seeds (Roelfs, 1992, Chen, 2005).

Loss in wheat yields is dependent upon factors such as cultivar susceptibility and the stage of infection, as yield loss can differ from 10% (where wheat varieties are less susceptible and initial infection occurred later), to 70% (where cultivar susceptibility is much higher and initial infection occurs early and continues during the growing season), with some cases even resulting in 100% yield losses (Chen, 2005). Currently, developing rust resistance in wheat is the most economical method of control (Hammond-Kosack and Jones, 1997, Gururani *et al.,* 2012).

The extent of disease epidemics can be due to a number of factors, such as global trade and climate change, with the latter inducing changes to biology, habitat and further affecting social-economic responses (Wilkinson *et al.*, 2011). Over time, wider prevalence of plant pathogens has been documented (Kumar *et al.*, 2002, Dean *et al.*, 2012) illustrating the constant challenge between new diseases and plant defence systems, and the growing need for more sophisticated control of diseases such as use of plant disease resistance genes (*R*) (Fig.1-1).

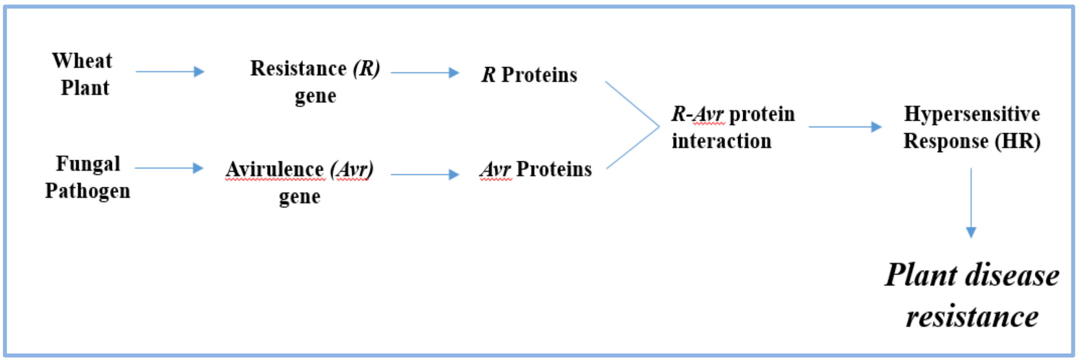


Figure 1-1 Illustration of plant-pathogen interactions. Plant resistant (*R*) genes, recognise a corresponding avirulence (*Avr*) gene in the pathogen, which activates a defence response.

1.1.1 The gene for gene theory of plant pathogen interactions

Following the independent genetic experiments of Flor and Oort in the 1940's the gene-for-gene theory was devised. Flor was working with *Melampsora lini* and flax in America, whilst Oort was working with loose smut (*Ustilago tritici)* and wheat in Europe. Both scientists noted that "for each gene determining resistance in the host, there is a specific corresponding gene in the fungus determining virulence". This simple statement was proposed and has now been refined as an established phenomenon in plant disease resistance mechanisms (Flor, 1971).

The gene-for-gene theory states that for every resistance gene (*R*) in the host there is a corresponding avirulence gene (*Avr*) in the pathogen. Plants producing specific *R* gene products are resistant towards a pathogen that produces the corresponding *Avr* gene product. Resistance genes in plants are normally dominant and constitutively expressed. ‘Avirulence’ genes in pathogens are also dominant and encode the specific ‘effectors’ recognised by the *R* gene products. Fundamentally, the plant is only resistant if it carries an *R* gene and is infected by a pathogen with a corresponding avirulence gene. This concept can be put into a quadratic check as shown in Table 1-1. This simple model holds true for most biotrophic pathogens, including fungi, viruses, and bacteria (Hammond - Kosack and Kanyuka, 2007). It is the nature of these resistance and avirulence genes that will be the focus of this thesis.

Table 1-1 Quadratic table which illustrates Flor’s gene-for-gene theory. Flax, similarly to wheat has two copies of every chromosome. Therefore, there are two alleles of every gene in the host i.e. RR (homozygous, resistant), Rr heterozygous, (one allele confers resistance whilst the other allele does not; the host will be completely resistant as ‘R’ is dominant over ‘r’), rr (homozygous, not resistant). This can also be explained for the pathotype genotype where; AA (homozygous, avirulent), Aa (heterozygous, one allele confers avirulence whilst the other does not; the pathogen will act completely avirulent because ‘A’ is dominant over ‘a’), aa (homozygous, virulent).

		Host Genotype	
		RR or Rr	**rr**
Pathotype Genotype	**AA or Aa**	Resistance	Disease
	aa	Disease	Disease

1.1.2 Plant defence mechanism and disease resistance

As pathogens have evolved to colonise living plants, plants have developed a wide range of passive and active defence mechanisms to supress their presence. The specificity of plant responses to pathogens can be classified into two categories, non-specific host resistance and specific resistance. Non-specific resistance is general resistance that is a response to all races of a particular pathogen and occurs in all cultivars of a potential host species (the entire plant species is resistant and therefore it is not a host to the pathogen). Specific host resistance on the other hand, is dependent upon the presence of a particular pathogen and a particular host plant cultivar. A single gene (or a small number of related genes) governs plant disease resistance in specific host resistance. The genetically inherited forms of defence are developed by the host species against the pathogen (Dickinson, 2003). For example, wheat is infected by yellow rust, *P. striiformis*, but certain genotypes will be resistant to specific races of the fungus.

Interactions between plant and pathogens are recognised as compatible interactions and incompatible interactions. In compatible interactions the pathogen is able to strive and dominate in the plant, causing disease. This is a result of the pathogen *Avr* product or the host *R* product being absent or changed. In incompatible interactions, recognition between the plant *R* gene product and the corresponding pathogenic *Avr* gene product lead to limited or no growth of the pathogen, resulting in very few symptoms, or complete absence of the disease. The result of an incompatible interaction is often a hypersensitive response (HR), a localised plant cell death as a plant defence response. Therefore, the presence of a resistance gene in susceptible wheat cultivars exerts a robust selection pressure for a mutation in the avirulence gene, resulting in severe epidemiological consequences.

1.1.3 Plant pathogen interaction and resistance

Understanding plant-pathogen interactions for crop improvement is important. Plants and pathogens have evolved to defend and manipulate, respectively. Fungal pathogens have special characteristics that allow them to colonise the plant host and infect. Like animals, plants have an innate immune system, which recognises pathogens and suppression of this host resistance will result in successful infection. Plants have two layers of active defence; the first is dominated by pattern recognition receptors (PRRs) which trigger the immune response known as PAMP triggered immunity (PTI) through recognition of conserved pathogen associated molecular patterns (PAMPs). The first level of defence requires the plant to recognise PAMPS, which triggers the plants general defence system known as PTI. The pathogen is able to supress components of PTI by ‘effector proteins’ delivered into the plant (Boyd *et al.*, 2013).

The second level of defence involves the activation of resistance (*R*) proteins upon contact of avirulence (*Avr*) effectors of the pathogen (Liu, 2013). This level of defence involves the effector proteins of the pathogen to be recognised by the resistance (*R*) genes of the plant. The specific effector proteins are referred to as Avirulence (Avr) proteins, and once recognised they produce a stronger resistance response known as effector-triggered immunity (ETI) or gene-for-gene resistance. If effector proteins successfully manipulate and alter the cellular environment of the plant to produce more suitable conditions for its growth and reproduction, successful infection will occur (Bernoux *et al.*, 2014).

This thesis will focus on this leading hypothesis of specific host resistance in yellow rust, where pathogens have developed effector proteins that can supress this broad system of resistance. Understanding plant-pathogen interactions involved in resistant cultivars particularly to yellow rust of wheat will have implications in breeding for ‘durable’ resistance i.e., resistance which remains effective for the long term in large epidemiological areas (Johnson, 1984).

1.2 Yellow Rust of Wheat

1.2.1 Taxonomy and Host Range

The wheat yellow rust pathogen, *Puccinia striiformis* Westend f. sp. *tritici* Eriksson (henceforth abbreviated to *Pst*) is the causal agent of yellow rust (also commonly known as stripe rust) on wheat. *Pst* belongs to the family *Pucciniaceae*, order *Uredinales* (more commonly known as *pucciniales*), class *Basidiomycetes* and genus *Puccinia*. This is common with wheat brown rust (*Puccinia triticina*) that shares a similar classification description (Chen *et al.*, 2014).

Pst has a wide host range including wheat, rye and barley (Table 1-2). *Pst* infects a range of different wheats including common wheat (*Triticum aestivum L.*), durum wheat (*T. turgidum var. durum L.*) and triticale (*Triticosecale*) causing major epidemics (Chen *et al.*, 2014). However, it can be further classified into formae speciales (forms), which are adapted to causing disease on specific host genera (Table 1-2) (Wellings, 2011).

Table 1-2 Host range of *Puccinia striiformis* f. sp. *tritici* (*Pst*) *Within the forms of *Puccinia striiformis*, *Puccinia striiformis* f. sp. *hordei* only attacks barley.

Formae speciales	**Major Hosts infected by *Pst***		**Minor Hosts infected by *Pst***	
	Scientific name	**Common name**	**Scientific name**	**Common name**
Puccinia striiformis* f. sp. *tritici	*Triticum aestivum* L.	Common wheat	*Secale cereale* L.	Rye
	T. dicoccum Schrank	Cultivated emmer wheat	*Hordeum vulgare* L.	Barley*
	T. dicoccoides Korn	Wild emmer wheat		
	T. turgidum var. *durum* L.	Durum wheat		
	Tritico secale	Triticale		

1.2.2 Life Cycle and Spore type

Rust fungi have complex lifecycles, which can include up to five different spore stages and two different host plants (Holtz *et al.*, 2013). Yellow rust is an obligate biotrophic parasite, and cannot be cultured on artificial media. *Pst* was long regarded as an autoecious microcyclic fungus, as it was thought to complete its lifecycle on a single host species, with urediospores and teliospores forming on the grass hosts. Several unsuccessful attempts were made in the 20th century to identify the alternate host of *P. striiformis* (Stubbs, 1985). However, recently, Jin *et al.* (2010; 2011) identified that *Pst* is a macrocyclic, heterecious fungus that requires two living hosts to complete its lifecycle. These include the primary host (wheat) for asexual reproduction and an alternate host plant (Barberry, *Berberis vulgaris* or *Mahonia aquifolium*) supporting pycniospores and aeciospores for sexual reproduction (Jin, 2011, Zhao *et al.*, 2013, Wang and Chen, 2015, Zhao *et al.*, 2011). Thus, this macrocyclic fungus produces five spore states during its sexual and asexual stages. These include pycniospores (spermatia), aeciospores, urediospores, teliospores, and basidiospores (Sajid, 2013). Table 1-3 below highlights the spore characteristics.

Table 1-3 Spore types produced during the yellow rust lifecycle and their characteristics.

Spore Type	Spore Classification
Pycniospores (Spermatia)	Haploid gametes produced within pycnia or pynidia.
Aeciospores	Non-repeating dikaryotic (n+n) vegetative spores, produced by an aecium (a specialised reproductive structure in rusts).
Urediospores	Repeating dikaryotic (n+n) vegetative spores produced by a uredinium that can re-infect the same host plant from which the spores were produced. They are red/orange/yellow and are characteristic of a rust infection.
Teliospores	Diploid spores produced by a telium that produce basidia. They are the survival stage of the life cycle.
Basidiospores	Haploid spores produced by a basidium that infect the alternate host.

Asexual urediospores are produced in the uredinia, passing through several asexual cycles. As temperatures rise later in the epidemic phase there is production of teliospores that are made in the telia. Each mature teliospore cell contains a diploid (2n) nucleus. At the telial stage there is no direct effect on the infection process and there is a small period of dormancy. Teliospores undergo meiosis and produce basidium, developing into 4 basidiospores, which as a result infect the alternate host (barberry). Basidiospores germinate and infect *Berberis* spp. or *Mahonia* spp. although there has been very limited confirmation of *Pst* on berberis with only two studies in China to date (Zhao *et al.*, 2011, Zhao *et al.*, 2013). Infection of berberis leads to the production of pycnia on adaxial leaf surfaces, showing disease symptoms. Fertilised aecia become apparent on the abaxial leaf, and dikaryotic aeciospores that are previously produced can repeatedly infect wheat plants to produce dikaryotic urediospores (Fig. 1-2) (Jin, 2011).

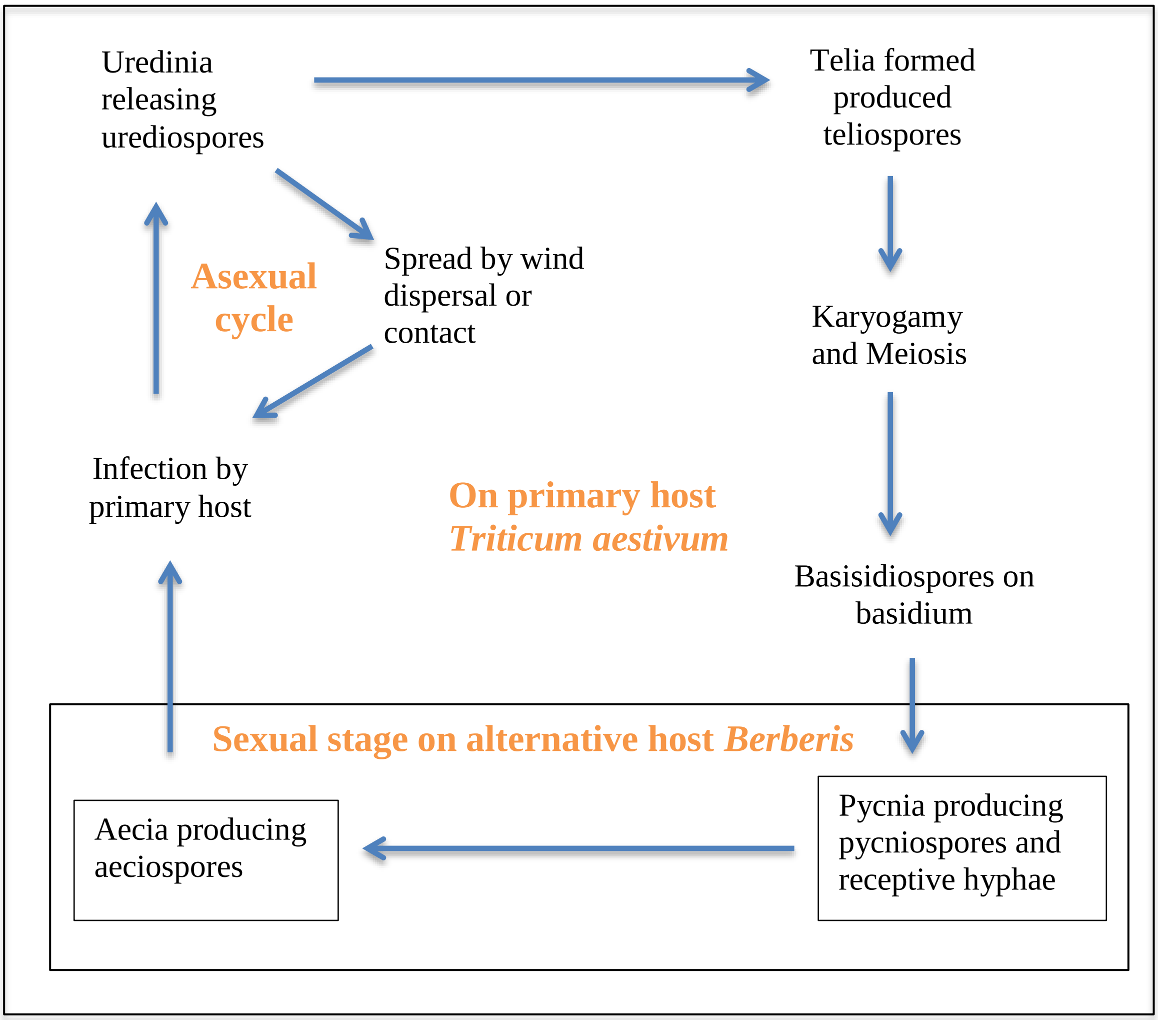


Figure 1-2 Schematic diagram of the heteroecial yellow rust lifecycle, showing the asexual life cycle of reinfection of urediospores - the only inoculum for primary and recurrent infection of wheat plants. Sexual stage on alternative host is also displayed.

1.1.1 Symptoms

Yellow rust takes its names from its appearance as it produces characteristic yellow to orange pustules (uredias), which consist of thousands of urediospores on the leaves of adult plants. Symptoms usually start in winter but often do not become clearly visible until the early spring. Symptoms are visible as symmetrical 'stripe' lesions along the leaf, in the epidermis. Pustules of yellow rust have some similarities to those of brown rust (Fig. 1-3), but a key difference is that whilst each individual brown rust pustule will be derived from a single initial spore/infection, it will be the entire 'stripe' that is derived from a single spore in the case of yellow rust. In resistant cultivars infection types can show no symptoms to small hypersensitive flecks, and the severity of necrotic lesions and chlorosis is dependent upon the susceptibility of the host plant. Symptoms usually occur after 1 week and sporulation takes up to 2 weeks (Chen, 2005).

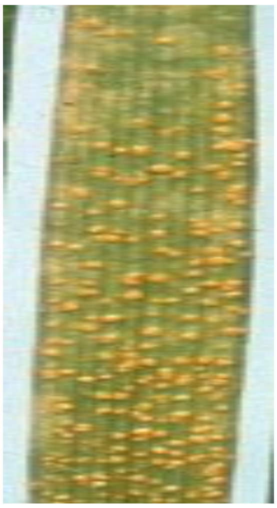
A.  **B.**  **C.** 

Figure 1-3 Yellow rust infections. **A.** Uredias on wheat seedling. The image demonstrates a seedling where the whole infection is derived from one spore (picture from M Dickinson) **B.** Characteristic stripes in an adult wheat plant where each individual stripe of infection on an adult plant is also derived from a single spore (picture from NIAB, 2012) **C.** A brown rust infection for comparative reasons, where each individual pustule has come from a single spore (picture from M Dickinson).

1.2.3 Infection process (asexual cycle)

As previously established, the primary disease-causing stage of the rustlife cycle is the asexual cycle through the production of asexual spores that are dispersed primarily through the air. Adhesion of germinated urediospores on the leaf surface initiates the infection process of yellow rust. They infect through the stomata using a germ tube, which penetrates through the stomatal cavity and different infection structures form, including germ tubes, appressoria, infection pegs, substomatal vesicles, infection hyphae, haustorial mother cells and finally, a specialised nutrient feeding structure known as haustoria. These specialised structures are formed as a result of specific genes being expressed in the first 60 hours of infection (Harder and Chong, 1984).

The common infection process involves the formation of these specialised structures that help entry of effectors into the plant. The infection process of yellow rust is initiated with germinated urediospores, which infect through the stomata using a germtube. This process leads to the formation of an appressorium above the stomatal opening and a penetration peg forms into the specialised nutrient feeding structures known as haustoria (Fig. 1-4).

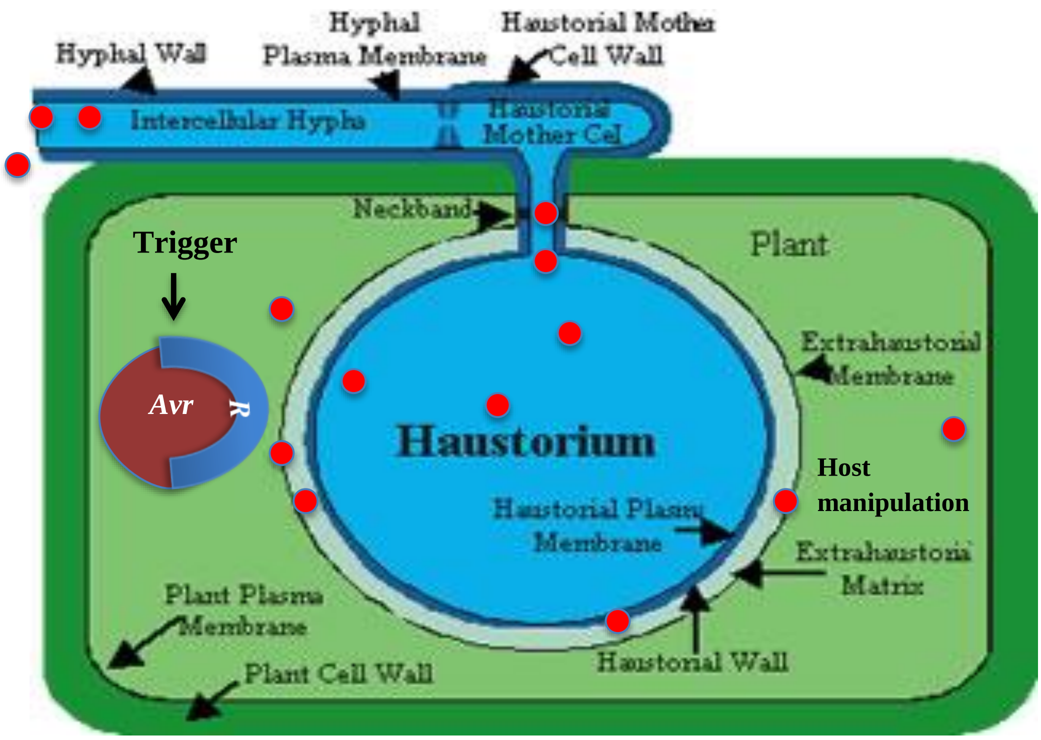


Figure 1-4 Schematic representation of the host-haustorium interface. The rust infection process is shown through cocktail of effectors (in red) modulating host defence and triggering host defence. The infection initiates as the haustorial mother cell (HMC) which differentiates from the intercellular hyphae. The pathogen penetrates the cell wall and invaginates the host plasma membrane, where the haustorium is developed. This allows the haustorium to remain separated from the host cell cytoplasm, surrounded by the extrahaustorial matrix and extrahaustorial membrane. During infection the pathogen effector proteins are secreted from the haustoria into the extrahaustorial matrix. Taken from Szabo and Bushnell (2001).

There is evidence suggesting that haustoria play an important role in nutrient acquisition and suppression of host defences (Voegele and Mendgen, 2003, Dodds *et al.*, 2009) (see Fig. 1-4). Haustoria are key players during infection as they secrete effectors into the host cytoplasm where they are thought to promote infection by manipulating the plant defence system (Garnica *et al.*, 2014, Petre and Kamoun, 2014). Haustoria are surrounded by the haustoria cell wall and then the extrahaustorial membrane surrounds the haustoria, which is structurally and biochemically different to the host plasma membrane. These structures along with the extrahaustorial matrix keep the haustoria separated from the host cell cytoplasm (Dodds *et al.*, 2009).

Haustoria aid infection as they invaginate the cell wall but remain extracellular, increasing the surface area of contact between the fungus and cell cytoplasm for infection (Garnica et al., 2013). The haustorium is located between the mesophyll cell wall and the cell membrane, allowing the fungus to deliver virulence proteins called effectors in to the host cytoplasm, subsequently suppressing host defences (Garnica et al., 2014). The haustorium is an example of rust fungi adaptation and this infection structure sustains the biotrophic association with hosts.

1.2.4 Disease development and epidemiology

Three different environmental factors affect development of yellow rust; temperature, wind and moisture. Temperature affects spore germination and infection, whilst wind plays an important role in spreading yellow rust. The wind can also dry urediospores, consequently reducing germination and infection. Spore germination is sensitive to light and moisture, and moisture is also important for infection and survival, thus irrigated crops are particularly sensitive to *Pst*. Although moisture in the environment promotes infection, it adversely affects spore viability; therefore, spores must be kept dry to maintain their viability especially during sample deliveries and storage.

Specific ranges of temperatures are imperative for the onset and development of yellow rust. Ideal growth conditions, where urediospores will germinate, include cool, humid nights, which are between temperatures of 10°C to 15°C (thus it is more of a spring disease) (Clark, 2015). Yellow rust is dormant below the threshold temperature 5°C as no sporulation occurs and it survives the winter as actively sporulating lesions or mycelium in over-wintering wheat plants. The time between infection and appearance of symptoms is known as the latent period, which can be up to 118 days in the winter (Zadoks, 1961); alternatively and more commonly, the latent period lasts up to 2 weeks. Daytime early spring temperatures (5°C and above) favour subsequent disease development, and it is important to note that summer temperatures of 20°C restrict spore germination (Sharma-Poudyal *et al.,* 2013). In the UK, the first foci of *Pst* infection usually appear in early spring resulting in epidemics until May/June. The complete lifecycle (initial infection to the production of urediospores) for yellow rust takes about two weeks.

1.2.5 Epidemics and Pathotypes

The most practical and economical means of control for yellow rust is the use of resistant cultivars. Since the introduction of resistant cultivars the frequency of epidemics has reduced globally. Baruss first introduced the idea of physiological races in 1911 (Barrus, 1911). It was suggested that different strains from a particular pathogen species have the ability to infect different varieties of the same host species, also known as 'differential hosts'. The different pathogenic strains (or pathotypes) are solely defined on their ability to cause disease on differential hosts based on differences in their resistances to the pathogen. These terms have been commonly used to describe fungal pathogens.

This pathogenic specificity on a restricted host range reflects on the deployment of resistance genes in host plant populations. Selective forces brought about by techniques such as selective breeding have led to changes in virulences. Breakdown of genetic resistance has increased complexity amongst pathogen populations, where new resistance genes can be impaired. The wider the geographical distribution of resistant cultivars, the greater the selection pressure for virulence will be. The likelihood that an epidemic will occur is therefore increased, as when avirulence is lost the new pathotypes will overcome the corresponding resistant cultivars. The ability of the pathogen to overcome resistance within that variety will be significantly increased, because the resistance gene product is no longer recognising the corresponding avirulence gene product (Avr) (Ellis *et al.,* 2014). A new resistance gene must therefore be deployed in the wheat population to avoid further yellow rust epidemics. Most resistance genes to date have singularly or variously combined been overcome.

The impact of yellow rust has been reported in different studies (Chen *et al.,* 2002, Duan *et al.,* 2010). Reduced yield of wheat grains is a result of plant pathogens and thus disease management procedures are required to alleviate pathogenic disease (Chen, 2005). Pathogenic spores spreading long distances by wind dispersal cause successful yellow rust epidemics, and overcoming host resistance has also been a large contributor to disease epidemics (Brown and Hovmøller, 2002). Epidemics of yellow rust affect wheat growing regions significantly and have been a major problem worldwide. For this reason there is a high dependence on effective resistant genes against yellow rust in the wheat population to prevent future epidemics in the UK and the rest of the world.

1.2.5.1 Epidemics in the UK

In the UK, yellow rust is the more important rust compared to brown rust (*Puccinia triticina*) and stem rust *(Puccinia graminis* f. sp. *tritici),* for the reason that brown rust occurs in the UK late in the season and therefore does not cause major problems on yield, and stem rust does not as yet infect cereal crops in the UK. Thus focus and efforts in the UK are largely towards the more destructive yellow rust.

The UK currently produces 2 million hectares of wheat annually and in 2010 it was reported that 30% of wheat was infected by yellow rust endemics causing crop yield to reduce significantly (Holdgate and Watts, 2015). Through research it has become evident that new races of *Pst* are evolving rapidly, overcoming host resistance and causing regional epidemics, which has reduced both yield and grain quality of UK wheat (Powell *et al.*, 2013). In 2011 for example, the 'Warrior' race was first detected to be virulent on adult plants and later in 2012, it had prevailed in 75% of the isolates found in the UK, illustrating how this pathogen has the tendency to become a major production constraint in the UK (Holdgate, 2016). ‘Warrior 1’ was the dominant race in 2011 but has since declined with ‘Warrior 3’ prevailing (Holdgate and Watts, 2015).

A recent study by Hubbard *et al.* (2016) described the dramatic shift identified in the *Pst* population in the UK. Through the Illumina whole-genome shotgun approach their findings displayed a clear distinction between the historical isolates, which clustered together in a single clade, and the newer isolates collected in 2013, which distantly related to these historical isolates. Most of the isolates from 2013 shared genetic similarities to the 'Warrior' pathotype that has been discussed above. This analysis further confirms the findings of Ali *et al.* (2014) who have also identified a new diverse sexually produced *Pst* population replacing an older population, and this has been discussed more thoroughly below (see section 1.2.5.2).

Furthermore, Hubbard *et al.* (2015) found that samples collected from host triticale clustered together, potentially signifying host specificity in the UK *Pst* population collected in 2013. A clear distinction between the old and new UK *Pst* isolates was determined through analysing an association between both genotypic and phenotypic diversity. Disease severity was recorded 16 to 20 days post inoculation and for each group distinct phenotypic characteristics were displayed between populations that shared similar genetic relatedness.

Yellow rust surveys have been carried out in the UK for many years in order to check epidemics across different sites in the country and Stubbs (1985; 1988) highlighted the importance of such distribution surveys. Data gathered from these sites enables population surveys to monitor epidemic risks through migration, changes to pathotype frequency and detection of new virulences. In the UK, the Cereal Pathogen Virulence Survey (UKCPVS) is conducted by NIAB and the CropMonitor team which is co-ordinated and led by the Food Environment Research Agency (FERA), to provide accurate and consistent feedback on surveys, a vital prerequisite for breeding programmes. Surveillance and information exchange between different regions of a country but also between countries is important to reduce the threat of this disease.

1.2.5.2 Epidemics worldwide

Wheat accounts for the majority of grain production in Australia, making it Australia's most important crop. About 22 million hectares are planted annually across Australia (Aegic, 2015). Yellow rust is common in Eastern and Southern parts of Australia, but recently there have been concerns with rust becoming more prominent in Western Australia (Australian Cereal Rust Control Program, 2015). Prior to 1979, the continent was devoid of yellow rust but introductions of this rust and changes in virulences have ensured close monitoring and reporting of this epidemiological area (Park and Wellings, 1992).

Yellow rust was first introduced into Victoria, Australia in 1979 spreading across Eastern Australia rapidly. The initial pathotype was 104E137, which had already been widespread in Europe, suggesting it was spread from Europe through clothing via plane travel (Wellings and McIntosh, 1990). This strain subsequently mutated and developed into different pathotypes of yellow rust, which spread in the east, but fortunately did not move towards Western Australia (Economic Development, 2015). At this point yellow rust had spread to New Zealand where the resistance gene *Yr9* was overcome in 1991 by pathotype 234E139A-; *Yr9* was still effective in Australia (Fig. 1-5). Figure 1-5 shows stepwise mutations for virulence in Australia from the first introduction of yellow rust in 1979, causing endemics despite being a pathogen that proliferates in cooler temperatures. Seventeen different pathotypes developed over the first 8 years that the rust was present in the country, indicating the rapidity with which the pathogen can evolve. From a period of 1979 to 1988, yellow rust became virulent against 10 specific wheat resistance genes (*Yr7*, *Yr6*, *Yr5*, *Yr8*, *YrA*, *Yr1*, *Yr9*, *YrSK*, *Yr2*, and *YrSpP*).

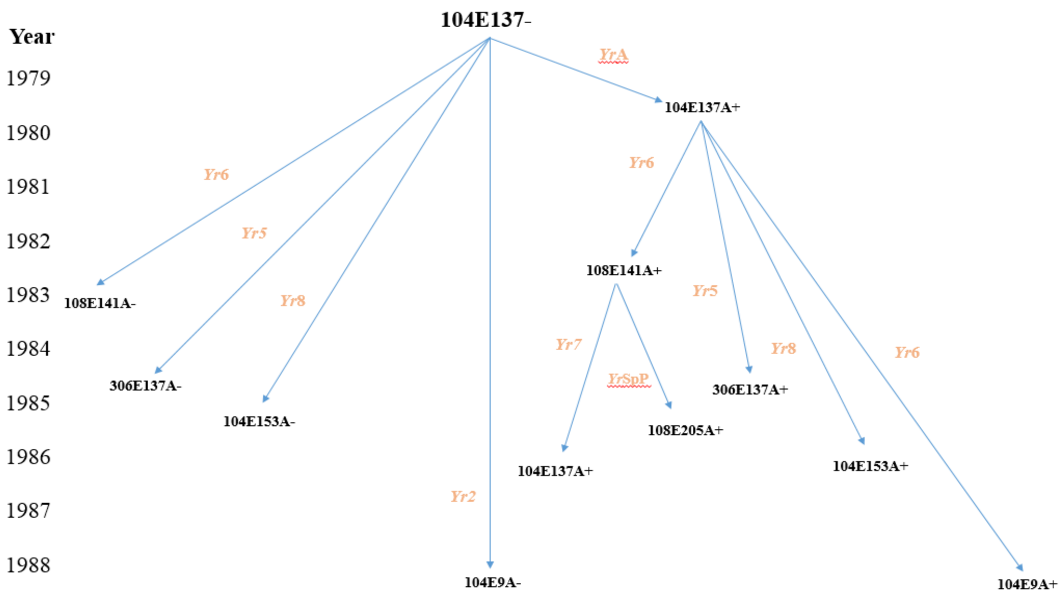


Figure 1-5 Stepwise mutation in Australia (Wellings and McIntosh, 1990).

The second introduction of yellow rust in Australia occurred in 2002 in Western Australia. By 2003 this pathotype commonly known as 'WA', became the prominent strain in Eastern Australia. 'WA' has since undergone many mutations, resulting in different varieties of this rust infecting wheat in Victoria (Wellings *et al.*, 2003).

It has been proposed that stepwise mutations from avirulence to virulence produced the new pathotypes from the original isolate (Wellings, 2007). New strains evolve by single gene mutations and significant population growth sizes contributed to large epidemics within 2 to 3 years (Wellings, 2007). Although a large number of mutant derivatives have been detected in Australia over the last 25 years, only a few specific pathotypes have been dominating within that time period due to their specific adaption to the grown wheat varieties (Wellings, 2007, Chen *et al.*, 2014). The observations from this unique study in Australasia give weight to the notion that existing pathotypes increase strains through mutation of avirulence genes. Therefore large-scale dependence on effective resistance genes leads to greater potential for damaging epidemics.

In China, winter wheat production was approximately 22.4 million hectares in 2015, again highlighting the worth of this crop as one of the world's most important food commodities. Pathogenicity surveys have been conducted since 1940 (McIntosh *et al.*, 1995) due to severe disease outbreaks, with crop losses of up to 100%, though more commonly in the range of 10-70% (Chen, 2005). Similar to the situation in Australia, the breakdown of a single resistant gene has been shown to lead to epidemics, as occurred in 1990, which led to damage in 6.67 million hectares in North China (Ji *et al.*, 1996). Wheat production in China has been heavily impacted by epidemics that occurred in 1950, 1964, and 2002 reducing yield dramatically by 1 to 6 million tonnes (Chen *et al.*, 2009, Zheng *et al.*, 2013). Currently 53 *Yr* genes confer adult plant resistance (APR), high temperature-adult plants (HTAP) resistance, and race specific all-stage resistance (ASR) (McIntosh *et al.*, 2007). Pathotype CY32 was detected in 1994 and pathotype CY33 was detected in 1997; both pathotypes have become dominant in recent years posing a threat to wheat cultivars in epidemic regions of China (Chen, 2005).

Because of the speed at which yellow rust can evolve, it is important to use a variety of resistance mechanisms including major-gene resistance, polygenic resistance, high-temperature resistance, and adult-plant resistance, as mentioned in the case of China, to stabilize the pathogen population and to reduce epidemics (Mundt and Browning, 1985). However the struggle remains as introductions of new pathotypes eludes resistance in cultivars. Careful deployment of extensive resistance genes is a vital pre-requisite to control this disease, one that requires detailed knowledge of which resistance genes are already present in wheat cultivars. The major challenge is in dealing with the pathogen's ability to mutate to new races, which may infect previously resistant varieties. Rapid development of new *Pst* races and the resistance genes that become ineffective to predominant races is a global issue that requires a more strategic use of resistance genes and faster screening of field isolates.

In order to mitigate threats and environmental implications of invasive pathogens such as yellow rust, it is important to understand the origin and migration of rust pathotypes. Losses are particularly high in cases when the epidemic is associated with an exotic strain that has not been considered in breeding for resistance locally (Walter *et al.,* 2016). Recently, the evolution of yellow rust pathotypes has been documented by (Hovmøller et al., 2015); this study details the migration of rust pathotypes worldwide. They illustrated the emergence of new exotic races displaying a significantly higher genetic diversity than a previous clonal population. Since 2011, Warrior and Kranich races have largely replaced a pre-existing European population. A third, Triticale aggressive race, from 2006 onwards was also detected in the post-2011 population. Due to the significantly higher genetic diversity identified between the pre-2011 isolates and the new exotic races, it dismisses the notion that mutations or recombinations are a possible source of origin in the post-2011 European population (Ali *et al.,* 2014). The Himalayan and near-Himalayan regions have recently been documented as the likely centre of diversity, as this is the place where most sexual reproduction appears to occur (Ali *et al.*, 2010, Ali *et al.*, 2014).

This concept is accentuated by Hovmoller *et al.* (2015) who identified alleles in the new exotic strains that displayed relatedness to isolates sampled in the centre of diversity i.e. Himalayan regions, but were absent in the clonal pre-2011 population. This is further supported by the findings of Hubbard *et al*. (2015) who also found the Warrior race significantly diverse to previous European populations. Therefore, studies have reached a consensus where the origin of post-2011 *Pst* pathotypes in European populations has been confirmed in the Himalayan regions. The evidence is also that these new introductions then become clonal populations that evolve by stepwise mutation away from the centre of origin (Wellings, 2007, Ali *et al.*, 2014, Hubbard *et al.*, 2015, Hovmøller *et al.*, 2015).

1.2.6 Economic Implications

As yellow rust continues to affect regions where wheat cultivation is extensive the social and economic implications are also devastating. Damage caused by yellow rust is of such magnitude that it poses a threat to rural livelihoods. There is an increased expenditure on the use of fungicides, consequently resulting in environmental damage, but also a loss of confidence in plant breeding companies as epidemics overcome new varieties (FAO, 2006). Economic significance of yellow rust follows from the extent to which it may reduce grain yield and stability, its ability to spread rapidly and reach epidemic proportions under favourable conditions, and the pathogens' potential to mutate rapidly to overcome the effects of current resistance genes.

Aggressiveness of this disease increases food prices and is a threat to food security in areas where wheat is the main source of their calorie intake. In the developing world, this can be as high as 40% (ICARDA, 2011). More recently, emerging races are showing stability in extreme temperatures, highlighting their ability to adapt to changes in temperature, a phenomenon not extensively seen previously (Bryant *et al.*, 2014). Yield reductions increase use and expenditure of fungicides causing environmental problems, but more recently changes in regulations has minimised chemical control options. Chemical control regulations have meant efforts are moving increasingly towards further use of gene resistance.

The global extent of yellow rust allows for researchers to investigate the economic cost of this disease. Occurrence and economic consequences of stripe rust have varied of the years globally, for example between 1961 and 1984 average annual losses were 0.88 million metric tons, resulting in annual losses of $157 million US dollars per year. With the use of fungicides and resistant cultivars between 1985-1999 (Chen, 2005), this low loss middle period signifies the large difference in the losses attributable to this disease, with only 0.34 million metric tons lost and costing an improved $61 million US dollars per year. This however drastically increased post 2000 where there was a measured change due to the emergence of new stripe rust pathotypes. Large differences in the losses attributable to this disease are evident with annual losses of 4.74 million metric tons, and $848 million US dollars per year. (Chai *et al.*, 2014).

1.2.7 What are effector proteins?

Fungal avirulence genes encode virulence factors called effector proteins, which are secreted into host plants and modulate host immunity to promote infection (Petre and Kamoun, 2014). Effectors have a broad functional spectrum; for example, necrotrophic pathogens release toxins, which cause plant cell death, and biotrophic pathogens (such as *Pst)* release avirulence gene products, which the plant immune system has evolved to recognise. Research from transcriptome studies of isolated haustoria of different rusts such as the flax rust pathogen *M. lini* (Kemen *et al.,* 2005), the common bean rust *U. fabae* (Link *et al.,* 2014), the stem rust pathogen *P. graminis* f. sp. *tritici (Pgt)* (Duplessis *et al.,* 2011), and the yellow rust pathogen *(Pst)* (Cantu *et al.,* 2011, Garnica *et al.,* 2013), have predicted a large number of candidate effector proteins. A subset of these effectors constitutes the avirulence gene products that are recognised by the *R* gene products. Fungal effector proteins can be categorised into two distinct secretion systems based on where they reside. If the proteins reside in the extracellular space (apoplast) they are known as apoplastic effectors, and if the effector proteins translocate in to the host cytoplasm they are recognised as cytoplasmic effectors (Stotz *et al.,* 2014).

Haustorial formation induces effector proteins to initiate the infection process. Virulent effectors are initially secreted into the apoplast and into the extrahaustorial matrix. They must pass the extrahaustorial membrane before entering the host cytoplasm where the effectors can be recognised by their corresponding *R* gene products to trigger an immune response. It is evident that the interface between the host and the pathogen is fundamental for such trafficking mechanisms and their role in virulence.

1.2.8 Characteristics identified in effector proteins

Until now effector proteins of three different rust species have been identified. Rust of flax caused by *M. lini* has been a major focus within such studies and to date, the majority of the *M. lini Avr* genes that have been identified encode secreted proteins, which are expressed in haustoria and are recognised by the corresponding flax *R* genes (more specifically, the NB-LRR receptors), in accordance to Flor's gene-for-gene theory (Dodds *et al.,* 2004, Catanzariti *et al.,* 2006, Stergiopoulos and de Wit, 2009). *AvrM*, *AvrL567*, *AvrP123* and *AvrP4,* at four different loci, have been cloned (Catanzariti et al., 2006, Lawrence *et al.,* 2007). Catanzariti *et al.* (2006) established 21 haustorially expressed secreted proteins (HESPs), four of which corresponded to the genes mentioned above. The avirulent properties (i.e. their ability to trigger a specific immune response) of these proteins were confirmed, as expression of these *Avr* genes in flax induced a HR (Catanzariti *et al.,* 2006). Induction of a HR also suggests that effector delivery in the cytoplasm is likely, implying that the effectors are delivered specifically across the extrahaustorial membrane and into the host cell cytoplasm during infection. However the mechanism of trafficking of such proteins and their role in virulence is still unclear.

Similarly, the Rust Transferred Protein 1 (RTP1) from *Uromyces fabae* was one of the first proteins to be detected in the pathogen-host interface (the extrahaustorial matrix (EHM)) but was also identified in the infected host cells. RTP1 was specifically expressed in the haustorium, to be then transferred into the host cytoplasm (Garnica *et al.,* 2014). RTP1 also shows similar traits to cysteine protease inhibitors, so therefore may inhibit proteases that play a role in host defence mechanisms (Pretsch *et al.,* 2013). Moreover, RTP1 also shows a structural role as it forms filamentous-like structures inside the EHM and the host cytoplasm, stabilising the host cell to allow haustorium positioning (Kemen *et al.*, 2013).

In wheat stem rust, (Upadhyaya *et al.*, 2014) predicted 1,294 secreted proteins, out of which 520 were classified as HESPs from an Australian *Pgt* isolate (pathotype 21-0 from the year 1954). PGTAUPSE-10-1, the *Pgt* protein, commonly known as the candidate effector AvrRs22, causes cell death in the host that carries the resistance gene *Rs22* (Upadhyaya *et al.*, 2014). This effector protein is also thought to translocate from haustoria into host cells. Two more *Pgt* isolates (field derivatives from 1982 and 1984) were used for comparative studies. Genomic comparative studies of these pathotypes illustrated that 25 HESPs showed mutations which were not apparent in the 520 HESPs, signifying that virulence had developed against four resistance genes (*Sr5, Sr11, Sr27, SrSatu*), increasing scope for new virulent varieties in the field.

Table 1-4 Effector proteins that have been identified in three different rust species till date. The listed effector proteins are secreted in to the host cytoplasm.

Species	Publication	Effector proteins
Melampsora lini	Ellis *et al.*, 2007	AvrM, AvrL567, AvrP123, AvrP4
Uromyces fabae	Kemen *et al.*, 2005	RTP1
Puccinia graminis* f. sp. *tritici	Upadhyaya *et al.*, 2014	PGTAUPSE-10-1

Trafficking mechanisms of effector proteins are better understood than their role in virulence, but have been disputed. As established above, it is noticeable that there is a large amount of evidence that haustoria are a common infection feature. Most pathogens have evolved their trafficking mechanisms over time, since structures and sequences amongst pathogens are so diverse. Despite their diverse mechanisms, they share a common feature, the N-terminal domain that is proposed to mediate entry into the host cell. These domains are located after a general secretory signal peptide (Petre and Kamoun, 2014). For example *AvrL567* and *AvrM* of the flax rust fungus *M. lini* require an N-terminal domain commonly known as RXLR. Likewise, approximately 87% of *Phytophthora* species effectors also carry a RXLR motif, thus defining the RXLR effector family (Petre and Kamoun, 2014). It has been suggested that RXLR motifs within *Phytophthora* are highly conserved and are evolving over time. In contrast, some effectors of haustorium-forming fungal pathogens lack an N-terminal signal, for example *Avrk1 and Avra10* of *Blumeria graminis* f. sp. *hordei* that causes powdery mildew on barley, suggesting an alternate and unidentified route into the host cell (Stergiopoulos and de Wit, 2009).

Given the lack of sequence similarities and conserved sequence motifs, it is difficult to identify or characterise specific effector proteins from the list of candidates, though some share structural similarities. To identify a candidate secreted effector protein, certain criteria are usually met. Certain criteria such as, effectors are small in size, are rich in cysteine residues, and possess a secretory signal peptide for secretion, as well as the absence of a transmembrane domain (Petre and Kamoun, 2014, Sperschneider *et al.,* 2015). Effectors must be expressed during infection and in specialised structures like haustoria.

Explicitly, these features assist in effector screening in apopolastic effectors that are thought to be rich in cysteine residues when compared to candidates of cytoplasmic effectors (Sperschneider *et al.*, 2015). However, it is important to note that although the majority of effectors fit this standard, some effectors, for example, AvrLm1 from the hemibiotrophic pathogen *Leptosphaeria maculans*, are delivered into the apoplast, and only have one cysteine (Gout *et al.*, 2006). This further explains that screening for effectors does not discriminate between the different classes of effectors, thus a uniform and systematic approach is required.

Sequence homology between known effectors has helped in identifying some novel effectors, though significant sequence similarity is not common. This suggests that many effectors evolve rapidly, and are adapted for host specialisation. For example, *Cladosporium fulvum*, tomato leaf mould, has had four *Avr* genes cloned (*Avr2, Avr4, Avr4E*, *Avr9)* that are recognised by the corresponding Cf proteins (Cf-2, Cf-4, Cf-4E, Cf-9) in tomato (Thomma *et al.*, 2005, de Wit *et al.*, 2009). Novel *C. fulvum* effector genes encoding for Ecp6 and Ecp7 (extracellular proteins that invoke an HR response in the tomato host plant) have also been cloned. Ecps are secreted by different pathotypes of *C. fulvum* during infection and Ecp6 has shown significant sequence similarity to the rice blast *Magnaporthe oryzae Slp1* gene (Bolton *et al.*, 2008). Although no clear orthologues in other organisms have been identified, the presence of LysM domains in these proteins (which are recognised as carbohydrate binding modules), suggest that the LysM domain of *C. fulvum* Ecp6 may be involved in chitin binding, similarly to the Avr4 effector (Bolton *et al.*, 2008, de Wit *et al.*, 2009).

Fungal effectors harbouring a protease inhibitor activity are also common, as seen with RTP1 (Guyon *et al.,* 2014) . Effector genes are also known to often be located in unstable regions within the genome, such as repeat-rich regions (Saunders *et al.*, 2012). Prediction of biochemical functions of secreted proteins and analyses of conserved amino acid motifs has also been insightful in suggesting their role in pathogenicity.

Many candidate effector proteins have been identified using the general, basic but essential criteria described above (Cantu *et al.*, 2011, Cantu *et al.*, 2013, Saunders *et al.*, 2012, Garnica *et al.*, 2013, Zheng *et al.*, 2013). However, a more sophisticated and uniform approach would need to be introduced for systematically applying a unified pipeline for future comparative studies in other fungi (Fig.1-6).

Table 1-5 Illustration of transcriptome studies revealing potential effectors. Adapted from Petre *et al.*, 2014.

Rust Specie	Expression	Effector Protein	Predicted CSEPs	Transcriptome approach	Publication	No. of predicted secreted proteins
Melampsora lini	Haustorium	AvrM, AvrL567, AvrP123, AvrP4			Ellis *et al.*, 2007	1549
Uromyces fabae	Haustorium	RTP1		Illumina	Kemen *et al.*, 2005	
Puccinia graminis* f. sp. *tritici	Haustorium	PGTAUPS E-10-1	1106 candidate secreted expressed proteins (CSEPs), out of which 442 expressed in planta and 29 up regulated genes	Custom whole genome oligoarrays	Dupleiss *et al.*, 2011	1852
Puccinia striiformis* f. sp. *tritici	Haustorium	PGTAUPS E-10-1	933 CSEPs, out of which 57 were inducded in haustoria and 31 in planta	Illumina Genome Analayzer II	Cantu *et al.* 2011	1188

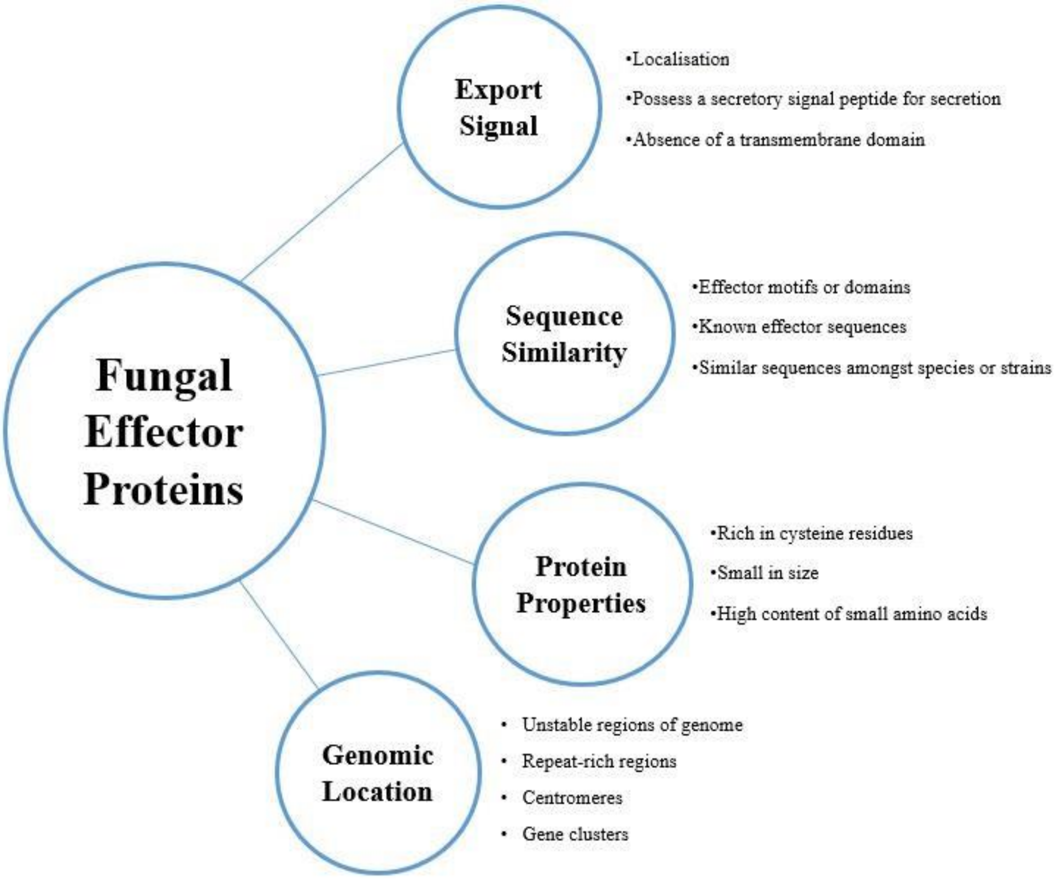


Figure 1-6 Schematic diagram illustrating the different pipelines used to identify effectors in different studies.

Pipelines for predicting effector proteins have been employed in different studies, which don't solely rely on the 'standard rules' listed above, i.e. small size and cysteine rich. For example, Saunders *et al.* (2012) developed a bioinformatics pipeline to identify a candidate effector protein. The criteria for identifying effectors listed in Figure 1-6 are not always met; some effectors may consist of properties that are absent in others.

Saunders *et al.* (2012) further confirmed identification of effectors through this pipeline. Using this approach, a total of 1549 and 1852 secreted proteins were predicted from *M. larici-populina* (*M. lini*) and *P. graminis* f. sp. *tritici* (*Pgt*) proteomes, respectively. However, although this pipeline proved to be successful in identifying high priority candidates, it failed to identify the PGTAUPSE-10-1 effector, which was instead found by Sperschneider *et al.* (2014), who identified 42 haustorially delivered effector proteins, including PGTAUPSE-10-1, which had previously been discovered (Upadhyaya *et al.*, 2014). Saunders *et al.* (2012) developed a sophisticated pipeline for haustorially secreted effector proteins according to criteria of verified effectors such as PGTAUPSE-10-1 and *M.lini* effectors. This pipeline failed to recognise a strong candidate effector such as PGTAUPSE-10-1, which was recovered as a top effector candidate by the Sperschneider *et al.* (2014) pipeline.

Sperschneider *et al.* (2014) established a pipeline that avoided existing effector prediction pipelines that focused on small-secreted, cysteine rich proteins, producing large numbers of potential effectors all of which may not be true effectors. Instead, the pipeline was based on multiple lines of evidence such as pathogen genes highly expressed during infection, rapidly evolving, and pathogen-specific. Incorporation of comparative analysis and evolutionary data through selection analysis was important for lines of evidence that looked at up regulation during *in planta* infection and diversifying selection.

Hierarchal clustering was employed to rank effectors:

1. Identified secreted proteins from the predicted proteomes (secretome prediction)

2. Secreted proteins and non-secreted proteins (to detect false positives) were put into tribes dependent on their sequence similarity scores to secreted proteins of two examined fungal species (Markov Clustering)

3. Explored sequence similarities (functional annotation)

4. Searched for conserved motifs (*De novo* motif search)

5. Ranked proteins to known effector properties (Hierachal tribe clustering)

Despite the different pipelines used for identifying effectors, thus far the only universal features for effector proteins are that they are secreted during infection *in planta*. How they are all secreted is still unknown, for example effectors that lack a signal peptide must have an unconventional mode of trafficking. Current knowledge on effector characterisation is still incomplete and there is a need to improve our understanding on their function in plant cells. It is important to look in to factors that can achieve a more uniform pipeline to screen for effectors.

1.2.9 Effector proteins of yellow rust

Transcriptomics promises identification of a larger number of proteins as more rust fungal genomes are becoming available. The genomes of four rust species have been published so far (Table 1-5).

Recently, availability of draft genomes such as PST-130 (origin, USA) through tools such as next generation sequencing, and sophisticated computational methods have made it possible to characterise effector complements from *Pgt* and *M. lini,* to identify and characterise putative effectors of *Pst.* For instance, the *Pgt* genome has been used as a reference for comparative studies to expand knowledge on the *Pst* effector catalogue (Cantu *et al.,* 2011).

It is evident that genomic analysis of these four species has revealed a large number of candidate secreted effector proteins through the pipelines discussed above. Currently, no yellow rust effector proteins have been validated; two different reports have predicted sets of candidate proteins that have been published. Cantu *et al.* (2011) annotated 22,185 putative coding sequences and predicted 1,188 genes coding predicted secreted proteins. From 1,188 potential secreted proteins, 1,088 putative *Pst* effectors from isolate PST-130 were selected. These proteins carry a conserved N-terminal RXLR motif. Likewise, Cantu *et al.* (2013) re-sequenced the isolate PST-130, along with four more isolates. Two UK isolates included PST-87/7 and PST-08/21 and two other isolates from the US included PST-21 and PST-43. Re-sequencing PST-130 along with these newer isolates led to a much larger number of 2,999 predicted candidate secreted effector proteins.

Studies such as these have improved our knowledge of the effector repertoire of *Pst,* and are slowly overcoming the challenge of identifying effector proteins. These two studies have allowed easier mining of effector proteins between closely related isolates, and subsequently relate these effectors back to their defined virulence profiles.

A bioinformatics pipeline (adapted from Saunders *et al.*, 2012) was used to identify heterokaryotic and homokaryotic SNP frequencies between the isolates. Non synonymous polymorphisms were identified between the two UK isolates; it is important to note that the two isolates differed in their virulence to wheat varieties Robigus (*YrRob*) and Solstice (*YrSol*) (both Avr genes are present in PST08/21 and absent in PST 87/7) but shared virulence to ten *Yr* genes. Of these 117 proteins that were polymorphic between the two UK isolates, 51% proteins were expressed in haustoria and 62% were expressed during infection and 6% were enriched haustoria. In accordance to the pipeline discussed above to characterise candidate secreted effector proteins, 15% were found to be cysteine rich and 15% also carried a conserved N-terminal motif, suggesting they are potential candidates.

Cantu *et al.* (2011; 2013) developed a ranking system in which the top 100 ranked proteins were positioned into tribes containing secretory candidate effectors. Fourteen secreted proteins with non-synonymous substitutions between the two UK isolates were thought to be ranking in the top 100. Furthermore, 5 proteins were found in the top 10% expressed in the haustoria, showing distinct single amino acid substitutions amongst the two UK isolates. It is evident that genomic comparative studies on the two UK isolates have been informative in increasing knowledge and detection of virulent effector proteins, which are specifically differentially recognised by *YrRob* and *YrSol.* Analysing polymorphic variants that show high expression in the haustoria, significantly reduced numbers of potential effectors during effector mining.

More recently however, Petre *et al.* (2015) have looked at processing bodies (P-bodies) to prioritise candidate effectors (P-bodies are plant and not pathogen proteins). P-bodies are cytoplasmic domains that contain proteins and are thought to have a role in mRNA decay in eukaryotic cells (Kulkarni *et al.*, 2010). In plants, P-bodies play roles in tolerance, salt-stress, and immune responses (Steffens *et al.*, 2015). Sixteen effector proteins of yellow rust were analysed using a new pipeline that involved the species *Nicotiana benthamiana* (Solanaceae), an experimental system used to study proteins (Bombarely *et al.*, 2012, Pais *et al.*, 2013).

The model plant *N. benthamiana* system involves selecting effectors *in silica* through protein-protein interactions. Nine candidate effectors targeted a region within the plant subcellular compartment and one candidate effector Pst02549 was notably found in P-bodies, suggesting that P-bodies are new cell compartments where effector proteins of yellow rust can be found, an association not made before this study (Petre *et al.* (2015). Accumulation in P-bodies may result in suppression of the host immune response and manipulate RNA metabolism. It is therefore clear that *N. benthamiana* is a novel system that can be used for further *in planta* studies.

1.3 Aims and objectives

The aim of this project has been to:

- ✓ Identify molecular markers linked to specific pathotypes of yellow rust through a random molecular marker approach such as Amplified Fragment Length Polymorphisms (AFLP).
- ✓ Identify molecular markers such as Single Nucleotide Polymorphisms (SNPs) in potential effector genes, using a targeted approach (Illumina Miseq).
- ✓ Explore the association of both types of markers with virulence/avirulence phenotypes of isolates, with the aim of developing markers that could be used to test collections of spore samples using rapid molecular techniques to identify their pathotypes.

Chapter 2

Amplified Fragment Length Polymorphisms (AFLP)

Chapter 2: Amplified Fragment Length Polymorphisms (AFLP)

2.1 Introduction

2.1.1 Detecting genetic variation with molecular markers

Understanding the genetic variability of plant pathogen populations is an important component of successful disease control. Before molecular techniques were introduced in the mid-1980s (Edal, 1980), pathologists relied on virulence and phenotypic differences to study plant pathogen populations. Conventional methods of detection are unreliable, slow and inefficient, thus research in plant pathology has become increasingly reliant on Polymerase Chain Reaction (PCR) based technologies. Appropriate identification of reputable molecular tools has since become essential for successful identification of genetic variation within fungal pathogenic populations.

PCR based molecular methods exploit variation of the DNA, allowing identification of strains and clonal lineages within a species (Arif *et al.*, 2010). Such methods allow for easy, reliable and less laborious means of addressing diverse nuclear gene sequences. PCR enables targeted amplification of specific regions within the genome from a very small amount of DNA, thereby circumventing the need for molecular cloning.

Genetic differences deriving from polymorphisms have proven to be very important in the study of obligate biotrophs. Polymorphism is a sequence variant present within a population emitting genetic variation. Most genes have low mutation rates, such that a change in sequences is not recognised unless they have severe effects (Strachan *et al.*, 2014). The most common type of mutational change is known as Single Nucleotide Polymorphism (SNP). SNPs, a variation in a single position in the genome involve a nucleotide change in which one purine base changes to another, for example, adenine to guanine (or vice versa); this is known as a transition. A single nucleotide change from a purine to pyrimidine base, for example guanine to cytosine (or vice versa) is called a transversion (Yang and Yoder, 1999).

SNPs have also been well studied in host plants as assessing genetic variation is understood to reveal the cause of susceptibility to disease. This has transformed the practice of species identification techniques; however, this chapter focuses on development of markers within the yellow rust pathogen. SNP analysis in pathogens has been well studied in yellow rust isolates (Cantu *et al.*, 2013; Hubbard *et al.*, 2015). These studies as well as SNP analysis obtained from sequencing data within this present study will be further discussed in Chapter 3.

2.1.2 Molecular Markers

Molecular markers have been extensively employed in yellow rust research to indicate specific genetic differences between isolates and for molecular analysis of plant microbe interactions (Akfirat *et al.*, 2010, Hubbard *et al.*, 2015). Molecular markers are used to follow a specific trait and should be located close enough to the gene of interest for both to be inherited together.

PCR based molecular markers have been widely applied in this area of research in order to characterise genetic diversity and discriminate against yellow rust pathogen populations. Molecular markers such as random amplified polymorphic DNA (RAPD) (Chen and Line, 1992), restriction fragment length polymorphism (RFLP) (Chen *et al.*, 1993), simple sequence repeats (SSR) (Bahri *et al.*, 2011), and Amplified Fragment Length Polymorphism (AFLP) (Hovmøller *et al.*, 2002, Hovmøller *et al.*, 2008) have been used successfully in previous work (Table 2-1).

Table 2-1 A comparison of the most widely used molecular markers in plant science. Adapted from (Aminha, 1998).

Molecular Markers				
Feature	**AFLP**	**SSR**	**RAPD**	**RFLP**
DNA quantity required	High	High	Little	Moderate
DNA quality	Moderate	Moderate	High	High
PCR based	Yes	Yes	Yes	No
Ease of use	Difficult	Moderate	Moderate	Difficult
Reproducibility	High	High	Moderate	High
Dominance	Dominant	Co-dominant	Dominant	Co-dominant
Number of markers scorable	Unlimited	Unlimited	Unlimited	Unlimited

2.1.2.1 Simple Satellite Repeat (SSRs)

Simple sequence repeat (SSR) markers, also known as short tandem repeats (STRs) and microsatellites, are small sequences, normally of 1 to 6 bp present adjacently in multiple copies within a DNA sequence. Microsatellites are tandemly repeated mono-, di-, tri-, and tetra- nucleotides (for example, (AT)n, (AAT)n, (GATA)n) found in coding and non-coding regions, and the number of these repeats is variable, though they are randomly and evenly dispersed in the genome (Collard *et al.*, 2005, Ijaz, 2011). SSR markers are highly polymorphic, as the number of repeats within a sample can vary tremendously, allowing detection of a range of alleles for that marker.

Conserved DNA sequences flanking SSRs are used to design suitable primer pairs that can amplify the microsatellite loci during PCR. This process produces PCR products that can be visualised on an agarose gel to identify SSR polymorphisms that will be in the form of differences in the length of the amplified product; each length will be indicative of an allele at that locus. SSR's are co-dominant markers where both alleles are expressed, distinguishing heterozygotes from homozygotes, thus establishing a rich source of allelic diversity by determining allele frequencies expressed at a particular loci (Kumar *et al.*, 2009).

SSRs have attractive features that make them suitable as markers in plants and pathogens, one being their highly polymorphic nature, which potentially allows detection of many different alleles for each marker. Additionally, due to the advancements made by PCR technologies, amplifying microsatellite repeats has become a simple process; however, SSR development requires prior sequence knowledge, thus difficulties lie in designing the SSR primers (as origin of variation is unknown), which will amplify the specific SSR repeat in a PCR reaction. As with AFLPs, SSR reproducibility is robust and also requires small amounts of template DNA (Morgante and Olivieri, 1993) .

As it has been discovered that SSR markers are highly abundant within pathogen species, they are therefore the preferred molecular marker used in pathogen genetic studies, alongside AFLPs (Bahri *et al.*, 2011, Cheng, 2012) . Therefore, microsatellites have been extensively used in characterising rust pathogen populations. For example, in crown rust (*Puccinia coronata* f.sp. *lolii*) of ryegrass, 55 SSR loci were identified with consistent tri-nucleotide repeats. From the 55 identified markers, 12 SSR markers were valuable for determining genetic diversity within and between the 16 isolates originating from the UK, Australia, and New Zealand (Dracatos *et al.*, 2006). Similarly, in brown rust (*Puccinia triticina*) a study using SSRs was used to show that *P.triticina* populations in North America could be placed into 4 groups. Pathotypes in a particular group shared identical genetic information, including the same virulence profiles (Ordoñez and Kolmer, 2009).

2.1.2.2 Amplified Fragment Length Polymorphism (AFLP)

The Amplified Fragment Length Polymorphism (AFLP) system provides a reliable method to visualise hundreds of amplified DNA restriction fragments concurrently (Paun and Schönswetter, 2012). AFLPs were first introduced by Vos *et al.* (1995); they are dominant markers that successfully allow simultaneous screening throughout the genome, and are preferably used as a fingerprinting technique, helpful in monitoring the identity of an isolate or analysing the degree of similarity among isolates. They are multilocus systems that allow primers to bind to numerous locations in the genome, thus allowing the amplification of multi loci at any one time. The ability of AFLPs to provide a powerful system stems from the fact that they can amplify a large number of bands per gel amplification (Vuylsteke *et al.*, 2007). Moreover, generation of a large number of markers without the requirement for prior sequence information demonstrates that this marker systems high potential. In contrast with other molecular markers, AFLP requires small amounts of template DNA to produce a large number of reproducible markers distributed in the genome.

Since Vos *et al.* (2005), AFLP protocols have been modified over the years, but preparation of samples to be genotyped for AFLP markers typically involves five main stages: 1. Digestion of purified genomic DNA with two restriction enzymes 2. Ligation of adaptors with sticky ends 3. Pre-selective PCR amplification of the restricted fragments with non-selective primers 4. Selective PCR amplification with selective primers 5. Electrophoretic separation of amplified DNA fragments. Once these stages are complete, a final stage of scoring and data interpretation is required. There are two types of polymorphism detected by AFLP markers; point mutations (a random change in genetic sequence along a specific point in the DNA strand), and insertions/deletions between them.

During the DNA digestion stage two restriction enzymes are generally used, *Mse*I (a four-base restriction cutter) and *Eco*RI (a six-base cutter). Small fragments of DNA are produced with “sticky ends”. *Mse*I, also known as a frequent cutter, plays a role in producing small DNA fragments that can be easily amplified and visualised on a gel, whereas *Eco*RI, a rare cutter, limits the number of fragments amplified (Weising *et al.*, 2005). All AFLP experiments require ligation of adaptors to complimentary sticky ends produced by the restriction enzymes to flank known DNA sequences. Restriction and ligation can be done in one single step or in two separate stages using suitable adaptors that are single stranded and must be annealed to each other.

The main advantage of AFLP markers is that the protocol is applicable widely, regardless of the species it is being applied to, such that generally little optimisation of the method is required. Owing to their greater reliability and high variability, using a range of different AFLP primer combinations, larger numbers of AFLP markers can be generated. For this reason, AFLP markers have been used extensively in fungal research, as they are very efficient in detecting polymorphisms.

Since AFLP markers have the ability to generate reproducible markers, they have thus proved to be suitable in fungi with low genetic variation. For example limited molecular variation was observed within rust isolates in Australia and New Zealand, using RAPD and AFLPs (Steele *et al.*, 2001). This study observed that the lack of variation detected between isolates was due to mutations arising from the introduction of a single pathotype, with no introductions of an exotic isolate. With their close genetic relationship, the Australasian isolates were subsequently compared with pathotypes from the UK in order to further identify variation, in which it was discovered that AFLPs could detect more reproducible polymorphisms, with fewer primer sets in comparison to RAPDs, a finding consistent with previous studies (Becerra *et al.*, 2007).

Studies that looked at yellow rust populations from Australia and Europe revealed that these isolates had low genetic diversity, contrary to Chinese populations that displayed high genetic diversity between isolates (Villaréal *et al.*, 2002, Duan *et al.*, 2010). Using AFLP markers enabled a molecular comparison of the two populations, illustrating that highly diverse populations from China strongly contrast with European isolates, which was reflected by the higher number of polymorphic bands detected in Chinese populations (Duan *et al.*, 2010).

In another study in Australia, AFLP analysis was used to assess the extent of genetic variation in five rust pathogens, (*Puccinia graminis* f. sp. *tritici (Pgt*), *P. triticina (Pt), P. striiformis* f. sp. *tritici (Pst),* the oat stem rust pathogen *P. graminis* f. sp. *avenae (Pga*) and barley grass yellow rust (*Bgyr*)) (Keiper *et al.*, 2003). The aim of this study was to understand the genetic makeup of these rust populations by comparing three molecular techniques; AFLP, and two techniques which were a modification to AFLPs (sequence-specific amplification polymorphisms (S-SAP) and selectively amplified microsatellites (SAM)). Their findings illustrated that although all three techniques detected polymorphism among the isolates, the original AFLP procedure produced the greatest number of polymorphisms per primer combination, and subsequently it offered the most informative fingerprinting data, proving to be the best marker type for discriminating species.

Furthermore, in a study in Iran which looked at wheat leaf rust, AFLP analyses revealed the high variability between the rust populations, suggesting that high gene flow has occurred between different provinces (Dadrezaie *et al.*, 2013). This research confirmed results of other studies, that there is high variability of different rust fungal pathotypes, particularly within and between *Puccinia* species (Majer *et al.*, 1996, Villaréal *et al.*, 2002, Balta *et al.*, 2014). They confirmed that gene flow and reproduction are important factors that determine the genetic structure of rust populations (Dadrezaie *et al.*, 2013).

Majer *et al.* (1996) adapted the AFLP protocol to be used in other fungal genomes in which the study analysed genetic variation in the fungal pathogens *Cladosporium fulvum* the causal agent of tomato leaf mould and *Pyrenopeziza brassicae*, which causes light leaf spot on oilseed rape. This study observed that this procedure was reproducible for the polymorphic markers found in the pathogens. They later used AFLP markers to determine the high level of heterogeneity in *P. brassicae* isolates from different regions (the UK, Germany, and France), results suggested high levels of sexual reproduction, because a clonal pathogen would otherwise display a lower level of heterogeneity. An average of 16 polymorphic markers were observed per primer combination (Majer *et al.*, 1998).

The consistency in AFLP studies of fungal pathogens illustrate that DNA based markers such as AFLPs allows efficient comparisons between different pathotypes of yellow rust. Therefore, in this study we chose to develop AFLP markers further for *Pst*, as these markers allow the opportunity for high rates of polymorphism detection within a species where genotypes are expected to be closely related.

2.1.3 Aims & Objectives

- ✓ AFLP markers used to explore the genetic relationships between different pathotypes of yellow rust.
- ✓ AFLPs will be used to differentiate old and new pathotypes of yellow rust.
- ✓ Identify markers, which correlate with specific virulence/avirulence phenotypes.

2.2 Methods and Materials

2.2.1 Fungal Material

Eight different isolates from the UK have been selected for AFLP analysis. The origins and virulence profiles are listed in Tables 2-2 and 2-3. Fungal spore material was received from John Innes Rust culture collection, UK and National Institute of Agricultural Botany (NIAB), UK. Spores were stored in -80°C freezer to preserve the DNA quality, before DNA was extracted from dry urediospores using the DNeasy plant mini DNA extraction kit (QIAgen) (See below).

Table 2-2 Five historical isolates of yellow rust species *Pst*

Isolate	Year	Origin	Natural Host	Virulence Profile on Wheat	Source
234E139	1995	UK	Wheat	*Yr* 2,3,4,7,9	John Innes rust culture collection, UK
169E136		UK	Wheat	*Yr* 1,2,3,9,17	John Innes rust culture collection, UK
237E141	1988	UK	Wheat	*Yr* 1,2,3,4,6,9	John Innes rust culture collection, UK
109E9		UK	Wheat	*Yr* 1,2,3,4,6,9	John Innes rust culture collection, UK
94/519	1994	UK	Wheat	*Yr* 1,2,3,9,17	John Innes rust culture collection, UK

.

Table 2-3 Three current isolates of yellow rust species *Pst.* All isolates originate from UK.

Isolate	Year	Natural Host	Virulence Profile on Wheat	Source
11/140 KWS Sterling (KWS)	2013	Wheat	*Yr* 1,2,3,4,6,7,9,17,32, Rob, Sol, Ste	Niab, UK
08/21 Solstice	2008	Wheat	*Yr* 1,2,3,4,6,7,9,17,32	Niab, UK
11/08 Warrior	2011	Wheat	*Yr* 1,2,3,4,6,7,9,17,32, Rob, Sol, Sp,Tim, War	Niab, UK

2.2.2 Production of wheat plants & Inoculation with *Pst*

The first strand of work involved sowing approximately 12-15 untreated seeds of two different wheat varieties (Robigus and Oakley) in Levington M3 compost. Plants were grown in the growth room and treated with 0.4 gL^{-1} maleic acid hydrazide on days 7 to 10 to ensure slow growth of seedlings.

On days 9-12 plants were ready for inoculation. Previously collected yellow rust urediospores are stored at -80˚C in Eppendorf tubes and left to reach room temperature, before diluting them with talc at a ratio of approximately 1 part spores to 20 parts talc. The talc and spore mixture were placed in a Sterilin tube, covering the top with a single layer of muslin. Plant pots are placed in a tray filled with approximately 1-2cm of water, and the talc/spore mix was shaken over the plants, ensuring plants are covered reasonably well. In order to provide an environment of high humidity, the inside of a propagator lid was sprayed with water and used to cover the plants overnight. The temperature regime was set to around 8˚C night and 18˚C day so that the yellow rust will germinate.

Since, despite several attempts, infections did not develop in the growth rooms, a second approach was taken in which wheat plants were grown in the growth room, but inoculated (as described above) in a cold glass house. They were then moved to a second glasshouse (with conditions of 15˚C and vents at 10˚C) the following morning. This proved to be successful and allowed spores to be collected for approximately two weeks and harvested every other day in the mornings. To harvest spores, plants are tapped over a foil sheet ensuring there is no water or soil on foil, collected into labelled Eppendorf tubes and stored at -80˚C for later use. This process was continuous, with the aim to bulk spore numbers.

Spore Preparation, DNA Extraction and Whole Genome AmplificationAlthough this method of inoculation was successful, spore numbers for all isolates were not always large enough to undergo AFLP experiments. Recent technological advances such as whole genome amplification, have been developed to amplify small amounts of DNA into larger quantities (Pinard *et al.,* 2006), and this method was therefore applied to the above isolates (Table 2-2 and 2-3) in order to preserve limited stocks and potentially improve the quality of amplified products. This was achieved through use of the REPLI-g mini kit (QIAgen), following the manufacturer's instructions). This method uses a Multiple Displacement Amplification (MDA) technology, which offers low DNA bias, but also ensures large coverage of the genome (Lasken, 2009). This approach was also applied in Chapter 3 where the DNA was used for Illumina sequencing. Whole genome amplification has been important in amplifying limited DNA samples, whilst also enhancing its quality. This process was undertaken once DNA was extracted from spore samples (Section 2.2.4).

Spore preparation involved placing the spores into liquid nitrogen for 3 min and breaking them using the FastPrep at a speed of 65 rpm for 45 sec. This process was repeated 3 times to ensure spores were fully broken, and checked under a microscope. For DNA extraction two methods were used. Initially, the QuickPick™ SML Plant DNA kit (Bio-Nobile) following the manufacturer's instructions was used as a quick and simple extraction method. However, the use of this extraction method did not seem sufficient for AFLPs to produce strong banding patterns. For this reason yellow rust DNA was thereafter extracted from spores using a DNeasy plant mini DNA extraction kit (QIAgen), following the manufacturer's instructions. Genomic DNA was extracted from similar amounts of yellow rust spores (approximately 100 mg) to ensure a comparable concentration of DNA extractions across all samples. All DNA concentrations were checked on nanodrop and also ran on gel electrophoresis to check DNA intensity. DNA concentrations for the rust isolates used for AFLP and sequencing experiments throughout this study were as follows for the rust samples:

- Sample 234E139 DNA Concentration was 37.5 ng/µl
- Sample 169E136 DNA Concentration was 37.7 ng/µl
- Sample 237E141 DNA Concentration was 46.3 ng/µl
- Sample 109E9 DNA Concentration was 55.6 ng/µl
- Sample 94/519 DNA Concentration was 19.0 ng/µl
- Sample KWS DNA Concentration was 69.2 ng/µl
- Sample Solstice DNA Concentration was 58.8 ng/µl
- Sample Warrior DNA Concentration was 48.7 ng/µl

2.2.3 Preparation of DNA digests and adapter-ligated fragments

DNA used was from REPLI-g mini kit discussed in section 2.2.3. Fifteen µl of genomic DNA was digested using 1 µl *Mse*I and 1 µl *Eco*RI (5 U of each enzyme) in the appropriate restriction enzyme buffer (5 µl of 10X Tango buffer) and 3 µl dH_20

in a total volume of 25 µl for 2 h at 37°C. The enzyme activities were terminated by incubating the reactions at 70°C and placing on ice for 15 min. Following digestion, double stranded adapters were ligated to the fragment ends. Twenty µl adapter-ligation mixture (1 µl *Eco*RI- adapter, 1µl *Mse*I-adapter µl, 1µl T4 DNA Ligase (1U), 6 µl T4 Ligase buffer (5X), 11 µl dH_20) were added to each digest sample, followed by incubation at room temperature (22°C) overnight.

The ligated products were diluted with TE buffer (1:10 dilution) before performing the pre–selective amplification reaction. The digested DNA was checked against undigested DNA on a 1% agarose gel to confirm successful digestion.

2.2.4 Pre-selective amplification

The initial PCR amplification reaction was carried out using the digestion/ligation products as a template with universal primers. This stage is known as pre-selective amplification. The PCR reaction was prepared in a 25 µl total volume containing 1 µl *Eco*RI-universal primer, 1 µl *Mse*I-universal primer, 5 µl diluted DNA mixture, 18 µl dH_20, and PuReTaq Ready-To-Go™ PCR Beads (GE Healthcare illustra™). The initial 30 cycles of PCR amplification were performed as follows: 94°C denaturation for 30 sec, 56°C annealing for 1 min, and 72°C extension for 1 min. PCR products were checked on a 2% agarose gel, they appear as a smear to illustrate the successful digestion. The pre-amplified DNA was diluted (5 µl digested DNA and 15 µl TE buffer) and used for selective amplification.

2.2.5 Selective amplification

Selective amplification was performed using the diluted products of the pre-amplification stage as the templates. Primers used in this reaction had additional selective nucleotides at the 3' end (adapter sequence plus 2 or 3 extra nucleotides) (see Table 2.5). One of the primers was also fluorescently labeled to allow visualisation (*Eco*RI primers were either labeled with 5',6'-carboxyfluorescein (FAM), or with 5',4',7',2',4',5',7-hexachloro-6-carboxyfluorescein (HEX) (WellRED fluorescent dye D3 and D4, Sigma-Aldrich).

Selective amplifications were performed in a 25 µl total volume reaction containing 5 µl of the diluted DNA, 1 µl of *Mse*I and *Eco*RI primers each having 2 or 3 selective nucleotides, 18 µl dH_20, and PuReTaq Ready-To-Go™ PCR Beads (GE Healthcare illustra™). The initial PCR programme was an 11-cycle touchdown programme: 1 cycle of 94°C for 30 sec, 69°C for 1 min, 72°C for 1 min. This was followed by 25 cycles of the following programme: 94°C for 40 seco, 56°C for 1 min, 72°C for 2 min, and a final elongation step at 72°C for 15 min. All PCR amplifications were carried out using a BioRad S1000 Thermocycler PCR machine.

All PCR products were checked on a 2% agarose gels containing ethidium bromide and were run at 120 volts for 45 min, to be visualized under UV light. Fluorescence-labelled PCR products were analysed using capillary electrophoresis on an automated DNA sequencer (CEQ Beckman Coulter (UK) Ltd), allowing results to be visualized as peaks on an electropherogram.

2.2.6 Data Analysis

The data from the CEQ Genetic Analysis System was exported to *MS Excel* and manually examined, cleaned up and edited before transformation to binary coding. The absence of a peak/band is indicated by '0' and the presence of a band is indicated by '1'. Monomorphic fragment peaks were not scored. The phylogenetic analysis of the AFLP data was done using *FreeTree* software using UPGMA (Pavlicek *et al.,* 1999). The distance matrix was calculated using the Nei and Li distance (Nei and Li, 1979). Resampling was done by boostrapping with 1000 replicates. The phylogenetic tree derived from *FreeTree* was viewed using Treeview and MEGA 6.0.6.

Table 2-4 AFLP primer combinations used to generate loci for addressing the genetic difference between yellow rust isolates, with the selective bases listed. (Working primers were all at 10 mM). *Shaded areas indicate the primer combinations presented within this study. The unshaded primer combinations were tested but did not consistently show good results.

Selective Amplification Primer Combinations (5'-3')			
	***Mse*I** –GACGATGAGTCCTGAGTAA + **2 or 3 selective nucleotides**		***Mse*I** –GACGATGAGTCCTGAGTAA + **2 or 3 selective nucleotides**
***Eco*RI –GACTGCGTACCAATTC + AA**	+AA	***Eco*RI –GACTGCGTACCAATTC + AC**	+AA
	+AC		+AC
	+AG		+AG
	+AT		+AT
	+CA		+CA
	+CC		+CC
	+CG		+CG
	+CT		+CT
	+GA		+GC
	+GC		+GA
	+GG		+GG
	+GT		+GT
	+TA		+TA
	+TC		+TC
	+TG		+TG
	+TT		+TT
	+CAA		+CAA
	+CAT		+CAT
	+CTC		+CTC
	+CTG		+CTG

Table 2-5 Adapters and primer sequences used in selective amplification and pre-selective amplification experiments

Adapters	**Sequence (5'-3' and 3'-5')**
*Eco*RI- adapter 1	5'-CTCGTAGACTGCGTACC-3'
*Eco*RI-adapter 2	3'-CATCTGACGCATGGTTAA-5'
*Mse*I- adapter 1	5'-GACGATGAGTCCTGAG-3'
*Mse*I- adapter 2	3'-TACTCAGGACTCAT-5'
Universal primers	**Sequence (5'-3' and 3'-5')**
*Eco*RI	5'-CGTAGACTGCGTACCAATTC-3'
*Mse*I	5'-GACGATGAGTCCTGAGTAA-3'
Selective primers	**Sequence (5'-3' and 3'-5')**
*Eco*RI	5'-GACTGCGTACCAATTC-3' (plus selective nucleotides and fluorescence labeled with dye D3 or dye D4)
*Mse*I	5'-GACGATGAGTCCTGAGTAA-3' (plus selective nucleotides)

2.3 Results

AFLP was carried out on eight yellow rust isolates (see Table 2-2 & 2-3). The selective primers (*Eco*RI and *Mse*I) (see Table 2-5 for sequences) were used in 21 combinations to assess the number of polymorphism that could be detected using the AFLP protocol.

2.3.1 DNA Digestion

DNA digestion is a crucial first step in the AFLP process. Successful DNA digests are illustrated in Figure 2-1, as the fragmented samples display clear smears. The figure differentiates between cut and uncut DNA with distinct solid bands apparent for uncut DNA.

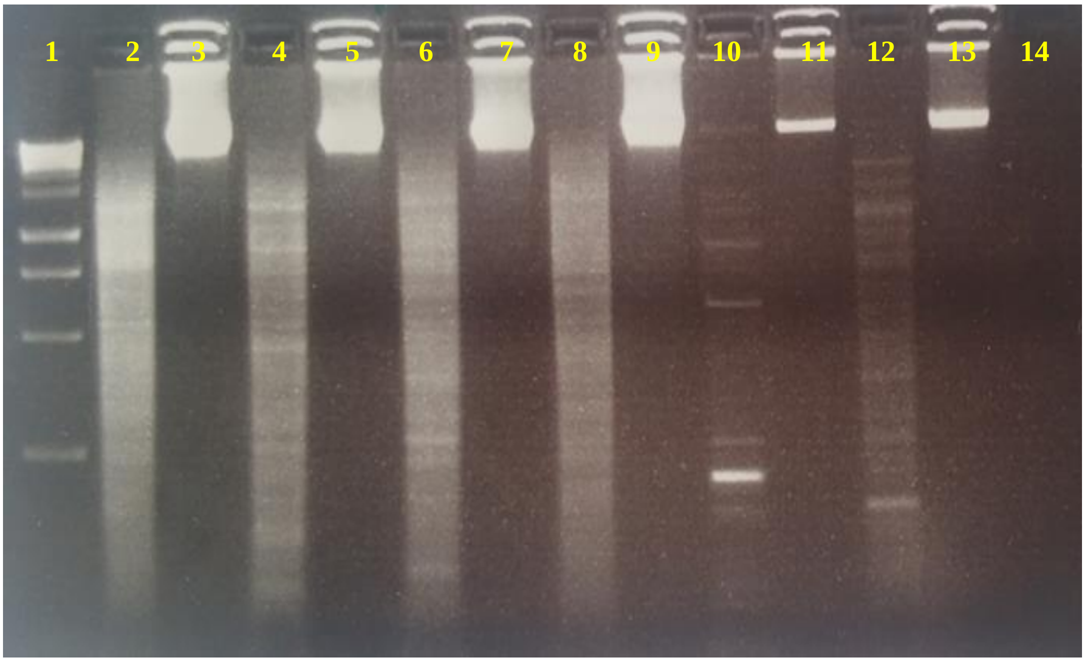


Figure 2-1 Example of genomic DNA digested with restriction enzymes *Mse*I and *Eco*RI. Lane 1: 1kb ladder. 5 µl DNA was loaded onto the gel, lanes 2, 4, 6, 8, 10, and 12show smears of digested DNA. For comparative reasons lanes 3, 5, 7, 9, 11, and 13 are also presented to illustrate the uncut DNA. Lanes include all historical isolates and KWS. 1: 1kb DNA ladder 2: 234E139 4: 169E136 6: 237E141 8: 109E9 10: 94/519 12: KWS 14: Water control.

2.3.2 Pre-selective amplification

In order to ensure maximum amplification, the samples were subjected to two rounds of PCR, the first referred to as the pre-selective amplification and the second known as the selective amplification. Pre-selective amplification is an optional step in most AFLPs, but for this study pre-selective amplification was used in each AFLP experiment. Pre-selective amplified restriction fragments with non-selective primers were checked on a 2% agarose gel.

Yellow rust DNA was successfully pre-amplified as illustrated in Figure 2-2 in which clear smears are visible, with a clear smear product in the 50 – 550 base pair range. The genomic DNA is fully restricted in all lanes (Fig. 2-2).

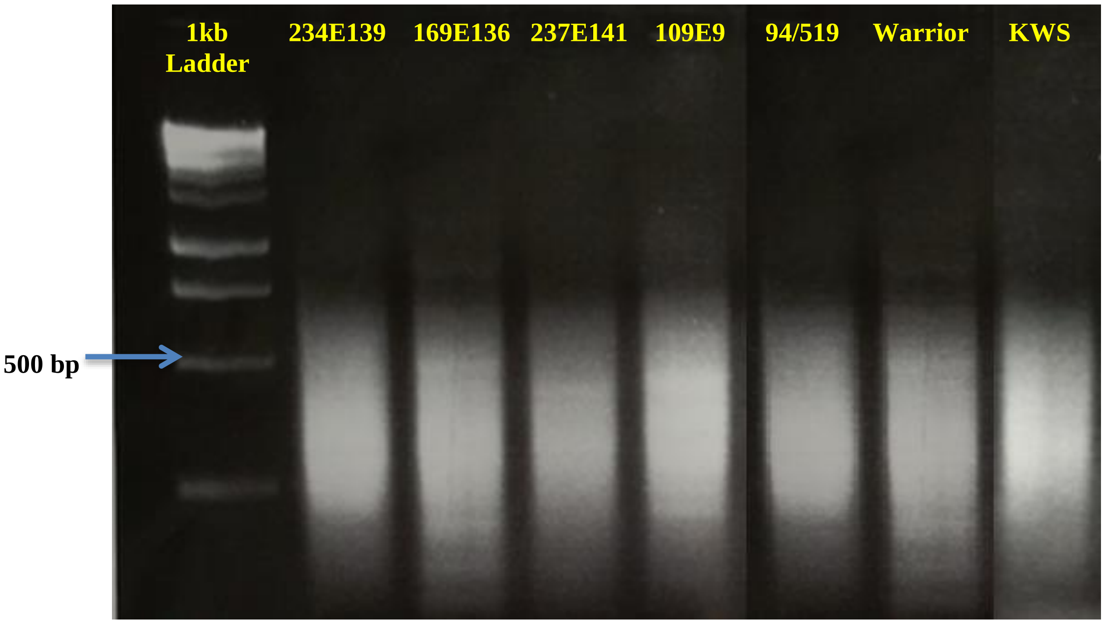


Figure 2-2 Pre-selective amplified restriction fragments with non-selective primers. Successfully pre-amplified yellow rust DNA samples were checked on a 2% agarose gel, with smears ranging from approximately 50 – 750bp.

2.3.3 Selective amplification

Isolates were tested with AFLPs to detect polymorphisms; common markers were visible between 250 bp and 650 bp amongst the isolates. For each of the eight isolates, the AFLP patterns were described through manually recording bands present on a 2% gel for each polymorphic marker (Fig. 2.3). Selective amplification PCR fragments were highly reproducible across the replicates. Using gel electrophoresis it was possible to detect polymorphic regions across the isolates with peaks ranging in size from 50 to 500bp were selected as polymorphic markers. On average, nine bands were observed per primer combination.

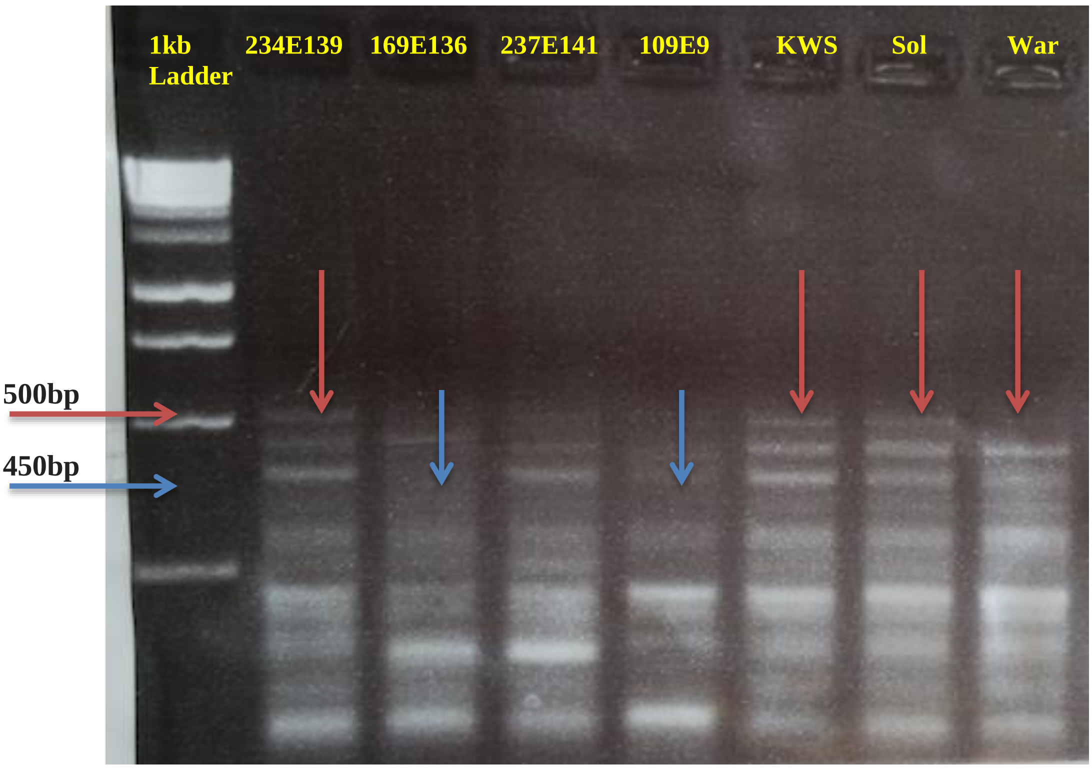


Figure 2-3 Successful selective amplification from primer combination *EcoR*I AA and *Mse*I GC. Isolate 234E139, 169E136, 237E141, 109E9, KWS Sterling, Solstice, and Warrior. KWS, solstice and Warrior show almost identical fingerprints with base pairs ranging from 50 to 300. The historical isolates vary in fingerprinting in

comparison to the more recent isolates. The red arrows indicate polymorphism being detected in this amplification at approximately 450 bp. At approximately 500 bp KWS, Solstice, & Warrior share a clear band which is not present in any other isolate, other than 234E139.

2.3.4 Fragment Analysis

Fragment analysis was conducted using a DNA sequencer (CEQ Beckman Coulter (UK) Ltd), which detects and groups fragments produced in AFLP reactions for better resolution of separated AFLP fragments. The CEQ software scores the absence and presence of AFLP fragments in an electropherogram, to allow better data comparisons (Fig. 2-4 & 2-5). Polymorphisms between the yellow rust isolates 234E139, 169E136, 237E141 are displayed in Figure. 2-4, in which specific peaks differentiate them. This enabled better understanding of fingerprinting patterns between the closely related species, so as to study phylogenetic relationships among the yellow rust species.

2.3.5 Polymorphisms between historical isolates and current isolates of wheat yellow rust.

Polymorphisms between yellow rust isolates are represented in Figure 2-4. Up to four polymorphic regions were detected between isolates using a single primer combination. The red dye represents the size standards and the green peaks are representative of the fragments. Peaks present in all historical *Pst* isolate, but absent in the KWS Sterling, Solstice, and Warrior were revealed. Similarly, peaks present in the newer set of isolates but absent in the older pathotypes illustrates the genetic diversity between the two *Pst* populations.

Out of the 9 successful primer combinations, all apart from one combination (*Eco* TA + *Mse*I AG) displayed polymorphisms. Some primer combinations illustrated clear distinct fingerprints, which easily differentiated between the isolates, whilst others did not demonstrate strong bands, though distinctions were possible. Only distinct and reproducible bands were included in the AFLP cluster analyses. On average nine bands were produced per primer combination, on an agarose gel, whilst 20 to 40 peaks were produced per primer combination for the older isolates using the CEQ software, in comparison to KWS Sterling, Solstice, and Warrior, which generated roughly 20 to 55 peaks per primer combination.

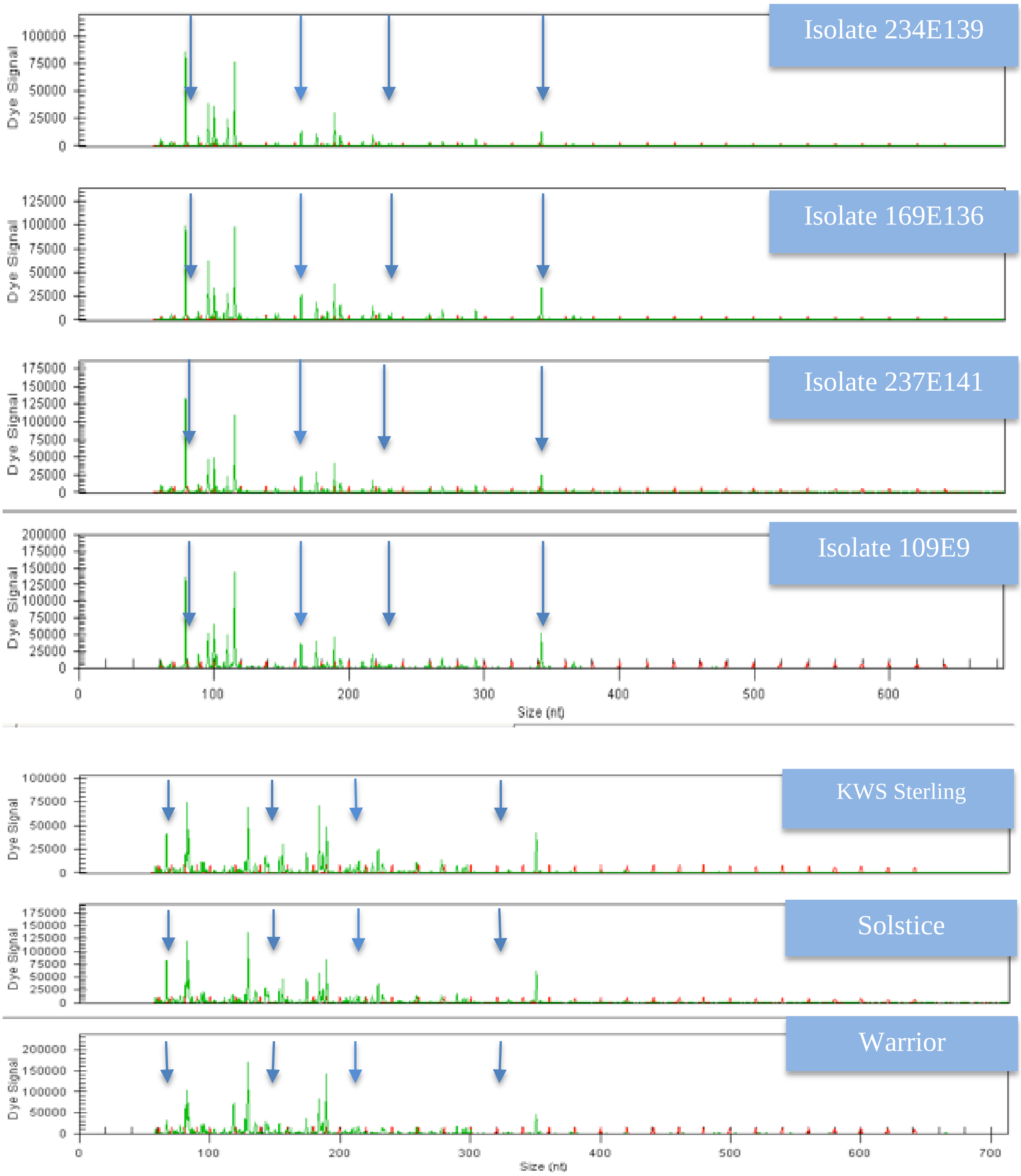


Figure 2-4 Example of AFLP electropherograms results. Four historical isolates 234E139, 1693136, 237E141, 109E9 and three recent isolates KWS Sterling, Solstice and Warrior respectively, highlighting four polymorphic regions. Primer combinations *Eco*RI AA + *Mse*I TA.

The fragment analysis process detected distinct peaks to differentiate between isolates 234E139, 1693136, 237E141, 109E9 and KWS Sterling, Solstice and Warrior. Figure 2-5 further support the results from Figure 2-4, as another primer combination (*Eco*RI AC + *Mse*I AC) which also displays distinctions between the old and new pathotypes (Fig. 2-5). The AFLP fingerprint of the Warrior isolate is very distinct to the older isolates, with a higher number of fragment peaks detected. This correlates with Fig. 2-3 and Fig. 2-5, which also amplify larger number of bands in the newer isolates in contrast to the historical isolates. Warrior is a good representation for the difference in genetic fingerprints expected to see in the newer isolates, when compared to the historical isolates. As expected, the fragment analysis and software produced better resolution of bands/ fragments and greater reliability than the gel picture.

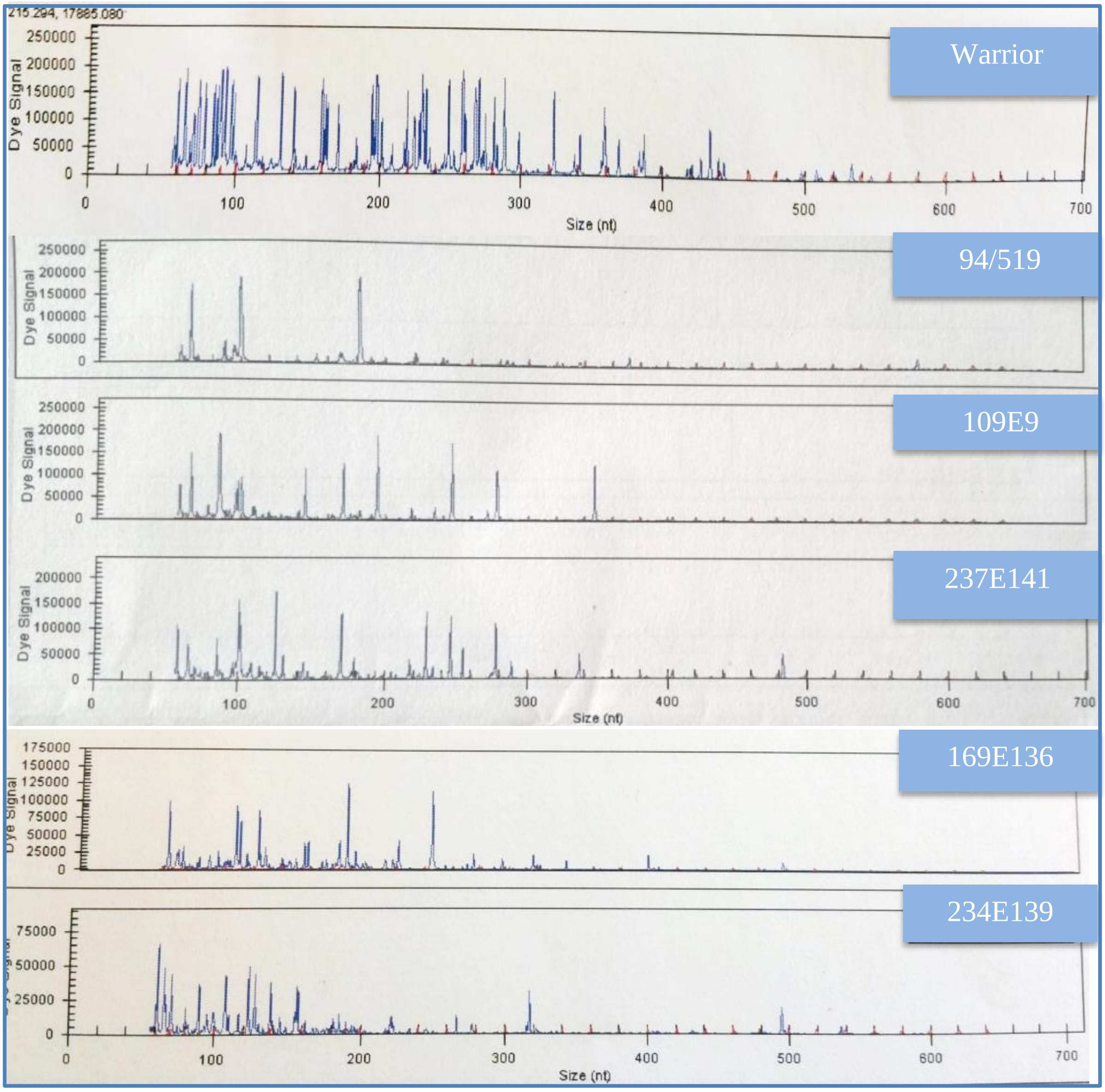


Figure 2-5 Electropherogram results from AFLP analyses showing the genetic differentiation between a modern isolate and historical isolates. Results from Primer combination *Eco*RI AC + *Mse*I AC displaying the same major peaks shared by all historical isolates

2.3.6 Phylogenetic Analysis with AFLP Markers

Markers developed from the AFLP analysis were manually sorted into a binary format using Excel Spreadsheets; in which all successfully amplifying primer combinations were put together (See section 2.2.7).

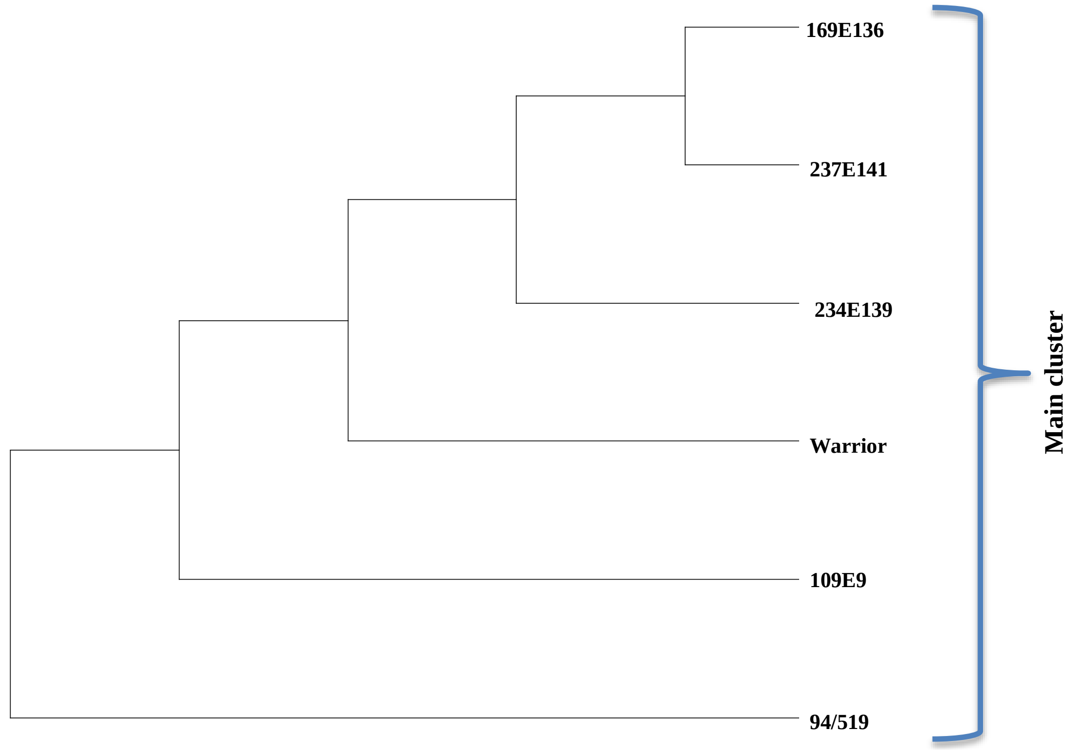


Figure 2-6 Phylogenetic tree displaying primer combinations that successfully amplified polymorphisms (*Eco*RI AC + *Mse*I AA, *Eco*RI AC + *Mse*I AC, *Eco*RI AC + *Mse*I AG, *Eco*RI AG + *Mse*I AT, *Eco*RI AC + *Mse*I CC, *Eco*RI AC + *Mse*I AA, *Eco*RI AC + *Mse*I CG, *Eco*RI AC + *Mse*I CT, *Eco*RI AC + *Mse*I GC). Samples included 234E139, 169E136, 237E141, 109E9, 94/519 & Warrior. One major cluster is displayed, with isolate 94/519 as an outlier.

Table 2-6 Similarity matrix values of the primer combinations that successfully amplified polymorphisms (*Eco*RI AC + *Mse*I AA, *Eco*RI AC + *Mse*I AC, *Eco*RI AC + *Mse*I AG, *Eco*RI AG + *Mse*I AT, *Eco*RI AC + *Mse*I CC, *Eco*RI AC + *Mse*I AA, *Eco*RI AC + *Mse*I CG, *Eco*RI AC + *Mse*I CT, *Eco*RI AC + *Mse*I GC). This table is in correlation with Figure 2.6, similarly highlights similarities between isolates 169E136, 237E141 and 234E139, Warrior

	234E139	**169E136**	**237E141**	**109E9**	**94/519**	**Warrior**
234E139						
169E136	0.221					
237E141	0.201	***0.245***				
109E9	0.179	0.181	0.184			
94/519	0.160	0.115	0.147	0.142		
Warrior	0.210	0.202	0.184	0.186	0.162	

0.245 underlines that isolate 169E136 and 237E141 share the highest similarity value & 0.186 representative of a historical isolate lacking similarity to a new isolate Warrior.

Figure 2-6 illustrates a phylogenetic analysis of the different isolates from the AFLP marker data developed in the present study. Coinciding to that is Table 2-6 that presents the similarity matrix values for the studied isolates. The similarity coefficients varied from 0.245 between isolate 169E136 and 237E141 to 0.186 between isolates109E9 and Warrior. The high similarity value is indicated by the bootstrapping value of 93. Isolate 169E136 and 237E141 share the highest similarity value, in comparison to other pathotypes when compared to isolate 169E136. This is a good indication of the relatedness between the older isolates, which by contrast is not seen with old and new pathoypes. For example the low matrix value highlighted in Table 2-6 between Warrior and 109E9 is a good representation to show the lack of genetic similarity between old and new isolates.

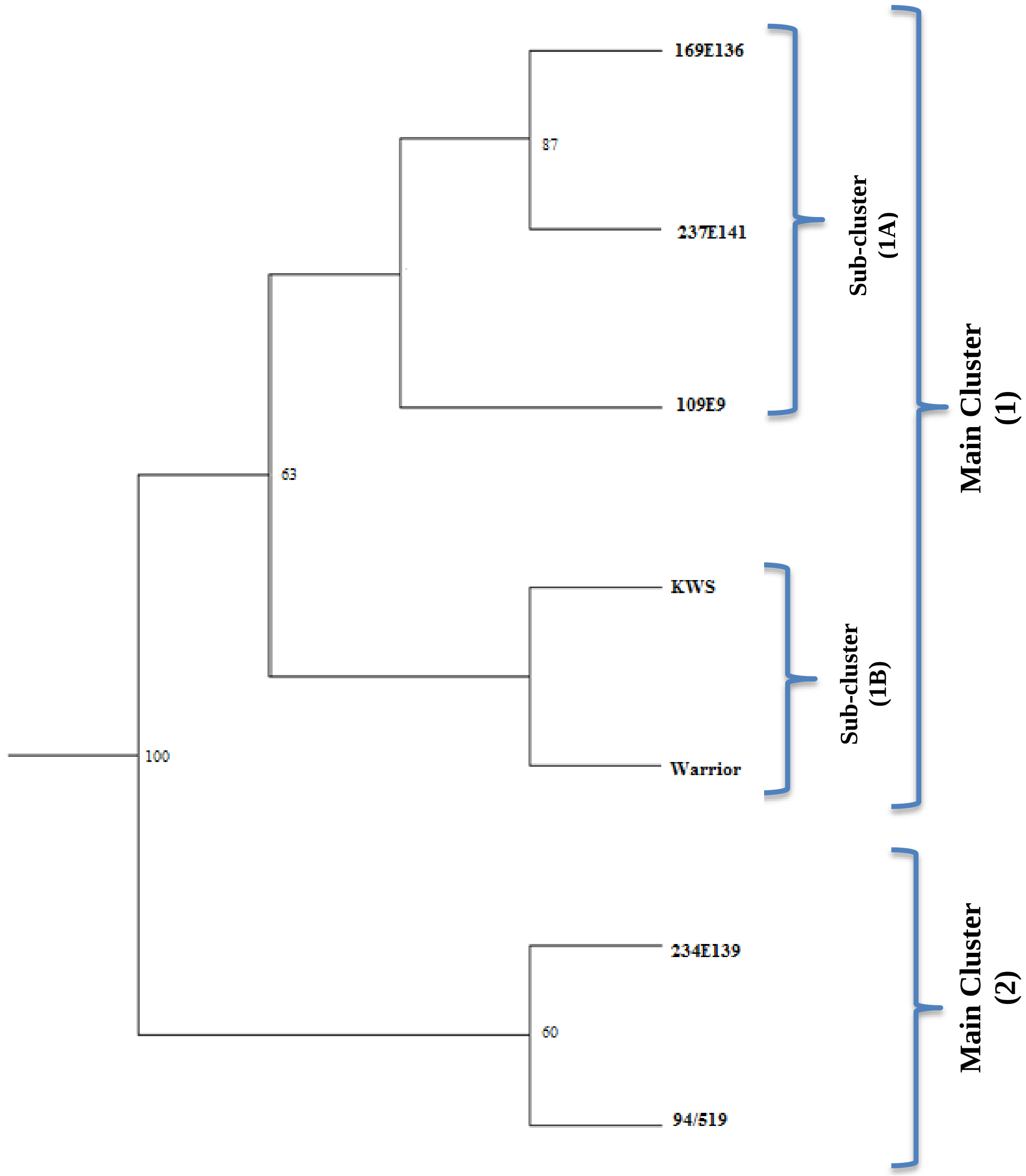


Figure 2-7 . Phylogenetic tree for the primer combination *Eco*RI AC + *Mse*I AG. This tree is an illustration of how an individual primer set represents the same result displayed Figure 2-6 (in which all successfully amplifying primer combinations were put together). For example, isolate 169E136 and 237E141 are closely together in sub-cluster 1A with a high bootstrapping value of 87.

Comparing phylogenetic trees with genetic similarity values is a good exemplification of the AFLP data obtained earlier. Overall, the phylogeny displayed distinct clusters. Figure 2-7 shows 2 major clusters with 2 sub-clusters in cluster 1. Within this sub-cluster (1A) isolate 169E136 and 237E141 have a bootstrapping value of 87, but also share similarities with another historical isolate 109E9. The distinction between sub-cluster1A and sub-cluster 1B clarify the genetic diversity seen between the historical and new isolates. Additionally, distinct clustering of isolate 234E139 and 94/519 further supports the notion that the old and new pathotypes are different from each other.

The common relationship between both sub-clusters is envisaged through a bootstrapping value of 63, demonstrating similarity between sub-cluster 1A and 1B. Moreover, isolate 94/519 is typically an outlier in most phylogenetic data sets and in the AFLP results because banding patterns were unclear (results not shown). However, within this primer combination (*Eco*RI AC + *Mse*I AG) it is showing similarities with 234E139 with a bootstraping value of 60 and a similarity matrix value of 13% (0.132). This may be due to specific DNA fragments amplified within this particular primer combination that highlighted the similarities between isolate 94/519 & 234E139. Overall the common relationship between both major clusters had a bootstrapping value of 100, signifying that if these AFLP experiments were to be repeated, the samples would share similarities to a reasonable degree.

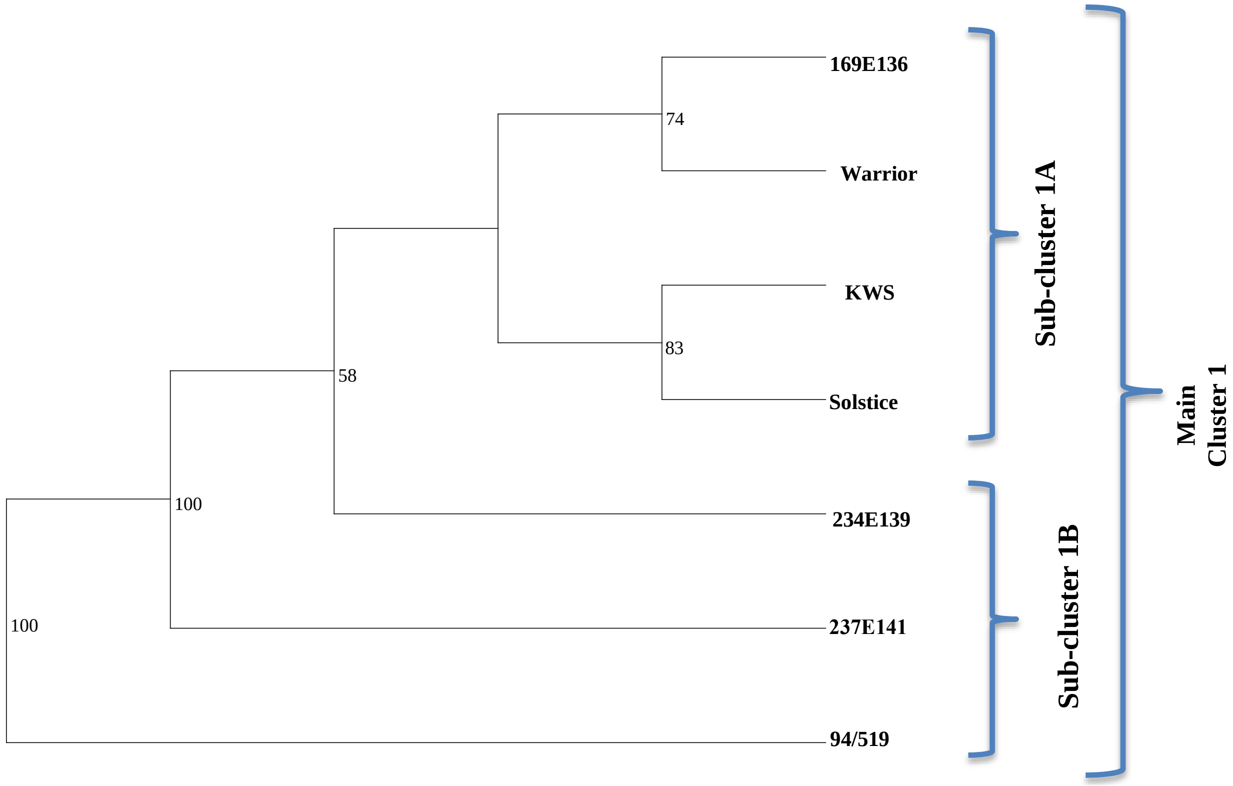


Figure 2-8 Phylogenetic tree for the primer combination *Eco*RI AA + *Mse*I TG.

Table 2-7. Similarity matrix value for the primer combination *Eco*RI AA + *Mse*I TG. (In correlation to Figure 2.8).

	234E139	169E136	237E141	94/519	KWS	Solstice	Warrior
234E139							
169E136	0.591						
237E141	0.578	0.621					
94/519	0.396	0.408	0.375				
KWS	0.685	0.682	0.627	0.470			
Solstice	0.644	0.669	0.513	0.449	0.754		
Warrior	0.644	0.726	0.651	0.389	0.680	***0.650***	

High genetic similarity values identified amongst these 3 isolates

Figure 2-8 demonstrates the genetic similarities within the pathotypes KWS Sterling, Solstice, and Warrior, all of which are part of 1 sub-cluster (Figure 2-8, sub-cluster 1A). They all share similarities with isolate 169E136, with similarity matrix values between 0.68369 and 0.72632 (Table 2-7). Within a sub-cluster, KWS Sterling and Solstice share a high bootstrapping value of 83, whilst Warrior and 169E136 also share a high bootstrapping value of 74. In accordance to the full AFLP data (Fig 2-6) the historical isolates 234E139, 237E141, and 94/519 form a sub-cluster of their own, whilst the recent isolates are in their own sub-cluster with greater similarity comparisons to isolate 169E136. The older isolates should display genetic variation from the newer isolates in distinct clusters; the achieved results determined this through an accurate representation of what was expected.

2.4 Discussion

2.4.1 Analyses of the genetic variation between and within *Pst* isolates

In this study, different *Pst* pathotypes from the UK population were studied, including both old and new isolates. The aim of this chapter was to determine the genetic relationship among the studied *Pst* population. The results revealed that there were genetic differences between the studied old and new isolates. Understanding the genetic variation of these isolates, rather than variation in physiological or morphological traits makes it easier to understand the population structure, and thus develop identification tools to recognise new strains in the future. As it is difficult to bulk the number of yellow rust spores, AFLP markers have been ideal because they require small amounts of DNA material.

The number of polymorphisms detected using this AFLP technique was ample enough to differentiate between particular isolates, and phylogenetic analyses displayed that all isolates, as expected, share similar genetic makeup, with close genetic relationships (Fig. 2-6, 2-7, and 2-8). This is consistent with other AFLP fungal studies, which displayed low genotypic diversity for *Pst* and reproducible bands ranging from 100 to 350bp (Villaréal *et al.,* 2002). In another study, Liu *et al.* (2011) estimated the genetic relationships between *Pst* populations within China, in which they also observed low level genetic diversity in one particular province, compared to the other region. This current study observed *Pst* isolates from the UK, with KWS Sterling, Solstice and Warrior originating from the UK more recently, suggesting that genetic variation exists amongst this group of newer strains of isolates and the older strains is potentially derived from different factors. The potential shift in genetic variation could possibly relate to the geographic or climatic features, such as the evidence found in Europe, which suggests that the *Pst* migration route was from Europe to Australia and Americas (Ali *et al.,* 2014), such emergences resulting in mutational changes.

Cluster analysis illustrated the overall genetic relationships amongst all the isolates. Results were consistent with AFLP polymorphic markers, which identified the contemporary isolates in one cluster, with likenesses to older isolates such as 234E139; KWS Sterling, Solstice and Warrior were closely grouped in one sub-cluster and the historical isolates were presented in another. Clustering explored the relationships between the isolates, and displayed an overall difference in genetic relationship. However, the lack of isolates used in this study limited exploration of larger groups of isolates, from outside of the UK for example. These results confirm that the AFLP technique is suitable for distinguishing variation between closely related pathotypes within a species.

Markell and Milus (2008) used a similar approach to this study looking at contemporary (isolates collected since 2000) and older (isolates collected before 2000) *Pst* strains from the US, and found distinct differences between the two groups. They looked at 101 different isolates that displayed clear genetic variations, clustered distinctively, such that they ruled out the hypothesis that any isolates after the year 2000 had emerged due to mutations from pre-2000 isolates. As this study looked at a much larger *Pst* population it was possible to determine clusters and understand the discrepancies. Thus, future work following this line of research can focus on observing larger data sets.

As Markell and Milus, (2008) study suggests that the new isolates were different from the old isolates, they imply that genetic differentiation between the pathotypes is likely to have derived from a new exotic introduction, rather than a mutation in isolates of the old population. Findings from this study support the same hypothesis that KWS Sterling, Solstice and Warrior are different from the studied older isolates. Since 2011, the Warrior race has largely replaced previous populations originating from North-West Europe (Hovmøller *et al.,* 2015, Hubbard *et al.,* 2015). Its recent incursion explains its ability to derive from a new exotic introduction and thus displays genetic differences to the older isolates.

SSR markers have previously been developed from complimentary DNA (cDNA) and expressed sequence tag libraries (EST) to study population genetics in rusts (Enjalbert *et al.*, 2002, Dracatos *et al.*, 2006, Bahri *et al.*, 2009). Bahri *et al.* (2009) analysed a published *Pst* EST library from Zhang *et al.* (2008) which contained 1,398 sequences, out of which approximately 15 sequences were selected that contained polymorphic microsatellites. Subsequently, Bahri *et al.* (2011) employed SSR technology in order to characterise yellow rust isolates in Pakistan. They discovered that genetic diversity was high amongst the 49 isolates found in Pakistan using the 15 microsatellite markers, and observed variability was higher than that found in North

West Europe, with the presence of recombination contributing to the high genetic variation in *Pst* populations. The spread of new virulent strains has also been reported by Hovmoller *et al.* (2008) who found migration of the pathogen and recombination as factors contributing to the high variability between isolates. Some preliminary SSR work was undertaken in this current study using the primers developed by Bahri *et al.* (2011). However, because of changes to personnel and closure of the fragment analysis facility, these studies could not be completed so have not been presented.

2.4.2 Avirulence Markers

Another major aim for this chapter was to identify markers, which correlate with specific virulence/avirulence phenotypes. Table 2-8 is representative of the virulence profiles of the isolates studied in this thesis. The avirulences for *Yr* genes 2, 3 and 9 can be discarded from the analysis, since all isolates overcame these resistance genes. However, there were differences in the capacity to overcome *Yr1, Yr4, Yr6, Yr7, Yr17, Yr32, YrRob, YrSol, YrSte, YrSp, YrTim*, and *YrWar* amongst the isolates.

With regards to the AFLP analysis, specific markers were maybe linked to avirulence ~ this was achieved by exploring the binary data obtained from the AFLP analysis of different primer combinations. We can certainly identify a pattern in which isolates are differentiated through their distinct peaks and are correlated with virulences in Table 2-8. For example, peaks, which were present in the all three new isolates, but absent in historical isolates can be associated to the *Avr*32 gene. This is highly anticipated due to the presence of this specific gene only within the three new isolates. Likewise, presence of the *Avr*32 gene can be potentially linked to markers identified in Figure 2-4, which highlights the similarities identified between isolates KWS, Solstice, and Warrior.

Table 2-8 Virulence profiles for the studied pathotypes. Virulence on *Yr* genes '2', '3', and '9' is present in all samples and is therefore highlighted as not significant. Empty boxes signify absence of virulence on a *Yr* gene. Five isolates are named using the binary system, this reveals which series (International or European) the pathotype is, as well as providing an understanding of which *Yr* gene is overcome. This, however, is not the case for KWS Sterling, Solstice, and Warrior. Virulence profiles refer to the *Yr* genes that are overcome, with the corresponding *Avr* genes absent or in its recessive form.

Isolate	**Virulence Profile (*Yr* genes overcome/*Avr* gene absent/recessive)**													
234E139		2	3	**4**		**7**	9							
169E136	**1**	2	3				9	**17**						
237E141		2	3	**4**		**7**	9							
109E9	**1**	2	3	**4**	**6**		9							
94/519	**1**	2	3				9	**17**						
Solstice	**1**	2	3	**4**	**6**	**7**	9	**17**	**32**					
KWS	**1**	2	3	**4**	**6**	**7**	9	**17**	**32**	**Rob**	**Sol**	**Ste**		
Warrior	**1**	2	3	**4**	**6**	**7**	9	**17**	**32**	**Rob**	**Sol**	**Sp**	**Tim**	**War**

A more thorough analysis of potential AFLP markers linked to virulence/avirulence is described in Chapter 4. The aim of this thesis was to develop a range of markers using a random approach (AFLP experiments) and a more targeted approach (SureSelect Illumina sequencing). A final list of potential markers derived through this study (specifically Chapter 2 and Chapter 3), have been analysed and discussed in Chapter 4 together. Here we conclude on the genetic differences analysed between the isolates, with some insight into developed markers. Chapter 4 further analyses the binary data obtained from AFLP data to look at the presence/absence of specific individual AFLP bands with the presence/absence of particular AVRs.

An important conclusion of this study is that AFLP is a powerful technique to gauge the level of variation between isolates and has thus been used to study many rust pathogens (Steele *et al.,* 2001, Justesen *et al.,* 2002, Villaréal *et al.,* 2002). Detecting polymorphisms using this technique has been successful in identifying markers that can be used for the development of management strategies.

Chapter 3

Bioinformatics

Chapter 3: Bioinformatics

3.1 Introduction

Next generation sequencing (NGS), also known as high-throughput sequencing has revolutionised the field of molecular biology. This method has been widely adopted due to the high quality of sequence data, robustness of the technique, and the ability to sequence DNA (and RNA) at an unprecedented speed (Buermans and Den Dunnen, 2014). With costs reducing and methods becoming simpler, the application of this technology has increased greatly. NGS offers inexpensive and accurate genomic information, as more genome sequences continue to be generated. It has therefore become the preferred choice for genomic analysis, slowly replacing microarray technologies and offering a new approach to analyse plant and pathogen genomes.

High-throughput technology with high coverage sequencing generates large data sets, resulting in demanding workflows (Schuster, 2007). Enormous data sets need to be analysed and this also requires terabytes of storage. Analysis of such data requires bioinformaticians with biological insights in order to interpret the high quality data. The general NGS technology pipeline includes 3 main stages; DNA (or RNA) library preparation, sequencing using a NGS instrument, and data analysis (bioinformatics).

Genome sequencing, assembling and annotating are the usual pipelines required in NGS analysis. The different sequencing technologies available have developed different strategies, all the way from preparing the DNA library through to detecting and reading the DNA sequences. What differs in the available NGS technologies is the specific changes in the protocol, which will be discussed below. The comparison of the main sequence systems commonly used is focused in (Table 3-1 P.79). Massively parallel sequencing systems all share good throughput performance and routinely generate reasonably low error rates with good read lengths, which permits researchers to draw reliable conclusions in understanding genomes (Liu *et al.*, 2012). In the following section, the major NGS technologies will be briefly described with a particular emphasis on Illumina, as the work presented in this chapter used Illumina technology to sequence different *Pst* isolates. By briefly understanding each NGS technology, this will further advance our knowledge on the primary concepts that they share. It will also highlight how these technologies are continuously under development to serve their widespread applications. In order to assess genome diversity it was important not to limit this study to molecular markers such as AFLPS, but to complement AFLPs, without hindering the possibility of consistently identifying polymorphic markers

3.1.1 Current Sequencing Technology

Roche 454 was developed in 2005 and uses pyrosequencing technology. Pyrosequencing determines the order of nucleotides in DNA, through the detection of released pyrophosphates (PPi) (which during replication are released due to addition of a deoxynucleotide (dNTP)). Pyrosequencing is the most time-efficient method; this technology is different from the traditional Sanger sequencing method (also known as the chain termination method) in that it relies on chain termination with dideoxynucleotides (ddNTPs) (Johnsen *et al.*, 2013). Although the Sanger sequencing method has been widely used since its development in 1977 (Sanger *et al.*, 1977), it has several limitations, such as the inability to analyse samples in parallel, high error rate and costs, and large number of gels required for larger sequences (Fakruddin and Chowdhury, 2012). Due to these restraints, Sanger sequencing has been superseded by NGS technology.

The Ion Torrent method is similar in concept to that of Roche 454 pyrosequencing but releases a H^+ ion as a dNTP binds with a DNA polymer. pH is detected and with every H^+ ion released, the pH is decreased, thus pH determines if a particular base is added to a sequence read (Johnsen *et al.,* 2013). Sequencing by Oligo Ligation Detection (SOLiD) is an additional NGS technology developed by Life Technologies, but unlike the technologies mentioned above, SOLiD uses sequencing by ligation, rather than DNA polymerisation (Metzker, 2010). Primary information is listed and compared with other massively parallel sequencing systems in Table 3.1.

3.1.1.1 Illimina Miseq

In this study Illumina Miseq has been used for sequencing. Illumina, which was initially developed by Solexa and later purchased by Illumina in 2007, adopts a sequencing by synthesis approach (Liu *et al.,* 2012). A unique 'bridged' amplification reaction occurs in the flow cell, which embeds the DNA on the flow cell surface for amplification. Clustering is the process whereby each DNA fragment is isothermically amplified. Adapters are ligated to the end of DNA fragments to get the sequence to anneal to where the DNA sequence can be determined by the MiSeq sequencer. DNA libraries have P5 and P7 binding regions, which ligate to complimentary oligo sequences on the flowcell surface, which is a glass slide with lanes composed of a lawn of dense oligos. Double stranded DNA (dsDNA) strands attached to the adapters are denatured to single stranded DNA (ssDNA) (Illumina Technologies, 2016). ssDNA fragments are embedded on to the flow cell and ligated to corresponding primer sequences on flowcell surface. Each end of every DNA molecule matches one of the two adapters on the surface of the dense lawn of primers. DNA that doesn't attach is washed away. Unlabeled dNTPs are added and DNA polymerase used to elongate and join DNA strands attached to the flowcell surface, to initiate bridge amplification. 'Bridges' of dsDNA are formed between the primers on the flowcell. dsDNA is denatured to form multiple copies of the same sequence (forming dense clusters) (Holt and Jones, 2008).

Sequencing by synthesis is the process in which sequencing begins with the extension of the first sequencing primer to produce the first read. Fluorescently labeled terminators (A,G,C,T) (that stop DNA synthesis) and primers (which attach to the DNA being synthesized) are added to the flowcell (Aird *et al.,* 2011). Illumina uses fluorescently labeled terminators to detect each nucleotide base. The first fluorescently labeled terminator is incorporated as the DNA polymerase binds to the primer. This prevents any more bases being added to the DNA sequence (until the terminator is removed). Each terminator base (A, G, C, T) gives off a different colour once lasers are passed over the flowcell to activate the fluorescent label. The first fluorescently labeled terminator is removed from the first base, and the next fluorescently labeled base is added alongside. This process is repeated multiple times to produce a dense amount of sequences per area on the flowcell. Incorporation bias is minimized as all four terminator bases are incorporated during the sequencing cycle (Illumina Technologies, 2016).

Paired-end sequencing initiates as polymerases extend the second flow cell oligo forming a double stranded bridge. dsDNA is linearized (1 forward strand and 1 reverse strand) and the 3' end is blocked. The original forward strand is cleaved off and washed away, leaving the reverse strand. Read 2 beings with the introduction of the read 2 sequencing primer. As with read 1, the sequencing steps are repeated until the desired read length is achieved. The read 2 product is washed away. This entire process will generate millions of reads representing all the fragments. Data analysis involves a process in which reads with similar base calls are locally clustered. Forward and reverse reads are paired creating contiguous sequences, and can be used for mapping reads and variant calling.

3.1.1.2 Agilent's SureSelect Target Enrichment System

Whole genome sequencing generates enormous amounts of data, which is laborious to analyse, expensive and extremely time-consuming. Sometimes, it is detailed analysis of specific genomic regions/genes that is required and therefore, techniques have been developed to improve NGS efficiency for such types of studies. Development of targeted enrichment methods allows studies to focus their efforts in identifying genetic variations within specific regions of interests. This results in a more time-efficient and inexpensive process. The combination of Illumina paired-end sequencing technology and the Agilent SureSelect Target Enrichment System is a platform that offers high-resolution targeted genome sequencing, by using targeted sequencing primers also known as amplification primers. This enables a subset of genes to be amplified within the chosen samples and effectively allows 'targeted sequencing' (Chen *et al.*, 2015).

For the purpose of this study, targeted sequencing was applied using the SureSelect Target Enrichment System for Illumina paired-end sequencing, in order to selectively capture DNA sequences. A SureSelect^{QXT} Target Enrichment kit was used in this project, which ensured a shorter hybridisation length of just 90 minutes. This enabled libraries to be ready for sequencing on the same day. With a lack of *Pst* genomes sequenced, application of such NGS technologies rapidly assists the development of molecular genetic information to develop a better understanding in pathogenicity and virulence factors.

The workflow for the genomic capture of regions of interest is done through the use of long RNA baits (biotinylated RNA Library). These baits are cRNA probes, with a length of 120bp, and are biotin labelled. The baits allow for efficient enrichment of targeted regions due to their increased sensitivity. The library is first prepared with the genomic sample, SureSelect HYB buffer and SureSelect RNA baits, preparing the gDNA sequencing libraries for target enrichment. The DNA samples are hybridised to a SureSelect capture library, and the prepared DNA library captured with use of streptavidin coated magnetic beads. A bead separation step ensures that the unbound fraction is discarded and the prepared DNA sample is separated from the magnetic beads successfully. Several steps of washing and eluting are undertaken to ensure successful amplification for sequencing.

Table 3-1 Comparison of the different sequencers available, with advantages and disadvantages displayed for each

NGS Technology	Sequencing mechanism	Read Length (bp)	Time Run	Run Type	Advantages	Disadvantages
HiSeq and MiSeq, Illumina	Sequence by synthesis	2 x 300 (MiSeq) 2 x 100 (HiSeq)	2-10 days	Single and Paired end	High throughput, low costs	Short read assembly and long run time
454 GS FLX, Roche	Pyrosequencing	700	1 day	Single end	Long & fast read length	High error rate, high cost, low throughput
SOLiD, Life	Ligation	50 + 35	7-14 days	Single and Paired end	Very low error rate	Short read assembly and long run time
Ion Torrent, Life	Proton detection	200-400	2 hours	Single end	Very short run time	High error rate

Table 3-2 A brief summary of different studies analysing yellow rust genetics with NGS technologies.

Study	Aim of study	NGS technology	Findings	Pst strains
Cantu *et al.* 2011	Sequence Pst strain, as this had not been done before. Sequencing to look at sequence similarities between different PST strains, as well as similar fungi.	Illumina Genome analyser II	RXLR motif present in potential effector proteins. This is important for delivery of effectos into host cells.	*Pst*-130
Cantu *et al.* 2013	Re-sequenced *Pst* isolates with distinct virulence profiles, and from distinct regions	Illumina RNAseq	Identified effector protein candidates	*Pst*-21 (US) *Pst*-43 (US) *Pst*-130 (US) *Pst*-87/7 (UK) *Pst*-08/21 (UK)
Garnica 2013	Isolated haustoria to gain better understand of effector proteins	Pyrosequencing 454 GS-FLX titanium platform	400 candidate effector proteins were identified in haustorial transcriptome. RT-PCR confirmed expression of 94 of these genes.	104E137A *Pst*-130
Hubbard *et al.* 2015	Developed a pathogenomics approach to analyse population structures.	Illumina HiSeq (GAIIx)	Discovered major shift in the *Pst* populations in UK.	*Pst*-78/66 *Pst* -12/83 *Pst* -12/86 *Pst* -11/128 *Pst* -11/13 *Pst* -11/08

3.2 Objectives

- ✓ To identify and develop SNP genetic markers that will differentiate between avirulent and virulent strains.
- ✓ To analyse genetic relationship between historical and new isolates.
- ✓ To analyse effector sequence properties

3.3 Methods

3.3.1 Genes of Interest

Genes of interest were based on general and known effector proteins (Saunders *et al.*, 2012) and a selection crieteria specific to *Pst* pathogen (Cantu *et al.*, (2013).

For sequencing with SureSelect and Illumina, with genes based on known effector features (see introduction section 1.2.8) (Cantu *et al., 2011,* Saunders *et al.,* 2012, Cantu *et al.,* 2013, Garnica *et al.,* 2013, Hubbard *et al.,* 2015) characteristics included:

- Similarity to known haustorial proteins, fungal Avirulence
- Presence/absence of effector motifs and Nuclear Localisation Signals (NLS)
- Presence of signal peptides
- Expression levels during infection
- Genome architecture
- Cysteine content

Pst specific criteria included:

- Sequence polymorphism
- Expression during infection
- Expression in haustoria

Genes that were ranked highly as candidate effectors by Cantu *et al.,* (2013) were chosen for this study. For targeted sequencing, targeted sequencing primers, also known as amplification primers, were custom designed by Illumina Technologies. Sequences of potential effector proteins selected as reference genes were taken from Cantu *et al.* (2013). This paper defined and classified a number or potential effector genes from five sequenced isolates, which followed the Saunders *et al.* (2013) pipeline. Potential effector protein sequences were amplified within the selected isolates, without sequencing the whole genome, which as a result would reduce coverage level, accentuating the importance of the 'targeted approach' sequencing.

We used highly ranked tribes to select genes with the above properties. These particularly included those genes that were highly expressed and enriched in the haustoria in infected tissue, and sharing sequence polymorphisms between *Pst* strains.For example, sequence polymorphism was seen in two UK isolates with virulences YrRob and YrSol; this as a result may help detect virulence/avirulence effector proteins that are recognised by YrRob and YrSol. In total 146 genes were selected from 1,088 potential effector candidates for targeted sequencing, we looked at predicted secreted proteins, and selected only the ones which contained RXLR motif variants within the first 100 N-terminal amino acids after the predicted cleavage site. Expression data from Cantu *et al.,* (2013) mRNA-seq analysis was then analysed to select genes expressed in infected tissue and in haustoria. Haustorial expressed secreted proteins

(HESPs) in the top 1%, 5%, and 10% expressed in haustoria were selected. Their characterized fungal avirulence is high, as well as their levels of sequence polymorphisms between isolate.

Table 4-4 (Section 4.4.5) lists the 146 genes used to design custom made probes for targeted sequencing. Design of probes only done for genes meeting the fungal effector protein criteria. All genes included in this study come from either Cluster VII or Cluster VIII (based on Cantu *et al.* (2013) hierarchal clustering), which focused on genes expressed during infection and proteins with expression in haustoria. Tribes with a large number of proteins with effector motifs and proteins with score for similarity of tribe members to HESP or characterised fungal AVRs were also included. Three genes (Beta tubulin, Elongation factor, Mitogen Activated Protein Kinase (MAPK) were also added as controls.

Applied tools and settings for bioinformatics analyses: The following section includes the computational tools and programmes used to analyse the data from the Illumina MiSeq (See Fig 3-1).

3.3.2 Sequencing Library Preparation

DNA was extracted from the eight isolates using the same method (DNeasy plant mini DNA extraction kit (QIAgen) and whole genome amplification) mentioned in Chapter 2, section 2.2.4. DNA was prepared in the laboratory using the SureSelect^{QXT} library preparation and capture system. This protocol was used for optimal Illumina paired-end sequencing, specifically optimized to enrich targeted regions of the genome. Further details can be found in the 'SureSelect^{QXT} Target Enrichment System for Illumina Multiplexed Sequencing' protocol (Agilent Technologies, 2015).

3.3.3 Paired end Sequencing

Illumina Miseq Sequencing produces both single-end reads and paired-end reads. Single-end only reads sequences at one end of a DNA fragment but paired-end sequencing is performed to produce a DNA fragment that is read from both ends and therefore generates a pair of reads (Illumina, 2016). In this study 'read 1' and read 2' was produced for each sequenced read, making paired-end DNA sequencing the chosen platform. The extra positional information obtained from these paired reads, increases the precision of mapping short reads to reference genes, as well as generating high quality, alignable sequence data.

3.3.4 FASTQ

The main output of the Illumina DNA Miseq are the millions of read sequences processed in a FASTQ format (file extension .fastq or .fq) (Cock *et al.*, 2010). FASTQ files are easily transferred and provide a numeric quality score associated with each nucleotide in a sequence. The FASTQ file formats normally have four lines per sequence:

- Line 1 '@' with optional sequence identifier
- Line 2 The sequence
- Line 3 '+' with optional sequence identifier
- Line 4 Quality score for sequence in line 2

An example of a FASTQ file would be:

```
@EAS54_6_R1_2_1_413_324
CCCTTCTTGTCTTCAGCGTTTCTCC
+
;;3;;;;;;@@;;;;7;;;;;;;88
```

PHRED quality scores are assigned to each nucleotide in a sequence (line 4 of a FASTQ file). The PHRED score is important in characterising the quality of the DNA sequence as each character encodes a specific quality score (Cock *et al.*, 2010). More commonly a PHRED score of 20 or above is assigned to the bases in a sequence, as this will offer a base accuracy of 99% (Ewing *et al.*, 1998).

Table 3-3 PHRED quality scores are logarithmically linked to error probabilities. A quality score of 30 & above was used, so the probability that a corresponding base call is incorrect will be 99.99%.

PHRED Quality Score	Probability base call is incorrect	Base call accuracy
10	1 in 10	90%
20	1in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%

3.3.5 Quality control

It is important to remove low quality reads by trimming them so data with poor quality bases are not included in the mapping analyses or for calling SNPs. The basic principle of trimming allows only the high quality bases to be included in the analyses by trimming the 3' end of a read (Del Fabbro *et al.*, 2013).

Sickle Master (https://github.com/najoshi/sickle) was the programme used to trim reads based on the quality score. Parameters with a maximum threshold quality score of 30 and length of 50bp were applied. Sickle master kept into consideration the paired-end reading, as a low quality base in 'read 1' was removed, the corresponding read ('read 2') was also removed. The process of trimming produces new files of trimmed sequences and reduces the need for downstream sequence analysis.

The command used for the trimming process is shown below and the explanations of the options included in the command are available from https://github.com/najoshi/sickle.

The command for trimming is listed below:

- Sickle pe -f input_file1.fastq -r input_file2.fastq -t sanger \ -o trimmed_output_file1.fastq -p trimmed_output_file2.fastq \ -s trimmed_singles_file.fastq

- ./sickle pe -t sanger -f 3_S3_L001_R1_001.fastq -r 3_S3_L001_R2_001.fastq -o 3_S3_R1trimmedsickle.fastq -p 3_s3_R2trimmedsickle.fastq -q 30 -l 50 -s S3trimmedsicklesingle.fastq

3.3.6 Alignment- aligning reads to the reference genes

Aligning the trimmed sequenced reads to the reference genome is significantly important for analysing genetic variation in sequencing data. This is because gaps and mismatches in the alignment will disclose the SNPs and indels within the sequences.

In order to align millions of short reads in a fast efficient manner, the standard BLAST algorithms cannot be used for this purpose, as they cannot support the vast amount of data generated from the Illumina Miseq. For this study, Burrows-Wheeler Alignment (BWA) was used as the ultrafast alignment tool. The algorithms in this programme allow ultrafast alignment speed (Li and Durbin, 2009).

After installing the chosen alignment programme, the reference genes were first to be 'indexed'. Indexing optimizes the speed at which the reads are aligned and produces a secondary file with an extension 'ebwt'.

- *The command used to index the reference genes is as follows:*
  ./bwa index ~ /yRust_References.fasta

Once indexing is complete, it is very important to 'sort' the reads. Sorting the reads will ensure the order of reads is correct.

- *The script used for aligning and sorting the trimmed reads in BWA is as follows:*

./bwa mem -M -t2 ~ /yRust_References.fasta 3-15-8_R1trimmedsickle.fastq 3-15-8_R2trimmessickle.fastq | ./samtools view -Sb - | ./samtools sort - 3-15-8_yrust.sorted

- ✓ Command in green – command which converts the trimmed FASTQ files to the SAM file, producing the ouput file '.sam'
- ✓ Command in blue – command which converts the SAM file to a BAM file, producing the ouput file a '.bam'
- ✓ Command in orange- command which 'sorts' the BAM file, producing the ouput file '.sorted.bam'

The pipe symbol '|' redirects one output into another output within the terminal. This allows programmes such as SAMtools and BWA to work simultaneously, allowing the output files to be transferred between them.

The outputs from the read aligner BWA are stored as BAM and SAM files as briefly discussed above. BAM is the binary form of Sequence Alignment Map (SAM). It is common to check how many alignments (reads) are present in the file. This is possible to do in the programme SAMtools, using the 'view' option. IDXSTATS is also an option available in SAMtools to see the mapped reads.

3.3.7 Filtering

The sequenced reads need to be filtered further to ensure confidence with them. Removing any secondary alignments using the FLAG option -F256 in SAMtools will ensure only primary alignments are included. The FLAG option –f2 will ensure unpaired reads are removed from the alignments, only leaving the sequences. Examples of these commands are listed below.

- *Command for removing secondary alignments:*

  ./samtools view -b -F256 3-15-8_yrustproperpair.bam > 3-15-8_noprimaryalign.bam

- *Command for ensuring reads mapped in 'proper pairs':*

  ./samtools view -b -f2 3-15-8_yrust.sorted.bam > 3-15-8_yrustproperpair.bam

As the reads have been filtered and aligned, the BAM file needs to be indexed to produce a secondary file with an extension of .bai. The indexed BAM file acts as a 'table of contents', so when it is run by the different programmes such as SAMtools or BCFtools, it won't have to read through all the sequences, instead it can go straight to the sequence required.

- *Below is the command used for indexing:*

  SAMtools index Warrior_S8_trimmed.sorted.bam. (This command produces the file: '2_S2_1902_trimmed.sorted.bam.bai)

3.3.8 Variant calling

Variant calling is the process used to analyse the genetic variation in sequenced data. SNPs and deletions (known as indels) in the sequence are changes identified in the variant calling process. SNPs are largely the easiest to ascertain genetic variation, thus the most important markers for such analysis.

Once the alignment has been generated into a BAM file, SNP calling can be done using SAMtools to produce a binary variant call format (BCF) file, followed by BCFtools to generate the variant call format (VCF) file. VCF files are the emerging standard for storing variant data and BCF files are the binary form, similar to the SAM and BAM format.

A VCF file is a text file format, which starts with meta-information lines that include all the data from the mandatory column 'INFO'. Seven other compulsory columns included in the VCF file follow this, and all header lines describe the genetic variants:

1. #CHROM (chromosome)- Identifier for the reference genes.

2. POS (position)- The position of the reference gene. The first base in the variant comes first as VCF files store positions numerically in order of increasing position, within each reference sequence CHROM. It is possible to have multiple entries of the same POS.

3. ID (identifier)- A semicolon-separated list of marker identifiers

4. REF (reference allele)– Is expressed as a sequence of A, G, C, or T nucleotide. Single or multiple bases are permitted (e.g. 'T' or 'TTA'). The value in the POS refers to the position of the first base in the string.

5. ALT (alternate allele)– Is expressed as a sequence of A, G, C, or T nucleotides or angle-bracketed ID String ("<ID>"). If there is more than one non-reference, alternate allele they will be comma separated. If there are no alternative alleles, then the missing value should be used.

6. QUAL (quality)– Phred- called quality score ($-10\log_{10}$(probability)) for the assertion made in ALT. High quality scores relate to high confidence in SNP calling.

7.FILTER – A filter status of 'PASS' refers to a call being made at that position, or a semicolon-separated list of codes for filters that fail due to low quality score (less than 10%).

8. INFO- Additional information should be specified in the meta-information as described above. Examples of the sort of information included in this column are the depth of the read (DP), the total number of alleles in called genotypes (AN), and strand bias for a particular position (SB).

The commands used to call the SNPs are as follows:

- ./samtools mpileup -g -f sequence.dna 2_S2_1902_trimmed.sorted.bam > 2_S2_1902_trimmed2.bcf

- ./bcftools-1.1/bcftools call -c -v 2_S2_1902_trimmed2.bcf > 2_S2_1902_trimmed2.vcf

3.3.8.1 Developing Phylogenetic Tree

The phylogenetic tree was done using *FreeTree* software using UPGMA (Pavlicek *et al.,* 1999). The distance matrix was calculated using the Nei and Li distance (Nei and Li, 1979). Resampling was done by boostrapping with 1000 replicates. The phylogenetic tree derived from *FreeTree* was viewed using *TreeView* and *MEGA 6.0.6.*

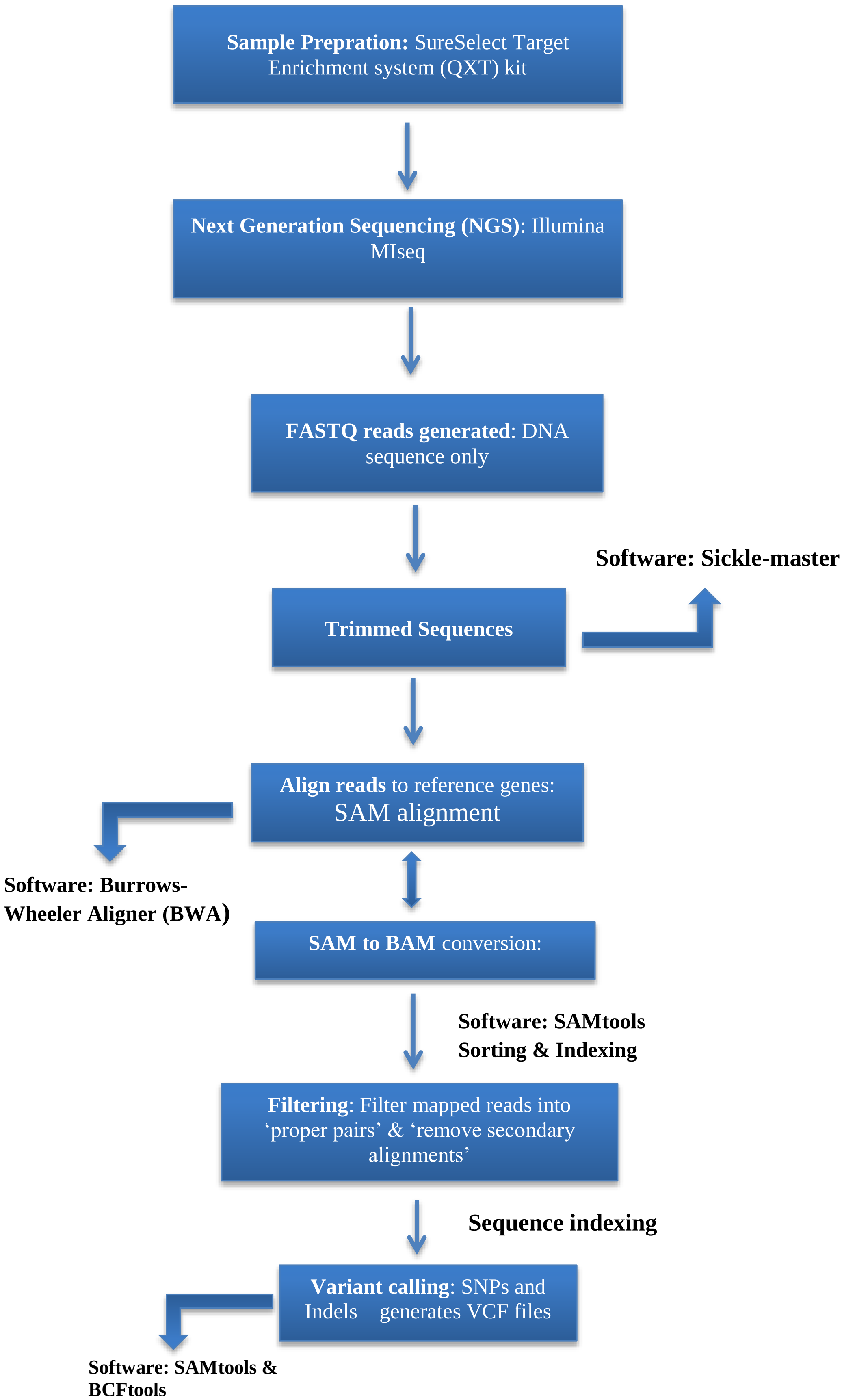


Figure 3-1 Flow diagram displaying the bioinformatics workflow.

3.4 Results

3.4.1 Quality Control

Once the data was obtained from the sequencer it was important to check the read quality of sequences. Quality control is very important in bioinformatics analyses, thus looking at the range of quality values across the bases before initiating the aligning and analysis process is essential. Quality scores across all bases were analysed for the raw data and trimmed data (this displayed the difference in quality of bases once they were trimmed). Here, the term *'trimming'* refers to trimming based on low quality scores while preserving the longest high quality sequence reads.

Figure 3-2 presents the quality scores for raw data and Figure 3-3 displays the data once the sequences were trimmed. For each box whisker plot, the yellow boxes are representative of the inter-quartile range (25-75%), the red line within each yellow box is the median value, the upper and low plots represent the 10% and 90% points, and the blue going across the box plots is representative of the mean quality. The quality scores are on the y-axis (refer to PHREDs quality score in methodology section) and the graph is divided into green (very good quality base calls), light orange (reasonably good quality base calls), and red (poor quality base calls) (see Figs 3-2 & 3-3).

Miseq raw data had a total of 2047921 sequences, of which the majority of the reads were very good quality or reasonably good quality (Fig. 3.2). The raw data also highlights some poor quality reads, which were eliminated once the reads were trimmed using Sickle master; this ensured only high quality reads were included. The process of trimming reads needs to keep a good balance between the increased quality of the score and reducing the quantity of reads, to maintain mapping efficiency. As both graphs display no poor quality reads (Figs 3-2 & 3-3), it can be concluded that this study included overall very good quality reads. It should also be mentioned that the quality of reads degrade as the run progresses, and this is visible on both graphs where the base calls are falling towards the orange/red area at the end of a read.

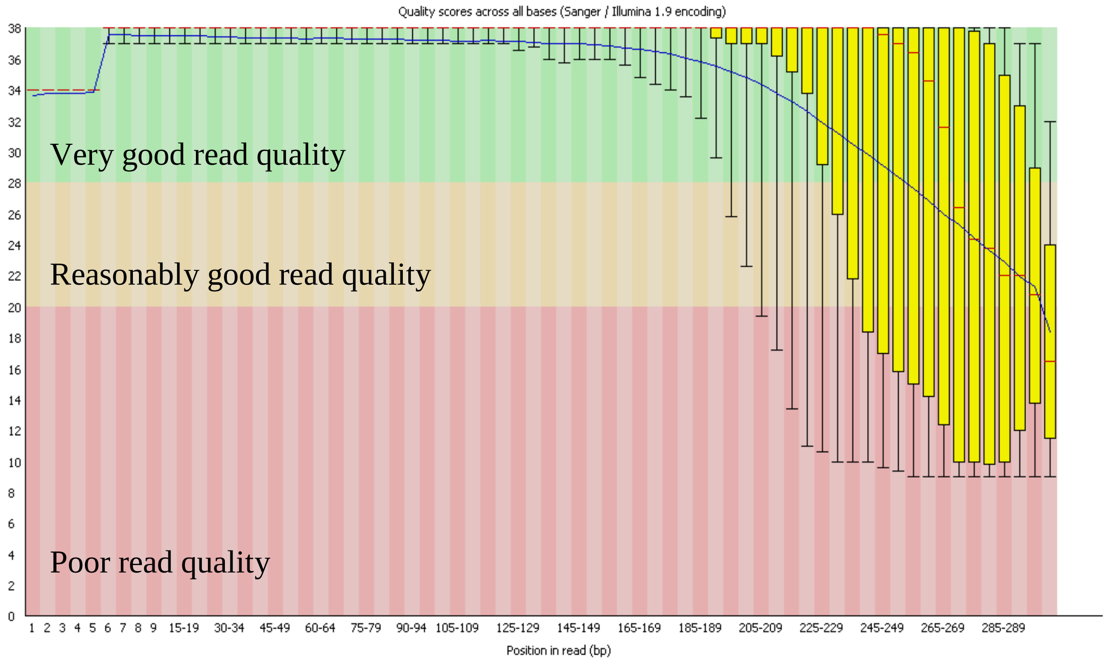


Figure 3-2 Quality control analysis of raw data from all samples using FastQC. A box-and-whisper plot is plotted for each read position. Total Sequences 2047921, sequences flagged as poor quality 0, Sequence length 301, %GC content 44.

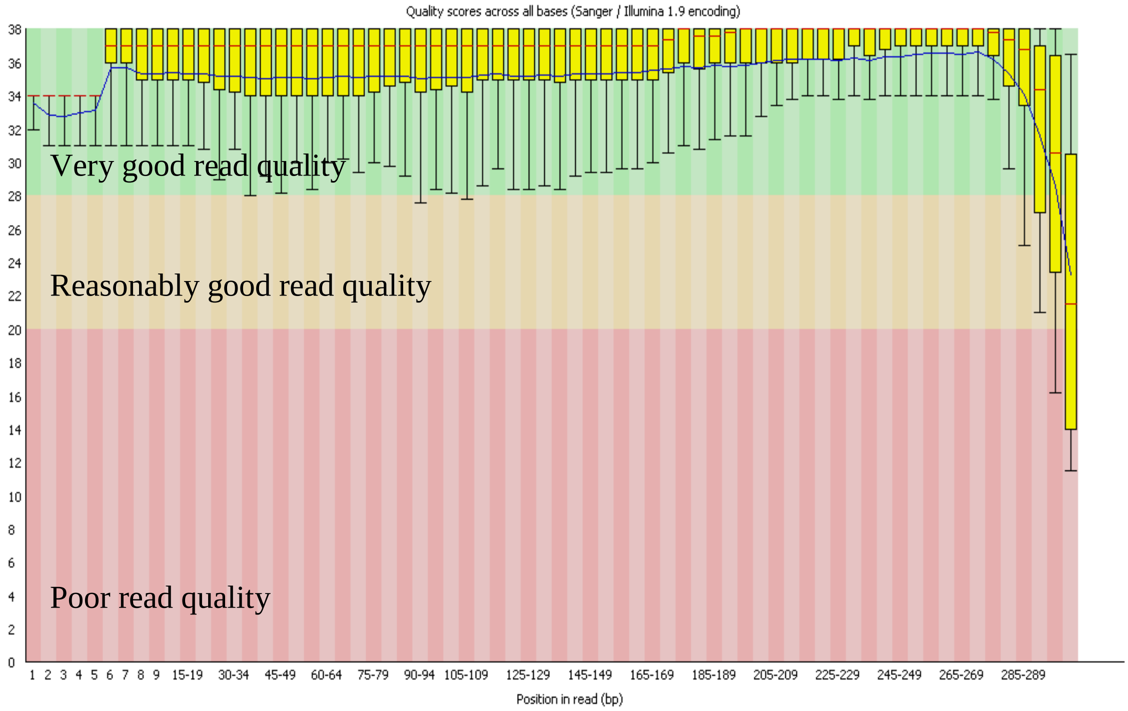


Figure 3-3 Quality control analysis of trimmed data from all samples using FastQC. A box-and-whisper plot is plotted for each read position. Total Sequences 70706, sequences flagged as poor quality 0, Sequence length 50-301, %GC content 47.

3.4.2 Filtering and Mapping Reads

After trimming, high quality sequence reads were maintained by filtering reads into proper pairs and by removing secondary alignments. The number of reads ranged from 2801496 to 5509482 base pairs; the raw and trimmed reads are summarized in Table 3.4. To assess genetic diversity between the *Pst* isolates the sequence reads of a particular isolate were aligned to reference genes.

Low quality calls contribute towards a higher number of unaligned reads, reducing the mapping efficiency as a result. In furtherance of good quality reads, reads were filtered to remove low quality sequences and consequently to reduce sequencing errors. This process ensured only primary alignments were included and unpaired reads were removed, again eliminating sequences with poor quality reads. Choosing specific thresholds to maintain a balance between the quality and quantity for data efficiency requires trimming and filtering of reads. As expected, the number of reads reduced considerably but it also improved the overall quality of the reads, for example, in terms of false positive SNP calls.

To characterise the genetic diversity of *Pst* isolates, 8 different pathotypes were sequenced using the Illumina Miseq SureSelect technology, allowing for efficient enrichment of regions of interest. After trimming and filtering sequenced reads, the obtained reads were mapped on to reference genes to deduce the genetic structure of any given sample. We identified that a larger proportion of the sequenced reads aligned to the reference genes; an average of 80% reads aligned across the isolates (Table 3-4). This indicated the vast genetic similarities to the reference genes, and, by extension, to each other. Mapping isolates on to the reference genes containing sequences of potential fungal effector proteins, suggests that the fungal pathotypes must share genetically similar effector protein sequences to candidate effectors of high interest.

An obvious aligning strategy would be to align the isolates to the previously published *Pst*-130 genome. However, this strategy would only be applicable if whole genome sequencing was used (Cantu *et al.,* 2011). Focusing on targeted regions that include potential effector sequences, promotes the chances of characterising fungal effector proteins with high quality sequence alignment to fungal sequences.

Table 3-4 Alignment of reads to reference genes using the BWA programme, after filtering reads to remove secondary alignments (The correct placement of a read may be ambiguous, e.g. due to repeats. In this case, there may be multiple read alignments for the same read. One of these alignments is considered primary. All the other alignments have the secondary alignment and filtered to ensure reads are mapped in proper pairs (the sequence quality in both directions). Alignment to reference genes was an average of 83% or more across the isolates, providing confidence in the analyses.

Sample	Number of reads before mapping (raw reads)	Reads mapped in proper pairs	Removing secondary alignments	Alignment to reference genes (%)
234E139	5312503	4597841	4536784	**86.85**
169E136	3880372	2970971	2943010	**77.12**
237E141	3925357	2831188	2801496	**72.54**
109E9	5317299	3949672	3922052	**74.70**
94/519	464	402	402	**88.15**
KWS Sterling	6397743	5509482	5424160	**86.98**
Solstice	3970289	3414885	3361898	**86.99**
Warrior	4743500	4453974	4361158	**93.86**
Average				**80%**

To avoid between sample variation and reduce the number of false positive and false negative reads for each gene, the median per sample was calculated. The median value of mapped reads was used as a cut off for gene presence and divergence/absence. Using this, the gene presence and divergence data was calculated. Sequence presence and divergence data is represented as heat maps and hierarchal clustering (Fig 3-4). The phylogenetic tree was done using *FreeTree* software using UPGMA (Pavlicek et al., 1999). The distance matrix was calculated using the Nei and Li distance (Nei and Li, 1979). Resampling was done by boostrapping with 1000 replicates. The phylogenetic tree derived from *FreeTree* was viewed using *TreeView* and *MEGA 6.0.6* (Fig. 3-7).

Appendix I present the presence and divergence of genes and this information is correlated in table 3.5. As mentioned previously in Chapter 2, isolate 94/519 has proven to be an anomaly in the dataset due to the very low read counts and thus absence of most genes was seen within this sample (See Appendix I). As a control, three house-keeping genes were included in the study; beta tubulin, elongation factor and mitogen activated protein kinase (MAPK), all of which were present in all isolates with large read numbers, reinstating confidence in results of mapping sequence reads. Table 3-5 highlights the presence and absence/divergence of genes. This table represents that most genes were absent in isolates 169E136 and 109E9. Table 3-5 displays 16 of 146 genes that were specifically absent amongst the particular UK isolates.

Targeted amplification with selected genes thought to be potential effector protein candidates, showed that some genes, such as *Pst*43_04767, were absent in all isolates used within this study. Although gene *Pst*43_04767 was present in *Pst* isolates taken from studies Cantu *et al.* (2011; 2013), lack of amplification of this specific gene in all eight isolates appears to eliminate *Pst*43_04767 as a yellow rust effector protein in UK isolates.

Table 3-5 The presence and absence/divergence of genes for potential secreted effector proteins within the 8 samples. Most of the listed genes are absent in sample 109E9 and 169E136.
**All genes are absent in isolate 94/519 but have not been included in the table.*

Gene Name	Gene Divergence/Absence
*Pst*877_19843	Absent in all samples (except Solstice and Warrior)
*Pst*_14831	Absent in 234E139, 169E136, 109E9, KWS, Solstice
*Pst*43_04767	Absent in all isolates
*Pst*877_00017	Absent in KWS
*Pst*130_02589	Absent in 234E139, 169E136, 109E9,
*Pst*43_03249	Absent in 109E9
*Pst*130_038899	Absent in KWS and 109E9
*Pst*43_08388	Absent in 169E136 and 109E9
*Pst*0821_09721	Absent in 109E9
*Pst*877_15982	Absent in 169E136
*Pst*21_17484	Absent in 109E9
*Pst*0821_09979	Absent in 234E139 and 169E136
*Pst*0821_16435	Absent in 109E9
*Pst*21_18223	Absent in 234E139 and 109E9
*Pst*130_06558	Absent in 169E136 and KWS
*Pst*21_08753	Absent in 169E136 and 109E9

3.4.3 Variant Calling

SNP calling is also referred to as ‘variant calling’, the process whereby all the polymorphic positions are ‘called’ and identified. It was important to undergo the above processes proficiently (base calling denoted at good PHRED scores, alignment, and filtering) before SNP calling took place. This is because SNPs are important molecular markers used in this thesis that will allow us to distinguish between different *Pst* isolates. Therefore, as mentioned above it was imperative to remove the possibility of any false positive SNPs through the filtering process.

3.4.3.1 Estimation of diversity between Pst isolates

In order to be able to identify genetic variation amongst the different isolates and to distinguish the virulent and avirulent markers within them, it was important to assess the SNP data. SNP data was generated using the bioinformatics workflow where they were identified and filtered using SAMtools and BCFtools software (Fig. 3.1). As expected, all eight isolates showed a variable number of SNPs within the genes (Table 3.6).

Bioinformatics analyses revealed that gene *Pst*21_16369 displayed the most variation with SNPs ranging from 335 to 652 across the isolates, with the exception of isolate 94/519, which displays zero SNPs within this particular example, confirming it as an anomaly due to the low DNA quality (Table 3-6). On the other hand, *Pst*130_06558 showed very low variation with one SNP per isolate, similarly to genes *Pst*0821_06060, *Pst*43_01402, *Pst*0821_06581. It is apparent that the SNPs are not evenly spread throughout the eight genomes, suggesting that the genes are highly variable. Genes *Pst*0821_05302 and *Pst*130_00285 are examples of genes, which exhibit lower variability due to the reduced number of SNPs present (Table 3-6). The number of SNPs displayed in Table 3-6 denotes to the actual number of SNPs identified within that gene, with '0'0 indicative of no real SNP identified. The heat map below is a good representation of the variation identified in the reference genes (Fig. 3-4).

Table 3-6 SNP frequencies for 8 *Pst* Isolates. Numbers are representative of the number of SNPs present within each isolate. For example, gene *Pst*21_16369 entails 335 SNPs in isolate 234E139. Includes examples of full set of genes that exhibit high variability and genes that are least variable. The right column indicates whether the number of SNPs were high, medium or low.

Gene Name	234E139	169E136	237E141	109E9	94/519	KWS	Solstice	Warrior	
***Pst*21_16369**	335	508	580	652	0	522	521	523	High
***Pst*21_10634**	442	372	307	97	0	459	226	253	
***Pst*130_07695**	61	52	49	5	0	34	71	54	Medium
***Pst*130_00860**	22	54	53	49	53	46	46	54	
***Pst*130_00285**	8	22	21	24	0	22	18	13	Low
***Pst*0821_05302**	1	2	1	46	0	4	1	1	

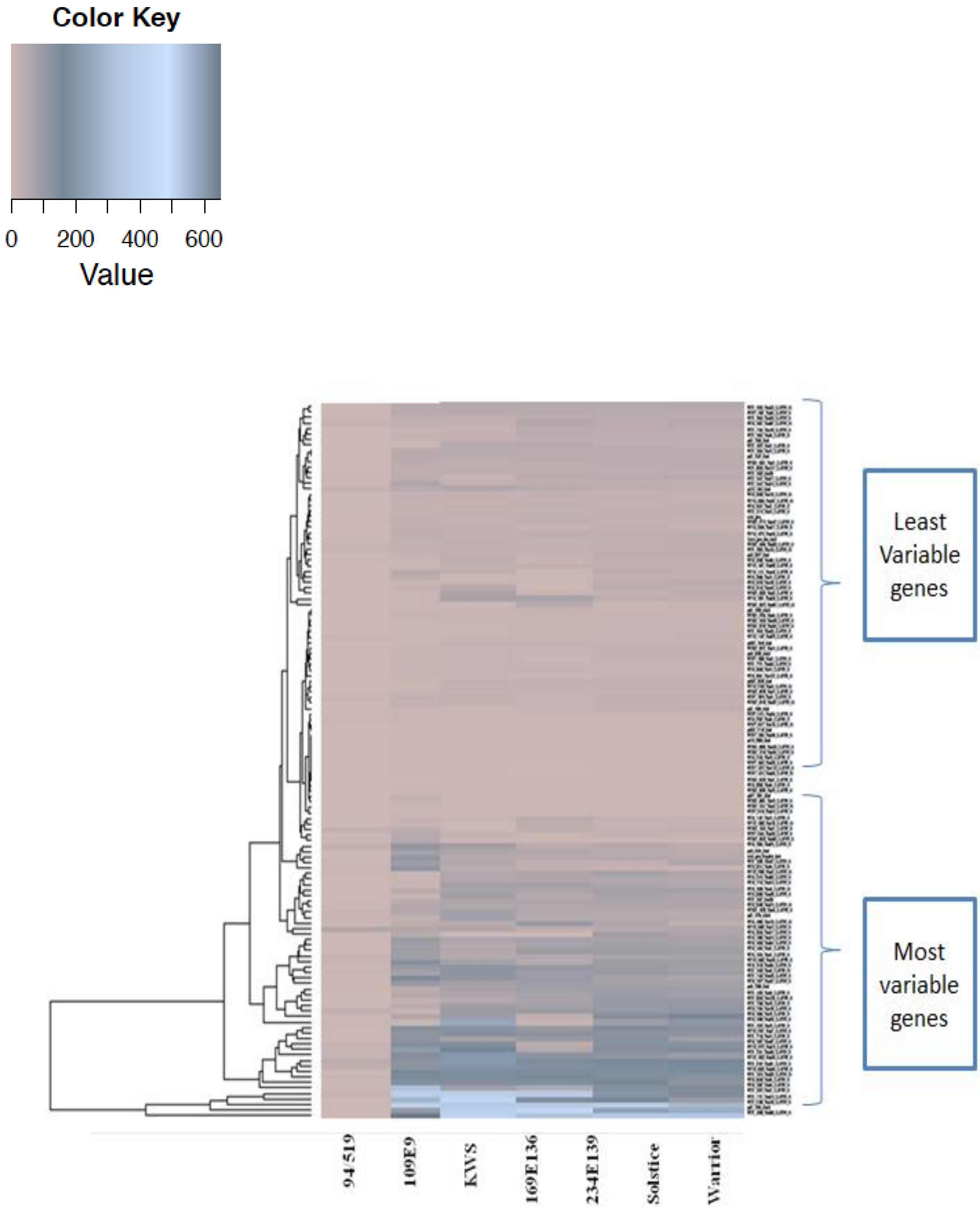


Figure 3-4 Heat map presents the SNP frequency present within each sample. Colour key highlights that pink shading is representative of presence low SNP numbers (approximately 0-200), light shading is representative of 200-400 SNPs within a gene, and dark shading is illustrative of approximately 400-600 SNPs. This correlates with the list of reference genes on the far right, as genes towards the bottom of the graph are characteristic of highly variable genes, compared to genes further to the top, which are the least variable. The generated heat map is a clear representation of which isolates entail the most variable genes. Isolate 94/519 appears to show the least variation, whilst Solstice and Warrior share high variation.

A heat map was constructed (using *R studio, version 3.3.1*) showing the SNPs generated in the study, this enabled a visual analysis of the genetic diversity between the different isolates. The light and dark shades of blue are representative of 200 to 600 SNPs respectively. The darker shades confer to higher genetic variation and the lighter shades represent SNPs with no or low genetic diversity. The reference genes are displayed on the far right of the heat map, and the isolates are listed at the bottom. *Pst*0821_05302 displays lower SNP variation as described in Table 3.6, whereas *Pst*21_16369 shows higher genetic variation, and is thus positioned towards the bottom of the graph. Phylogenetic analysis on the far left of the heat map is displaying the relationship between the reference genes. Reference genes have been clustered to provide an insight into their effector properties. Highly variable genes are clustered together, and this cluster includes genes within tribe 238, 79, and 351 which include genes expressed in the haustoria, and tribes 43 and 9 which are similar to known haustorial proteins (Cantu *et al.,* 2011, Cantu *et al.,* 2013).

Using BCFtools 'isec' option it was possible to identify the common SNPs present in all the isolates, and by using the BCFtools merge option, SNPs present in at least 1 sample among the 8 isolates could be found. These were valuable tools to categorise any distinct SNPs, subsequently placing them in to clusters to classify the SNPs that were either common or unique amongst the genes mentioned above. A Venn diagram was generated using Venny 2.1, presenting the percentage of overlapped and non-overlapped SNP sites that were identified against the new isolates. With *Pst*21_16369 and *Pst*21_10634 being two of the more variable genes, a pattern is seen, as KWS, represented in the blue circle, contains the largest number of SNPs within both genes followed by Solstice and Warrior respectively. It is apparent that SNPs are largely common amongst all three isolates, indicating the close genetic relationship.

The Venn diagram also reveals that within *Pst*21_16369, Solstice and Warrior share more SNPs than in *Pst*21_10634. Analysing SNP markers within these samples provides a better understanding of their genetic differences. SNPs unique to one particular isolate such as KWS, differ from what SNPs are present in Warrior for example.

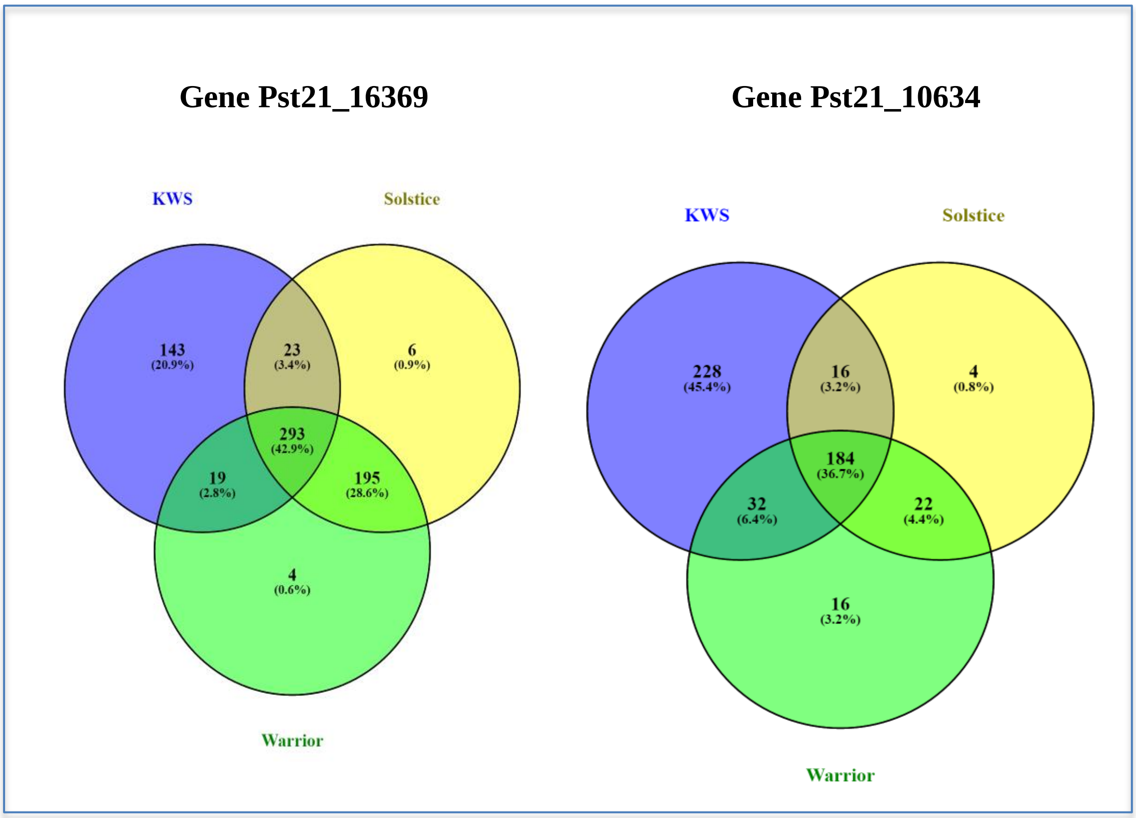


Figure 3-5 Venn Diagrams comparing two highly variable genes within samples KWS, Solstice and Warrior. A common pattern seen amongst the samples is that, KWS contains the most SNPs within both genes, and the least amount of SNPs are unique to Solstice and Warrior. Solstice and Warrior share the most SNP positions at approximately 29% in gene *Pst*21_16369, but share a much lower amount in gene *Pst*21_10634 (4.4%).

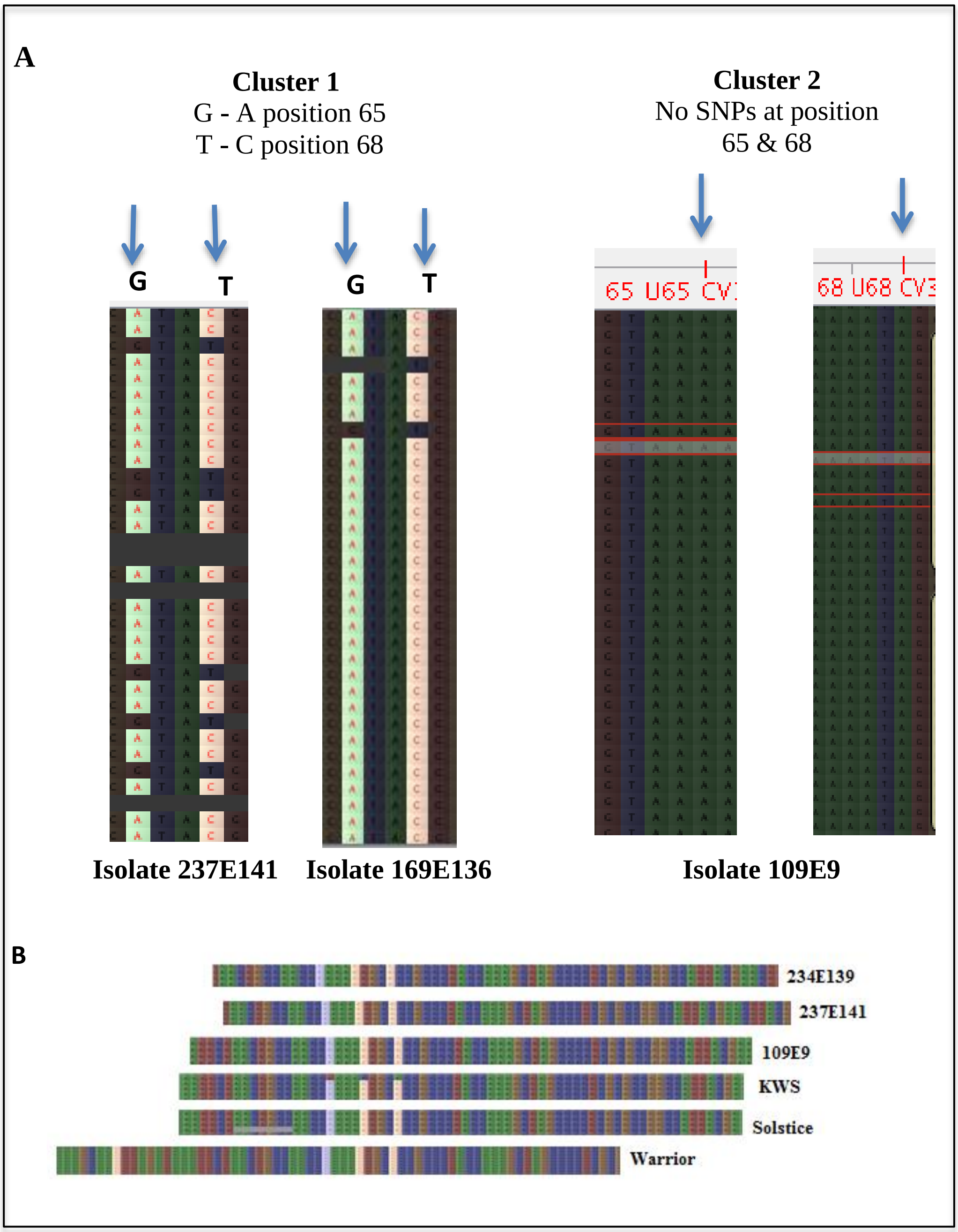


Figure 3-6 **(A)** Representation of SNP discrimination between isolates in different clusters in gene *Pst*0821_00728. Following phylogenetic analysis, isolates were clustered and it became apparent that SNPs in cluster 1 are unique to SNPs in cluster 2. **(B)** Example of sequence polymorphism in SNP positions 447, 451, and 455 in gene *Pst*21_1564. Sequence polymorphism is present in all isolates, but absent in samples 169E136 and 94/519, representing the genetic differences between samples.

To further characterise the genetic relationship between the sequenced isolates, the number of SNPs present distinctly within each isolate was recorded. Analyses from the VCF file provides the identified markers in all isolates within gene *Pst*21_10634. The sequence length of this particular gene was 4743 bp, which gives a better understanding of its variable nature. The number of SNPs unique to isolate 234E139 was the greatest followed by 169E136 and 237E141 respectively. SNP sites obtained from the VCF file correlates with the Venn diagram above (Fig. 3-5).

To determine the relationship between previously prevalent isolates and new isolates phylogenetic analyses was undertaken (Fig. 3-4). The reference genes have been clustered and focusing on the most variable genes in the bottom largest clade gave an insight into the close lineage between genes *Pst*21, *Pst*130, *Pst*43. This clade is absent of any *Pst*_0821 and *Pst*_877 genes. The sequence lengths of all these genes are large and contain the largest numbers of SNPs. The genes in this particular clade are present amongst all isolates. Genes that clustered together in Figure 3.4 contain specific effector properties as show in previous research (Cantu *et al.*, 2011; 2013). For example, gene Pst43_10686 has been clustered into tribe 238 which associates the members of this cluster with genes expressed during infection and thus in the haustoria, a specific *Pst* effector trait.

A

	234E139	169E136	237E141	109E9	95_519	KWS	Solstice	Warrior
234E139		0.62222	0.76829	0.62222	0.47423	0.62222	0.69767	0.65909
169E136	0.62222		0.68605	0.58696	0.50526	0.60440	0.65909	0.65909
237E141	0.76829	0.68605		0.64773	0.49474	0.68605	0.81250	0.76829
109E9	0.62222	0.58696	0.64773		0.45918	0.60440	0.65909	0.67816
95_519	0.47423	0.50526	0.49474	0.45918		0.64368	0.58889	0.55435
KWS	0.62222	0.60440	0.68605	0.60440	0.64368		0.84810	0.75904
Solstice	0.69767	0.65909	0.81250	0.65909	0.58889	0.84810		0.87179
Warrior	0.65909	0.65909	0.76829	0.67816	0.55435	0.75904	0.87179	

B

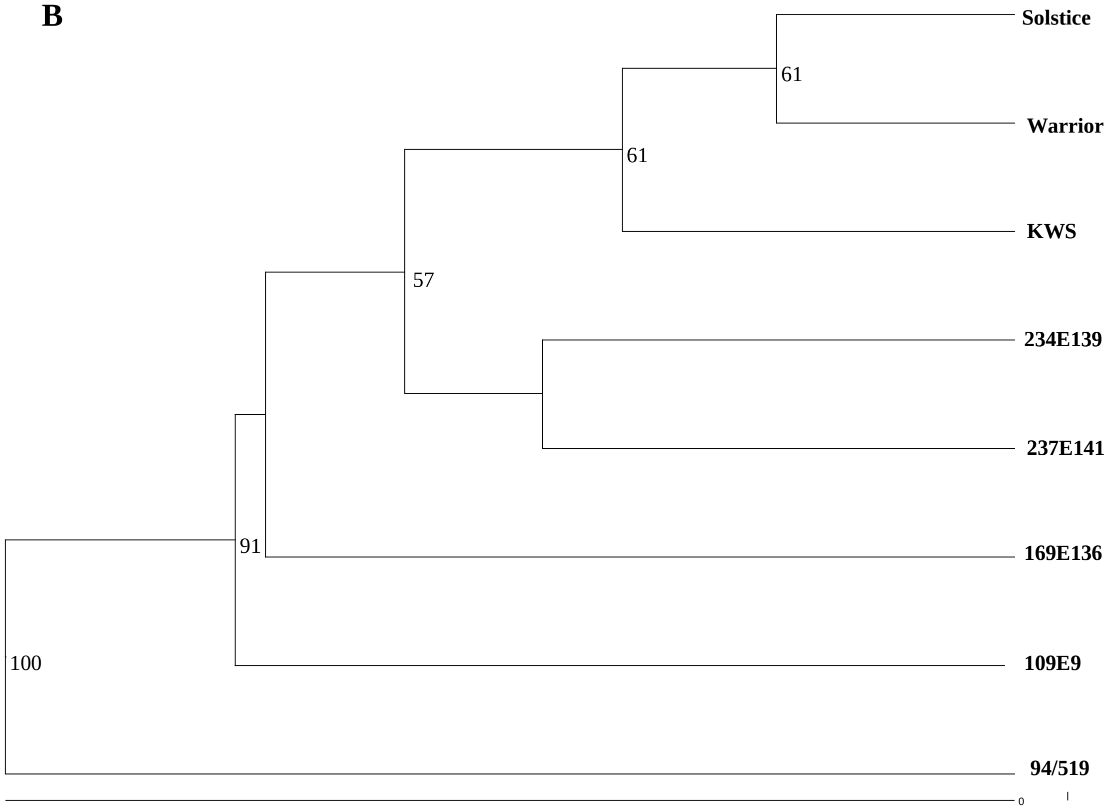


Figure 3-7(A) Similarity matrix values given for each isolate enabling further understanding of the likeness shown in the phylogenetic tree **(B)** Phylogenetic tree clustering closely related isolates. New isolates (KWS Sterling, Solstice, and Warrior) clustered together, whilst the older isolates produce a separate clade. Isolates 234E139 and 237E141 share genetic similarities with the new isolates. (See Section 3.3.8.1).

Phylogenetic analysis reveals that isolate KWS, Solstice and Warrior are distinctly part of one cluster, whilst Warrior and Solstice share more similarities to one another than with KWS Sterling. This correlates with Figure 3-7, as the number of SNPs in Warrior and Solstice are much lower in comparison to KWS Sterling. Isolates 234E139 and 237E141 share genetic similarities with the new isolates, whilst 169E136, 109E9 and 94/519 are within one particular sub-cluster, despite 94/519 being an anomaly.

3.5 Discussion

Traditionally, studying fungal plant pathogens has been limited to phenotypic assessments and molecular marker analyses. This has previously hindered the possibility of identifying new *Pst* pathotypes arising in the population (Park *et al.*, 2011). With new *Pst* strains arising more commonly due to migration, recombination and mutations it has become increasingly important for surveillance methods to be modernised. For example, traditional methods would make it difficult to distinguish between an exotic pathotype sharing phenotypic characteristics with an existing population, and more specifically lacking the inability to explore any genetic relationships amongst the population subset. Therefore SNP analyses from NGS technologies offer more discrimination and are becoming popular for distinguishing genetic differences amongst pathogens.

In this study eight different *Pst* isolates were studied and targeted regions within the samples were sequenced. These targeted regions included sequences from potential effector proteins. SureSelect^{xt} Target Enrichment System for Illumina paired end sequencing was undertaken at FERA directly from DNA extracted from urediospores. This method gave an insight into the genetic relationships between these rust fungi, also enabling us to characterise potential fungal avirulence genes (See Chapter 4).

The major objective of this chapter was to identify genes that could form the basis for developing genetic markers linked to avirulence genes, and for this purpose we looked at SNP markers.. Sequencing effector regions within the chosen isolates gave an insight of common SNPs shared within the selected areas. Difference in SNPs between *Pst* samples provided clarity on how such molecular markers was predicted to differentiate between closely related isolates, but also shed light on their virulence and avirulence properties.

3.5.1 Mapping Genes

The data generated from the 146 effector candidate genes in this chapter demonstrated that alignment with reference genes was on average 80% across the isolates. This result verifies that effector protein sequences are present amongst the isolates. Furthermore, absence of particular predicted secreted proteins (Table 3-5) within all isolates was limited to only one gene (*Pst*43_04767). Absence of any particular gene was verified by a lower limit cut off based on the median of mapped raw read counts per sample. This particular gene was clustered into tribe 40 by Cantu *et al.* (2013), who identified it as a potential effector protein, which lacks a host-translocation motif (and a nuclear localisation signal (NLS)), which are essential features to mediate entry in to the host cell cytoplasm (Whisson *et al.*, 2007). However, Cantu *et al.* (2013) recognised this gene as a candidate effector protein, based on expression data from RNAseq analysis, which indicated that *Pst*43_04767 was expressed in infected tissue (top 5% and 10%).

For comparative reasons, this research included this particular gene. The absence of this gene in all the selected *Pst* isolates used in this study makes it tempting to speculate that a lack of a motif and its absence, means it has no role in ‘full’ virulence and can be disregarded as a potential effector protein candidate in the isolates tested here. This supports the work of Saunders *et al.* (2012) who used motifs as a criterion for selecting effectors of *M. lini* and *C. fulvum.* With yellow rust being an haustoria forming fungus, Y/F/WXC sequences have been commonly associated with such effectors (Godfrey *et al.,* 2010), and the lack of it makes it difficult to confirm its effector status.

Interestingly, divergence of genes through sequencing highlighted that some potential effector proteins were absent in isolates 169E136 or 109E9, both of which were historical. If comparisons are made to the modern isolates, they display high sequence similarities to particular reference genes. For example *Pst*130_02589 is absent from 234E139, 169E136, and 109E9, but present in all currently prevalent pathotypes tested. This finding is in support of findings from Chapter 2, where difference between isolates is probably due to a new introduction since 2011 (Hubbard *et al.,* 2015). Hubbard *et al*. (2015) found a dramatic shift in *Pst* populations in the UK. Through genetic analysis, they concluded that pathotypes from 2013 did not display genetic similarities to most of the isolates collected between 1978 and 2011.

3.5.2 Identifying genetic relationship between Pst isolates

Comparative genomic studies are powerful tools for management of rust diseases. This approach can distinguish between isolates at a molecular level through marker analysis. SNPs were generated from VCF files using mpileup from SAMtools. Heat map (Fig. 3.4) illustrates the genetic diversity between all the different pathotypes, and also clusters the isolates together. As expected, the newer isolates clustered together, while the older isolates shared similarities. It is important to note that bootstrapping values were large enough to present clear similarities between all eight isolates. In conjunction with Chapter 2, where AFLP markers were used to explore genetic differences, similarly in this chapter we identified the relatedness between isolates 169E136 and 237E141 (a bootstrapping value of 87 was previously identified in Chapter 2). Exploring matrix values in Figure 3-7 (A) can further conclude the likeness between old and new isolates. High matrix values of Solstice, Warrior and KWS compare with similar values to 237E141. For example 237E141 and solstice hold a value of 0.81 between them, which is similar to 0.87, a value that compares the likeness between Solstice and Warrior.

In order to further classify the samples, SNPs were analysed within each cluster. This approach permitted unique SNPs to be mined out of each cluster, for example SNPs in cluster 1 were found to be unique to SNPs in cluster 2 for gene *Pst*0821_00728, at positions 65 (G-A) and 68 (T-C). Looking at sequence polymorphisms within clusters was highly informative in addressing any genetic relationship, but more importantly to characterise the yellow rust population by looking at markers such as SNPs.

Focus was placed on genes displaying higher genetic variation such as *Pst*21_16369, than those displaying lower SNP variation, such as *Pst*0821_05302. This illustrates that SNP distribution varied vastly across the genes. Differences in SNP numbers between the three new isolates were apparent, as KWS Sterling would generally comprise of larger number of SNPs in comparison to Solstice or Warrior. Distribution of SNPs was representative through the heat map in Figure 3-4.

Throughout this bioinformatics analyses isolate 94/519 has proved to be an outlier due to the poor sequencing data, producing significantly low mapped read counts and fewer shared similarities with other isolates. Therefore we cannot rely on data for this isolate. The position of this isolate is illustrated in the phylogenetic tree (Fig. 3-7 (B)) and coincides with findings from AFLP analysis, Chapter 2.

3.5.3 Conclusion

This analysis illustrated that four of the historical isolates clustered together in a single clade with little genetic variation. By contrast, the newer isolates, collected between 2008 and 2013 were indistinctly related to the historical isolates. The use of SNP markers enabled the characterisation of yellow rust population, by identifying the genetic variations. By using targeted sequencing primers it permitted us to achieve coverage of 500x-1000x more than what we would have achieved with whole genome sequencing (30x-50x) (Ilumina Technologies, 2016), emphasising the importance and applicability of this method in future plant-pathogen studies. In order to explore sequence polymorphisms through targeted sequencing technology is an easier and effective method than whole genome sequencing. Moreover, for the purpose of this chapter, looking into potential effector protein sequences is more effective through targeted sequencing. It can be concluded that with the high mapping percentage between the 8 isolates and the reference genes (which included the potential effector sequences) it can be strongly suggested, that the selected isolates in this study also contain and express effector proteins, enabling entry into host cells for infection. Future studies should focus on comparative sequence analysis using larger data sets to determine genetic differences between isolates and this data will aid the improvements of better surveillance strategies and improved durable resistance.

Chapter 4

Design and Validation of Potential Avirulence and Virulence Markers

Chapter 4: Design and Validation of Potential Avirulence and Virulence markers

4.1 Introduction

Conventionally, monitoring variation amongst pathotypes of fungi has been through observing the disease phenotype after inoculating differential hosts (with known resistance genes), with pathogenic fungal strains. This has provided insight into relationships between pathogen populations and host resistance, assisting decisions on breeding mechanisms.

Genetic variations amongst pathogen populations are due to many factors such as migration, mutations, and recombination, all of which play a role in shaping the genetic make-up of pathogens and affecting their evolution. With the use of molecular markers it has become easier to identify the genetic changes within these fungal pathogens and thus improve disease management. In order to fully understand and identify molecular markers correlating with avirulence phenotypes, sequenced fungal pathotypes must be easily accessible. In accordance with the gene-for-gene theory, in which interactions between avirulence gene products and resistance genes products determine the disease outcome, focusing on avirulence genes has become more important.

This chapter aimed to focus on identifying markers linked to avirulence genes/phenotypes, and to develop primers that can be used to check for the presence/absence of specific avirulences in *Pst* pathotypes. Ultimately, this objective intends to ease the process of identifying *Pst* pathotypes in the field using DNA assays.

As previously mentioned in chapter 2, DNA based methods have proven to be valuable tools for obtaining efficient and successful results. Early detection of pathogens avoids the introduction new pathogens and spreading the disease in new areas, demonstrating the importance of identifying virulent fungal pathotypes earlier. Using molecular techniques such as PCR allows closely related isolates to be distinguished, in conjunction with detecting and identifying species such as *Pst* that cannot be cultured, to be identified with high specificity.

As seen in Chapter 3, sequencing targeted regions of the genome has made it possible to genetically distinguishing the pathotypes. This gave an insight into their genetic make-up and provides the potential to develop DNA-based diagnostic assays for pathotypes. Focusing on SNPs in candidate effector genes rather than AFLPs enabled a targeted approach to identify SNPs, rather than a random approach as with AFLPs. This targeted approach enabled identifying unique and common SNPs within candidate effector genes that are markers associated with presence/absence of genes in yellow rust isolates.

Moreover, this Chapter focuses on developing primers from 72 candidate effector proteins identified by Cantu et al., (2011). As previously discussed in Chapter 1, Section 1.2.7 translocation of effector proteins into the host cell is important to supress host immune responses. For this translocation an important N-terminal RXLR motif is required, which was identified in 72 candidate *Pst* effectors. These 72 proteins were the only ones from the identified 1,088 potential effectors that contained these related motifs within the first 100 N-terminal amino acids after the predicted cleavage site (Cantu *et al.,* 2011). The hypothesis was tested that the 72 predicted candidate effector genes from Cantu *et al.,* (2011) are candidates for the *Avr* genes in the test *Pst* isolates.

The objective of this chapter is to develop primers based on AFLP markers identified in Chapter 2 and SNP markers identified in Chapter 3. This would ultimately characterise the pathotypes with specific primers enabling the detection of a particular pathotype, or group of isolates comprised of specific avirulence markers, which correlate to their virulence profiles.

4.1.1 Selecting *Pst* isolates

Eight *Pst* isolates were used in this study with different virulence profiles. All isolates were provided in spore form. As established in the earlier chapters, pathotypes 234E139, 169E136, 237E141, 109E9, and 94/519 are historical dating back to 1994 to 1998. Despite these differences, the more recent isolates (KWS Sterling, Solstice, and Warrior) had common virulences (*Yr1, Yr2, Yr3, Yr4, Yr6, Yr7, Yr9, & Yr17*) to eight *Yr* genes.

Isolate 94/519 and 169E136 share virulence to *Yr17*, which is present in all 3 of the recent isolates. Virulence to *Yr3* and *Yr9* is present in all 8 isolates; conversely 234E139 differs in its virulence to *Yr1* in comparison to the other isolates, since 234E139 does not share virulence to this *Yr* gene unlike the other isolates. Distinct differences in virulence profiles are apparent in KWS Sterling, Solstice, and Warrior. If compared to the historical isolates, they differ in their virulence to certain wheat varieties, but all 3 share common virulence to *Yr32*. In contrast, KWS Sterling and Warrior include extra new virulences (*YrRob, YrSol, YrSte*) and (*YrSp, YrTim, YrWar*) respectively (Table 4-1). Studying virulence profiles and highlighting the obvious distinctions will assist in exploring patterns of absence/presence of SNPs within the *Pst* genes.

In Chapter 2, AFLP markers were used to identify markers that were present in different pathotypes of *Pst*. Here, we aim to identify whether any of the specific AFLP markers correlate to virulence/avirulence traits. Table 4-1 is a representation of the virulence profiles for isolates 234E139, 169E136, 237E141, 109E9, 94/519, and Warrior. Virulence profiles refer to the *Yr* genes that are overcome, with the corresponding *Avr* genes absent or in its recessive form.

Table 4-1 Virulence profiles for the studied pathotypes. Virulence on *Yr* genes ‘2’, ‘3’, and ‘9’ is present in all samples and is therefore highlighted as not significant. Empty boxes signify absence of virulence on a *Yr* gene. Five isolates are named using the binary system, this reveals which series (International or European) the pathotype is, as well as providing an understanding of which *Yr* gene is overcome. This, however, is not the case for KWS Sterling, Solstice, and Warrior. Virulence profiles refer to the *Yr* genes that are overcome, with the corresponding *Avr* genes absent or in its recessive form.

Isolate	**Virulence Profile (*Yr* genes overcome/*Avr* gene absent/recessive)**													
234E139		**2**	**3**	**4**		**7**	**9**							
169E136	**1**	**2**	**3**				**9**	**17**						
237E141		**2**	**3**	**4**		**7**	**9**							
109E9	**1**	**2**	**3**	**4**	**6**		**9**							
94/519	**1**	**2**	**3**				**9**	**17**						
Solstice	**1**	**2**	**3**	**4**	**6**	**7**	**9**	**17**	**32**					
KWS	**1**	**2**	**3**	**4**	**6**	**7**	**9**	**17**	**32**	**Rob**	**Sol**	**Ste**		
Warrior	**1**	**2**	**3**	**4**	**6**	**7**	**9**	**17**	**32**	**Rob**	**Sol**	**Sp**	**Tim**	**War**

4.2 Aims and Objective

- ✓ The purpose of this chapter is to identify suitable variants, design and optimise markers obtained from AFLPs (Chapter 2) and candidate gene sequencing (Chapter 3)
- ✓ Develop DNA-based diagnostic assays for detection and discrimination between pathotypes.
- ✓ Link novel SNP markers to any of the *Avrs* identified in the studied pathotypes.

4.3 Method and Materials

4.3.1 Primer design for validating potential SNP markers

The SNP markers that assisted in designing the primers were taken from the bioinformatic analyses (see Chapter 3). The markers were identified on the Tablet programme from the James Hutton Institute (Milne *et al.*, 2012) and SnapGene viewer version 2.1 was used to develop the primers. Distinctive SNP markers were manually highlighted onto the gene sequence, and primers were designed around the variable regions. For comparative reasons, primer design was based on SNPs that may potentially correlate to virulence markers. For this reason, an isolate rich in SNPs in one particular region may differ from the same number of SNPs within a different isolate. This method allowed detection of particular isolates that could be differentiated on the basis of which avirulence they contain. Designing primers around specific targeted regions would make it easier for isolates containing specific avirulences to be detected in the field.

Because of the large number of genes identified as potential avirulence genes, initial focus was placed upon the highly variable genes *Pst*21_17389, *Pst*43_19594 and *Pst*43_10686. Chapter 3 identified these specific genes as highly variable and present in all isolates, in adequate read numbers. The gene length for *Pst*43_10686 is 4323 bp, *Pst*43_19594 is 1747 bp and gene length for *Pst*21_17389 is 1341 bp, allowing for efficient primer designing. Focusing on variable genes, allowed for variation between samples to be more easily identified, in comparison to genes with shorter read lengths and those not present in all pathotypes. Although, comparing variation between samples with genes for shorter read lengths were also analysed, but the most variable genes were selected, as larger number of interesting SNPs would increase the chance of distinguishing between the isolates. Primers were usually 18-28 bp to increase their specificity, with sufficient GC content to withstand higher annealing temperatures.

4.3.2 Developing primers from 72 proteins predicted by Cantu *et al.* (2011)

Primers were developed from the initial predicted 72 potential candidate proteins containing RXLR motifs, published by Cantu *et al.* (2011). The primers were not designed around SNPs, but taken from the protein sequence as a whole. Primers were designed using the programmes Molecular Evolutionary Genetics Analysis (MEGA) version 5, which was used to identify the DNA sequence from the original protein sequences from Cantu *et al*. (2011). Once the DNA sequences were obtained the programme Primer3 was used to design the 72 primers. These primers were used as a possibility to identify any markers, and thus tested on all eight isolates systematically.

4.3.3 Preparation of genomic DNA from *Pst*

DNA was extracted from the eight spores using the DNeay plant mini DNA extraction kit (QIAgen) according to the manufacturer's instruction (See Chapter 2, Section 2.24). The integrity of the extracted DNA was observed using ethidium bromide stained gels. DNA was stored at -20°C until use in PCR.

4.3.4 Optimisation of PCR amplification conditions

The newly designed primers were used in PCRs to determine their efficiency in amplifying the target regions. The T_m of the primers was tested at 55°C to 65°C. The 30 cycles of PCR amplification were performed as follows: 94°C denaturation for 1 min, 55°C to 65°C annealing for 1 min, and 72°C extension for 1 min, using a BioRad S1000 Thermocycler PCR machine. The master mix contained 1 µl DNA, 9.5 µl water, 10 µl MangoMix, 1 µl forward primer, 1 µl reverse primer (working primers were all at 10 mM). The optimised PCR conditions were used throughout this experiment, with varying Tm, which was in accordance to the nucleotide composition of the specific primer (Table 4-5). The PCR product for each primer pair was analysed on a 1.5-2% agarose gel, 35 minutes, 60 voltage.

4.4 Results

4.4.1 Correlating absence/presence of AFLP markers with avirulence. Primer combination *EcoR*I AC + *Mse*I AA and Primer combination *EcoR*I AC + *Mse*I AC

The binary data from the primer combination *EcoR*I AC + *Mse*I AA and *EcoR*I AC + *Mse*I AC was analysed. For each of the isolates, the AFLP patterns were described as present (1) or absent (0) of each polymorphic marker. The presence/absence of specific individual AFLP bands with the presence/absence of particular *Avrs* is also included.

Table 4-2 Displays the markers generated in AFLP analysis for isolates 234E139, 169E136, 237E141, 109E9, 94/519, and Warrior, using primer combination *Eco*RI AC + *Mse*I AA (with marker name ACAA) and *Eco*RI AC + *Mse*I AC (with marker name ACAC). For each of the isolates, the AFLP patterns were described as present (1) or absent (0) of each polymorphic marker. The correlations of presence/absence of specific individual AFLP bands with the presence/absence of particular *Avrs* is also included. The marker name is correlated to the size of fragment in base pairs (bp).

AFLP band ID Marker	234E139	169E136	237E141	109E9	94/519	Warrior	Presence/absence of *Avrs*
ACAA85	**1**	1	1	1	1	1	Present in All
ACAA100	**1**	1	1	1	1	1	Present in All
ACAA170	**1**	1	1	1	1	1	Present in All
ACAA73	**0**	0	0	0	0	1	*Yr 32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA79	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA106	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA117	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA150	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA162	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA166	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA223	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA267	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA280	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA312	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**

ACAA320	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA330	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA386	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA403	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAA90	**0**	0	0	1	0	1	*Yr6* **(present)**
ACAA119	**0**	0	0	1	0	1	*Yr6* **(present)**
ACAA108	**1**	0	1	1	0	0	*Yr17* **(absent)**
ACAA110	**0**	0	0	1	0	1	*Yr17* **(absent)**
ACAA277	**1**	0	1	1	0	0	*Yr17* **(absent)**
ACAA145	**1**	0	1	0	0	0	*Yr1* **(absent)**
ACAA178	**1**	0	1	0	0	0	*Yr1* **(absent)**
ACAC65	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAC75	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAC89	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAC117	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAC132	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAC167	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAC177	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAC191	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAC208	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAC259	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAC267	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**

ACAC275	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAC404	**0**	0	0	0	0	1	*Yr32, Rob, Sol, Sp, Tim, War* **(present)**
ACAC172	**1**	0	1	1	0	1	*Yr4*
ACAC241	**1**	0	1	1	0	0	*Yr17*
ACAC253	**0**	1	0	0	1	1	*Yr17*
ACAC110	**1**	0	1	0	0	0	*Yr1*
ACAC140	**1**	0	1	0	0	0	*Yr1*
ACAC216	**1**	0	1	0	0	0	*Yr1*
ACAC242	**1**	0	1	0	0	0	*Yr1*

Table 4-2 compares the presence and absence of avirulences in isolates 234E139, 169E136, 237E141, 109E9, 94/519, and Warrior with binary data from primer combinations *EcoR*1 AC + *Mse*I AA and *EcoR*I AC + *Mse*I AC. With primer combination *EcoR*1 AC + *Mse*I AA, the presence of markers ACAA73, ACAA79, ACAA106, ACAA117, ACAA150, ACAA162, ACAA166, ACAA223, ACAA267, ACAA280, ACAA312, ACAA320, ACAA330, ACAA386, and ACAA403 in the isolate Warrior but absence of bands in the other isolates, support the hypothesis that these markers are associated to *Avr32, AvrRob, AvrSol, AvrSp, AvrTim, AvrWar*. This correlation can be identified in table 4-2, in which the presence of a marker is described with '1' and absence with a '0'. Table 4-2 displays that the sample Warrior consists of markers present with '1', and the other isolates consist of '0's. Similarly, AFLP markers ACAA90 and ACAA119 were also present in the two isolates 109E9 and Warrior, which were not present in any other isolate. Presence of these markers can correlate to gene *Avr6*, as seen in Table 4.1.

Correlations of *Avr* marker absence was however, seen in isolates 169E136, 94/519, and Warrior. Table 4-2 displays absence with a '0' for these isolates, but presence is noted with a '1' for the other isolates (234E139, 237E141, and 109E9). The absence of AFLP markers ACAA108, ACAA110, and ACAA277 in isolates 169E136, 94/519, and Warrior, interestingly associates them to presence of *Avr*17 because it is specifically present amongst these three pathotypes. Likewise, 234E139 and 237E141 have presence/absence of bands for markers ACAA145 and ACAA178, which are not present in all other isolates (Table 4-2). Correlating the presence of these markers to Table 4-1 indicates that this marker could be linked to *Avr1*.

Table 4-2 also includes the presence and absence of avirulences from primer combination *EcoR*I AC + *Mse*I AC. Similarly to primer combination *Eco*RI AC + *Mse*I AA, it is apparent that majority of the identified AFLP bands (ACAC65, ACAC75, ACAC89, ACAC117, ACAC132, ACAC167, ACAC177, ACAC191, ACAC208, ACAC259, ACAC267, ACAC275, ACAC404) have been generated that correlate to presence of *Avr*32, *AvrRob, AvrSol, AvrSp, AvrTim, AvrWar.* On the other hand, the absence of virulence on *Yr4* is hypothesised to correlate with absence of the AFLP marker ACAC172 is absent in isolate 169E136 and 94/519, but present in the other isolates, which consist of the *Yr4* gene. Isolates 234E139 and 237E141 have the markers ACAC140, ACAC216, and ACAC242 present, which are absent in the other isolates. These two pathotypes are avirulent on *Yr1*, so likely have the *Avr1* gene, whilst the other pathotypes (which do not have these markers present) are virulent on *Yr1*. Moreover, bands (ACAC241 and ACAC253) present in pathotypes 234E139, 237E141, and 109E9 correlate to absence of *Yr17*, similarly how the absence of *Yr17* in samples 234E139, 237E141, and 109E9 potentially correlate to the absence of virulence on *Yr17* gene.

Section 4.4.2 gives an insight into how specific primer combinations can correlate presence/absence (1/0) of markers with the different *Avrs.* The above data illustrates that majority of the potential markers were identified in Warrior, establishing an association with *Avr*32, *Avr* Rob, *Avr* Sol, *Avr* Sp, *Avr* Tim, *Avr* War (Table 4.2).

4.4.2 All samples with primer combination *Eco*RI AA + *Mse*I TG

Table 4-3 below is an illustration of the binary data obtained from primer combination *Eco*RI AA + *Mse*I TG used for all eight isolates from the AFLP analyses.

Table 4-3 Highlights the markers generated in AFLP analysis for isolates 234E139, 169E136, 237E141, 109E9, 94/519, KWS, Solstice and Warrior. These markers are attributed to *Eco*RI AA + *Mse*I TG. For each of the isolates, the AFLP patterns were described as present (1) or absent (0) of each polymorphic marker. The presence /absence of specific individual AFLP bands with the presence/absence of particular Avrs is also included.

Marker	234E139	169E136	237E141	109E9	94/519	KWS	Solstice	Warrior	Presence/ absence of *Avrs*
AATG81	**1**	1	1	1	1	1	1	1	**Present in all**
AATG85	**1**	1	1	1	1	1	1	1	
AATG90	**1**	1	1	1	1	1	1	1	
AATG93	**1**	1	1	1	1	1	1	1	
AATG100	**1**	1	1	1	1	1	1	1	
AATG104	**1**	1	1	1	1	1	1	1	
AATG120	**1**	1	1	1	1	1	1	1	
AATG137	**1**	1	1	1	1	1	1	1	
AATG159	**1**	1	1	1	1	1	1	1	
AATG161	**1**	1	1	1	1	1	1	1	
AATG173	**1**	1	1	1	1	1	1	1	
AATG240	**1**	1	1	1	1	1	1	1	
AATG247	**1**	1	1	1	1	1	1	1	
AATG280	**1**	1	1	1	1	1	1	1	
AATG321	**1**	1	1	1	1	1	1	1	
AATG400	**1**	1	1	1	1	1	1	1	
AATG421	**1**	1	1	1	1	1	1	1	
AATG440	**1**	1	1	1	1	1	1	1	
AATG479	**1**	1	1	1	1	1	1	1	
AATG499	**1**	1	1	1	1	1	1	1	
AATG519	**1**	1	1	1	1	1	1	1	
AATG539	**1**	1	1	1	1	1	1	1	
AATG560	**1**	1	1	1	1	1	1	1	
AATG600	**1**	1	1	1	1	1	1	1	
AATG621	**1**	1	1	1	1	1	1	1	
AATG641	**1**	1	1	1	1	1	1	1	
AATG110	**0**	1	0	1	1	1	1	1	*Yr1* (absent)
AATG189	**1**	0	1	0	0	0	0	0	*Yr1* (absent)
AATG381	**1**	0	1	0	0	0	0	0	*Yr1* (absent)
AATG166	**0**	0	0	0	0	1	1	1	*Yr32* (present)
AATG303	**0**	1	0	0	1	0	0	0	*Yr4* (absent)

The markers identified in Table 4-3 correlate to the virulences in Table 4-1. Table 4.3 presents all eight isolates with primer combination *Eco*RI AA + *Mse*I TG. This analysis found an AFLP marker (AATG166) that is correlated to *Avr*32, which discriminates between the old and new pathotypes. The *Yr*32 gene is the only Avr present only in isolates KWS, Solstice, and Warrior and thus the absence of virulence on *Yr*32 amongst isolates 234E139, 169E136, 237E141, 109E9, and 94/519 discriminates between these isolates. As identified above, *Yr*32 gene was correlating with an AFLP marker that was absent in isolates 234E139 and 237E141 (Table 4-2). Similarly, the inverse was seen as the marker was identified as absent in isolates 234E139 and 237E141, and present in the rest. For *Avr*4, marker AATG303, was present in isolate 169E136 and 94/519, but absent from the other isolates, therefore presence of the band correlates with absence of the Avr4 gene, as also seen in Table 4-3.

4.4.3 Developing primers from 72 proteins predicted by Cantu *et al.* (2011)

The hypothesis was tested that the 72 predicted candidate effector genes from Cantu *et al.*, (2011) are candidates for the *Avr* genes in the test *Pst* isolates. Figure 4-1 shows the amplified *Pst* products using some of the developed primers. It was found that half of the 72 primer pairs did amplify the eight *Pst* isolates, but none of which showed any discrimination between the different pathotypes. It was important to associate the ability of these 72 primers with the eight isolates containing different avirulence genes. As no discrimination was identified, it can be concluded that the selected 72 proteins were not any of the avirulence genes.

Referring back to Table 4-1 will help to understand this concept. The concept was to identify one of the 72 genes as an avirulence gene; for example, if one of the 72 genes were recognised as *Avr*4, and this gene is lost by deletion from isolates that have become virulent on *Yr*4, for example 234E139, 237E141, 109E9, Solstice, KWS, and Warrior, the developed primer would only amplify isolates 169E136 and 94/519 (which still theoretically contain *Avr*4). Unfortunately this distinction was not made through these primers, and thus correlation with known avirulences was not possible to identify. Figure 4-1 shows that examples of one of these primers amplified all isolates tested and one that lacked the ability to amplify any product.

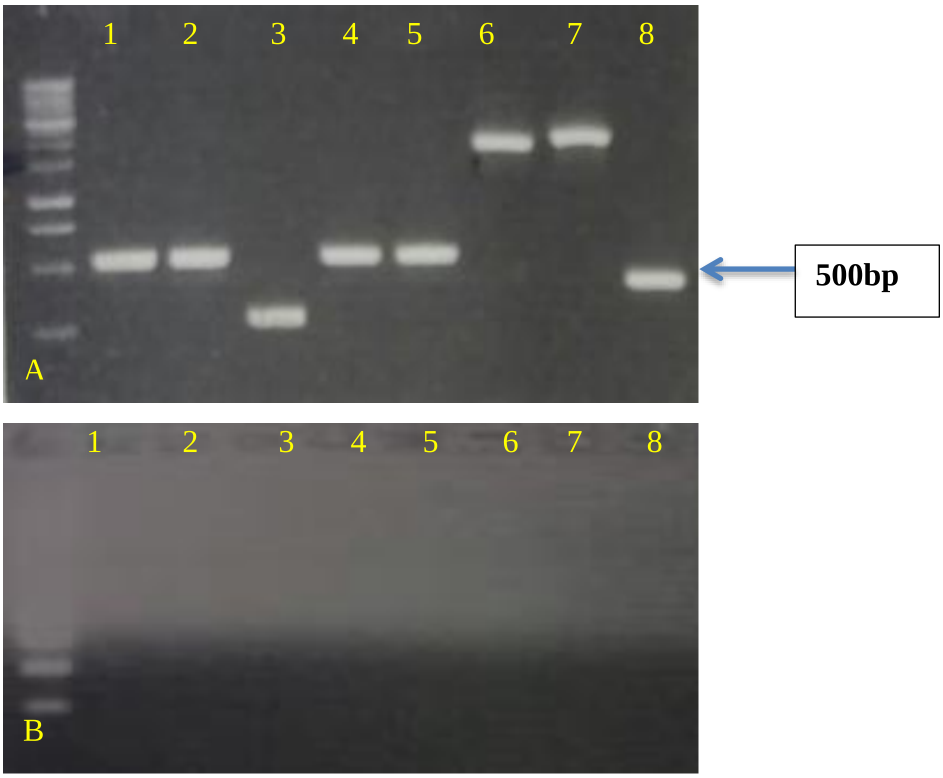


Figure 4-1 Representation of two genes taken from Cantu *et al.* (2011) **(A)** Gene *Pst*130_28359 did not show any discrimination between the different pathotypes and **(B)** Gene *Pst*130_9117_4 failed to amplify any product at all. Samples in lanes 1) 234E139 2) 169E136 3) 237E141 4) 109E9 5) 94/519 6) KWS 7) Solstice 8) Warrior. 1 Kb ladder displays products amplified at approximately 500bp.

4.4.4 Presence/absence of genes correlated with presence/absence of avirulence genes

This section will look into genes that were present/absent in the sequencing data obtained from Chapter 3, which correlated with presence/absence with avirulence genes. Table 4-4 lists the 146 genes used in Illumina sequencing, and these have been positively or negatively correlated with specific avirulence genes (Chapter 3). The first section of the table displays genes that were present in all eight isolates. We then see specific correlations to *Avr*1, *Avr*4, *Avr*6, *Avr*7, *Avr*17, *Avr*Ste. Table 4-4 displays the potential *Avr* genes identified through this study.

Gene *Pst*43_3149 and *Pst*877_18061 can be identified as potential avirulence genes, specifically *Avr1,* because both genes were present in isolates 234E139 and 237E141 but absent in the others. The absence of virulence on *Yr1* (Table 4.1) suggests the presence of *Avr1*. Likewise, *Avr4* can be identified as potentially one of the genes *Pst*0821_12746, *Pst*877_3038, *Pst*43_20465 or *Pst*43_2877. These genes were present in isolates 169E136 and 94/519 but absent in 234E139, 237E141, 109E9, KWS, Solstice, and Warrior. The presence of these genes in these two isolates correlates them to *Avr4*. *Pst*43_2877 however, shows an inverse correlation with *Avr4* in comparison to genes *Pst*0821_12746, *Pst*877_3038, *Pst*43_20465. For gene *Pst*43_2877 we see that the potential avirulence gene *Avr4* is absent in isolates 169E136 and 94/519 and present in the rest.

Gene *Pst*21_14230 is negatively correlated with *Avr6* due to its absence in pathotypes 234E139, 169E136, 237E141, 94/519 and presence in isolates 109E9, Solstice, KWS, and Warrior. Isolates 109E9, Solstice, KWS, and Warrior consist of the *Yr6* gene, and the absence of virulence in pathotypes 234E139, 169E136, 237E141, 94/519 indicates the potential correlation between gene *Pst*21_14230 and *Avr6*. Gene *Pst*21_18506 can be potentially identified as *Avr7*. The findings from chapter 3 highlight the presence of this particular gene in pathotypes 169E136, 109E9, and 94/519.

Another negative correlation was identified between gene *Pst*877_4028 and *Avr*17. The absence of this gene amongst isolates 234E139, 237E141, 109E9 highlight that it has an association with *Avr17*. Pathotypes 169E136, 94/519, KWS, Solstice, and Warrior all possess the *Avr17* gene. Thus the presence of gene *Pst*877_4028 in these isolates implies a negative correlation between *Pst*877_4028 and *Avr17*. Similarly, gene *Pst*130_4892 is absent in isolate KWS, but present in all other isolates. This again implies a negative correlation between *AvrSte* and gene *Pst*130_4892.

This study found that many of the genes correlated with the older virulences, and only one gene (*Pst*130_4892) shared any association with a new virulence (*AvrSte*). The other new virulences identified mainly in isolate KWS, Solstice, and Warrior, did not appear to correlate to any of the studied genes. This data has been extremely informative in identifying potential avirulence genes, and therefore markers for the presence/absence of these genes in yellow rust isolates.

Table 4-4 The presence/absence of 146 genes used in Illumina sequencing. Analysis of the presence/absence of genes, positively or negatively correlated with a specific avirulence gene. Correlation with *Avr1, Avr4, Avr6, Avr7, Avr17, AvrSte* have been identified and listed.

Gene Name	G_length	234E139	169E136	237E141	109E9	94/519	KWS	Solstice	Warrior	*Avr*
PST43_13507	1609	1	1	1	1	1	1	1	1	
PST43_18126	1697	1	1	1	1	1	1	1	1	
PST21_12873	3706	1	1	1	1	1	1	1	1	
PST0821_13334	760	1	1	1	1	1	1	1	1	
PST43_13004	9334	1	1	1	1	1	1	1	1	
PST21_18360	4390	1	1	1	1	1	1	1	1	
PST43_13577	1873	1	1	1	1	1	1	1	1	
PST130_11467	784	1	1	1	1	1	1	1	1	
PST43_15964	1094	1	1	1	1	1	1	1	1	
PST21_11846	3103	1	1	1	1	1	1	1	1	
PST21_16433	2246	1	1	1	1	1	1	1	1	
PST130_7070	14091	1	1	1	1	1	1	1	1	
PST877_4136	3381	1	1	1	1	1	1	1	1	
PST130_12532	6825	1	1	1	1	1	1	1	1	
PST43_10686	4323	1	1	1	1	1	1	1	1	
PST21_7457	9438	1	1	1	1	1	1	1	1	
PST43_1402	1586	1	1	1	1	1	1	1	1	
PST21_10879	5202	1	1	1	1	1	1	1	1	
PST43_12897	6817	1	1	1	1	1	1	1	1	
PST0821_18029	2075	1	1	1	1	1	1	1	1	
Pst21_12116	1286	1	1	1	1	1	1	1	1	
Pst130_860	1073	1	1	1	1	1	1	1	1	
Pst21_17139	1649	1	1	1	1	1	1	1	1	
Pst43_17354	918	1	1	1	1	1	1	1	1	
Pst21_1358	3562	1	1	1	1	1	1	1	1	
Pst43_20558	1279	1	1	1	1	1	1	1	1	
Pst21_10341	3025	1	1	1	1	1	1	1	1	

Gene Name	G_length	234E139	169E136	237E141	109E9	94/519	KWS	Solstice	Warrior	*Avr*
Pst21_15602	678	1	1	1	1	1	1	1	1	
Pst21_14919	4425	1	1	1	1	1	1	1	1	
Pst21_13429	1042	1	1	1	1	1	1	1	1	
Pst43_18621	805	1	1	1	1	1	1	1	1	
Pst21_10828	5758	1	1	1	1	1	1	1	1	
Pst43_12895	21300	1	1	1	1	1	1	1	1	
***Pst*43_3149**	1176	1+	0-	1+	0-	0-	0-	0-	0-	***Avr1***
***Pst*877_18061**	313	1+	0-	1+	0-	0-	0-	0-	0-	***Avr1***
***Pst*0821_12746**	318	0-	1+	0-	0-	1+	0-	0-	0-	***Avr4***
***Pst*877_3038**	1298	0-	1+	0-	0-	1+	0-	0-	0-	***Avr4***
***Pst*43_20465**	1155	0-	1+	0-	0-	1+	0-	0-	0-	***Avr4***
***Pst*43_2877**	1647	1+	0-	1+	1+	0-	1+	1+	1+	***Avr4***
***Pst*21_14230**	926	0-	0-	0-	1+	0-	1+	1+	1+	***Avr6* (inverse)**
***Pst*21_18506**	275	0-	1+	0-	1+	1+	0-	0-	0-	***Avr7***
***Pst*877_4028**	1063	0-	1+	0-	0-	1+	1+	1+	1+	***Avr17* (inverse)**
***Pst*130_4892**	1424	1+	1+	1+	1+	1+	0-	1+	1+	***AvrSte* (Inverse)**
PST43_18806	840	1	1	1	0	0	0	0	0	
PST877_15584	372	0	0	0	0	0	0	0	0	
PST21_17701	558	1	1	1	1	0	0	0	0	
PST21_9249	1780	1	1	1	1	0	1	1	1	
PST21_5006	546	0	1	0	0	0	0	0	0	
PST43_5641	246	1	1	0	1	0	0	0	0	
PST877_4381	2436	0	1	1	1	1	1	1	1	
PST877_12370	884	1	0	1	1	0	0	1	1	
PST0821_4858	2251	1	0	1	1	1	1	1	1	
PST21_18447	568	0	1	0	0	0	1	1	1	
PST43_17740	2305	0	0	1	0	0	0	0	1	
PST877_1814	305	0	0	0	0	0	0	0	0	
PST130_7531	3637	1	1	1	0	0	1	1	1	
PST0821_25	1233	0	0	1	1	0	0	0	1	
PST43_14940	1235	1	1	1	1	0	0	1	1	

Gene Name	G_length	234E139	169E136	237E141	109E9	94/519	KWS	Solstice	Warrior	*Avr*
PST43_4247	351	0	0	0	0	0	0	0	0	
PST0821_3901	1169	1	1	1	1	0	1	1	1	
PST21_18570	765	0	0	0	1	0	0	0	0	
PST877_16576	572	0	0	0	1	1	0	0	0	
PST877_15596	415	0	0	0	0	0	0	0	0	
PST0821_7170	532	0	0	0	0	0	0	0	0	
PST21_16369	5229	1	1	1	1	0	0	1	1	
PST130_14831	315	0	0	0	0	0	0	0	0	
PST130_14131	683	0	0	0	0	0	0	0	0	
PST130_16778	376	0	0	0	1	0	0	0	0	
PST0821_9560	494	0	0	0	1	1	1	0	0	
PST21_11721	4448	1	1	1	0	1	1	1	1	
PST43_10655	4080	0	1	1	1	1	1	1	1	
PST21_15642	866	0	1	0	1	0	0	0	1	
PST877_19843	414	0	0	0	0	0	0	0	0	
PST877_4484	2564	1	1	1	0	1	1	1	1	
PST130_16051	706	0	1	0	1	0	0	0	0	
PST21_18431	797	0	0	0	0	1	0	0	0	
PST130_285	406	0	0	0	0	0	0	0	0	
PST43_3249	1058	0	0	0	0	0	0	0	0	
PST21_17389	1341	1	1	1	1	0	1	1	1	
PST43_19594	1747	1	0	0	0	1	1	1	1	
PST0821_6060	876	0	0	0	0	0	0	0	0	
PST0821_6581	1614	1	0	1	0	1	1	1	0	
PST43_7857	211	0	0	0	0	1	0	0	0	
PST877_15391	1023	1	0	0	0	0	0	0	0	
PST0821_2920	4030	1	0	1	1	1	1	1	1	
PST21_15640	838	0	0	0	0	1	0	0	0	
PST0821_7254	533	0	1	1	0	1	1	1	1	
PST43_5258	374	0	0	0	0	0	0	0	0	
PST43_9436	3768	0	0	0	1	0	1	0	1	

Gene Name	G_length	234E139	169E136	237E141	109E9	94/519	KWS	Solstice	Warrior	*Avr*
PST43_20314	2851	0	0	0	0	1	1	1	1	
PST43_11457	980	0	0	0	0	1	0	0	0	
PST130_3899	343	0	0	0	0	0	0	0	0	
PST43_8388	250	0	0	0	0	0	0	0	0	
PST43_698	6394	1	1	1	1	0	1	1	1	
PST21_12834	917	1	1	1	1	0	0	1	1	
PST0821_9721	377	0	0	0	0	1	0	0	0	
PST0821_6756	848	0	0	0	0	0	0	0	0	
PST877_15982	315	0	0	0	0	0	0	0	0	
PST43_3362	419	0	0	0	0	0	0	0	0	
PST130_9320	6852	1	0	1	1	1	1	1	1	
PST0821_728	357	0	0	0	1	0	0	0	0	
PST21_17484	717	0	0	0	0	0	0	0	0	
PST43_4708	679	1	1	0	1	0	0	0	0	
PST0821_9979	403	0	0	0	0	1	0	0	0	
PST0821_5302	1848	0	1	1	0	1	1	1	1	
PST43_1515	2552	0	1	1	1	0	1	1	1	
PST877_10294	481	1	0	0	0	1	1	0	0	
PST0821_16435	305	0	0	0	0	0	0	0	0	
PST0821_5185	1061	1	1	1	0	0	0	0	0	
PST130_02589	301	0	0	0	0	0	0	0	0	
PST43_3536	834	1	1	1	0	0	1	1	0	
PST130_12956	1385	1	1	1	0	1	1	1	1	
PST877_00017	315	0	0	0	0	0	0	0	0	
PST0821_13679	526	0	0	0	0	1	0	0	0	
PST43_16562	1336	1	1	1	1	0	1	1	1	
PST21_19352	225	0	0	0	0	0	0	0	0	
PST877_14212	1580	1	1	1	0	1	1	1	1	
PST0821_217	424	0	0	0	0	0	0	0	0	
PST130_7695	1148	1	1	1	1	0	0	1	1	
PST43_16095	1245	1	1	0	1	0	1	0	0	

Gene Name	G_length	234E139	169E136	237E141	109E9	94/519	KWS	Solstice	Warrior	*Avr*
PST877_240	336	0	0	0	0	0	0	0	0	
PST43_04767	444	0	0	0	0	0	0	0	0	
PST130_1536	1412	1	0	1	1	1	1	1	1	
PST21_18223	225	0	0	0	0	0	0	0	0	
PST0821_10041	631	0	0	0	0	0	0	0	0	
PST21_15030	561	0	0	0	0	0	0	0	0	
PST877_9083	625	1	0	0	0	0	0	0	0	
PST0821_3872	1308	0	0	0	1	1	1	1	1	
PST21_18221	1459	1	0	1	1	0	1	1	0	
Pst21_18220	478	0	0	0	0	1	1	0	0	
pst21_10634	4743	1	1	1	1	0	1	1	1	
pst21_18784	981	1	0	0	1	1	1	1	1	
pst43_8462	285	1	0	0	0	0	0	0	0	
pst0821	505	0	0	0	0	0	0	0	0	
pst130_14643	354	0	0	0	0	1	0	0	0	
pst43_4004	943	0	1	0	0	1	1	0	0	
pst130_6558	307	0	0	0	0	0	0	0	0	
pst21_18427	717	0	0	0	0	1	1	1	0	
pst21_18994	454	0	0	0	1	0	0	0	0	
pst0821_20235	4110	0	1	1	0	0	1	1	1	
pst21_8753	319	0	0	0	0	0	0	0	0	
pst0821_16446	403	0	0	0	0	0	0	0	0	
pst0821_1106	497	0	0	0	0	0	0	0	0	

4.4.5 Designing primers based on SNPs to develop PCR markers

A large number of primers were developed to test against the isolates, to specifically amplify targeted regions of the genes *Pst*21_17389, *Pst*43_19594 and *Pst*43_10686 within isolates 234E139, 106E139, 237E141, 169E136, 109E9, KWS Sterling, Solstice, and Warrior. The aim was to determine whether the ability of the primers to amplify from the different pathotypes correlated with the presence of any specific virulences within these pathotypes.

Table 4-5 Primer sequences designed to amplify SNPs used in the development of PCR amplifications.

Gene Targeted	Primer Name	Sequence (5' – 3')	Size (bp)	Tm (°C)
*Pst*43_10686	P2_Pst43_10686KWS_F	AGGATTCATTGACGATGCAC AAC	23	60
*Pst*43_10686	P3_Pst43_10686KWS_ R	TGGATTTATGTGCCCTAGGC G	21	60
*Pst*43_10686	P3_Pst43_10686KWS_F	CCCGAAAACCGGAAAGCG	22	60
*Pst*43_10686	P5_Pst43_10686KWS_ R	CTGTTGGGATCGTCTCTTCA T	21	60
*Pst*43_10686	P4_Pst43_10686_F	AGCAGAAAAGGAATGGTTC GAG	22	60
*Pst*43_10686	P5_Pst43_10686_R	CTGTTGGGATCGTCTCTTCA T	21	60
*Pst*43_19594	P2_Pst43_19594WK_F	TCGATTTCCAAGGAACCCAA GGG	23	60
*Pst*43_19594	P4_Pst43_19594WK_R	GCATCGGCCACCTCCA	16	60
*Pst*43_19594	P5_Pst43_19594KS_F	CAAGTACAAGGGAGTCTGTT TGCT	24	60
*Pst*43_19594	P6_Pst43_19594KS_R	GTGAGTCGTTGCTAGAACCG CCGCCA	26	60
*Pst*21_17389	P1_Pst21_17389_Sol_F	ATCCATTCCACGGAGGGGTA TGT	23	64
*Pst*21_17389	P4_Pst21_17389_Sol_R	CCATGGTGATGGGTACCCAA AGTTT	25	64

4.4.5.1 Primers tested around SNPs but showed no discrimination

During this study many different primers were designed, Appendix II lists all the primers that were initially tested to try and discriminate between pathotypes. In most cases the primers either amplified from no isolates or from all isolates, as illustrated in Figure 4-2. However, there were some primers that did discriminate between isolates, and these are discussed in section 4.4.6.2. All of these primers have strong positive attributes, and amplify the targeted regions.

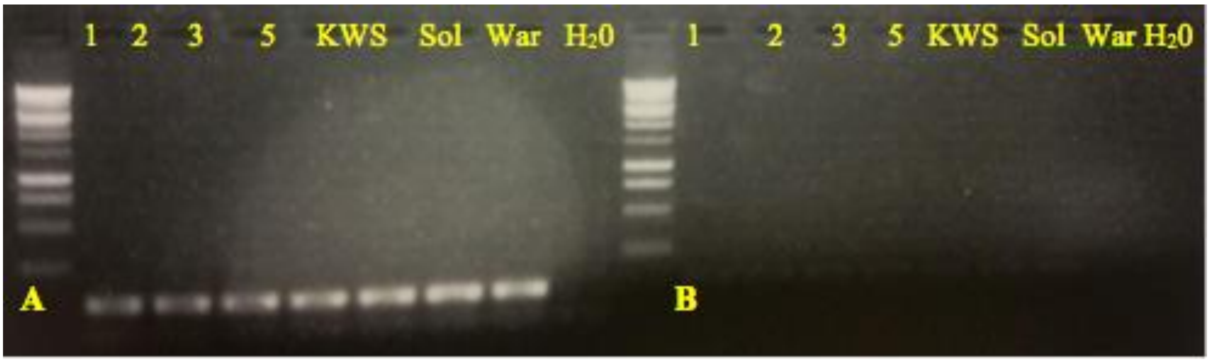


Figure 4-2 Representation of some of the results gained from the many primers that were screened that did not show any signs of correlation. In order to find primer pairs that did show potential, various primer combinations were screened. Appendix II lists the different primers designed and tested but failed to display any discrimination. **(A)** Primer pair P2_Pst43_19594_Sol_For (TCCGACAGACTCCACTCTCA) and P2_PST43_19594_Sol_Rev (GTGGGGTGTCGAGATTGACT) Tm 60°C **(B)** Pst21_17389_War_For P11 (TAACACAAGTCCTGATCCCTTTGGGA) and Pst21_17389_War_Rev P8 (GAGCTGAGGACAATGTCCAGAA) Tm 66°**C.** Lanes 1) 234E139 2) 169E136 3) 237E141 5) 109E9 & KWS, Solstice, Warrior and water control.

4.4.5.2 Primer pairs which successfully discriminated between pathotypes

A large number of primers were developed to test against the isolates, to specifically amplify targeted regions of the genes *Pst*21_17389, *Pst*43_19594 and *Pst*43_10686 within isolates 234E139, 106E139, 237E141, 169E136, 109E9, KWS

Sterling, Solstice, and Warrior. The aim was to determine whether the ability of the primers to amplify from the different pathotypes correlated with the presence of any specific virulences within these pathotypes.

4.4.5.3 Primer pair P2_Pst43_10686KWS_F/P3_PST43_10686KWS_R and P3_Pst43_10686KWS_F/ P5_Pst43_10686KWS_R

The first primer pair to show any discrimination between isolates was the forward primer P2_Pst43_10686KWS_F (Table 4-5)_ and reverse primer P3_PST43_10686KWS_R (Table 4-5). This primer pair was designed to target gene *Pst*43_10686 and produced products at annealing temperature of 60 °C.

A second primer pair, forward primer P3_Pst43_10686KWS_F (Table 4-5) and reverse primer P5_Pst43_10686KWS_R (Table 4-5) were also designed to target gene *Pst*43_10686 with product amplification also at an annealing temperature of 60 °C.

These two primer sets were designed around SNPs present in KWS Sterling as described in Section 4.3.1. These primers did not amplify isolate 234E139, 169E136, 237E141, and 109E9 due to the absence of particular SNPs, but amplified all three of the newer pathotypes (Fig-4-3), suggesting that the SNPs present in KWS were shared by Warrior and Solstice. Both primers were designed particularly close to the targeted region and thus both amplified the expected products at approximately 750bp (Fig. 4-3).

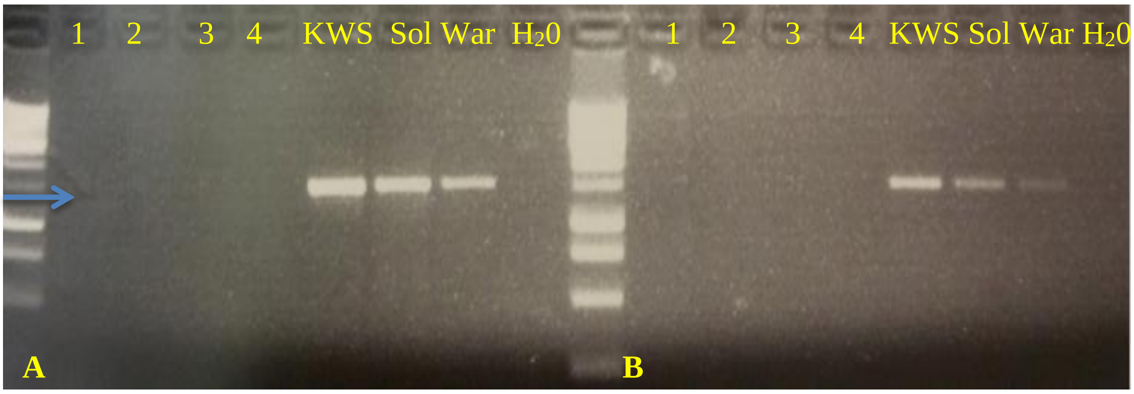


Figure 4-3 Amplification of samples KWS Sterling, Solstice, and Warrior by primer pairs **(A)** P3_*Pst*43_10686KWS_F/P5_Pst43_10686KWS_R and **(B)** P4_*Pst*43_10686KWS_F/P5 Pst43_10686KWS_R. These two specific primers do not amplify the older isolates, and it is also important to note that primers used in Fig. 4-3**A** produce slightly stronger bands than 4-3**B**. 1 Kb ladder displays products amplified at approximately 750bp. Lanes 1) 234E139 2) 169E136 3) 237E141 4) 109E9 & KWS, Solstice, Warrior and water control also used as a negative control.

As these two primer pairs distinguish between the old and new pathotypes, it is therefore a possibility that these markers are linked to virulence on *Yr*32, since this is the only one that all three of these pathotypes overcome that none of the old ones overcome.

4.4.5.4 Primer pair P2_Pst43_19594_WK_F and P4_Pst43_19594_WK_R

Another set of primers were designed to identify variations between the new isolates. This primer was designed around SNPs that were unique to KWS and Warrior, but absent in Solstice. Forward primer P2_*Pst*43_19594_KWS_F (TCGATTTCCAAGGAACCCAAGGG) and reverse primer P4_*Pst*43_19594_KWS_R (GCATCGGCCACCTCCA). This primer pair was generated to target gene *Pst*43_19594 and had an annealing temperature of 60°C. The primer pair amplified KWS and Warrior but did not amplify Solstice (Fig 4-4).

550bp

Figure 4-4 Gel electrophoresis of PCR-amplified products using DNA from isolates KWS, Solstice and Warrior at an annealing temperature of 60°C. Primer pair, **P2 *Pst*43_19594_WK _F and P4 *Pst*43_19594_WK _R** amplified from samples KWS and Warrior. Sample Solstice did not amplify, confirming the lack of SNP markers present in this isolate, which were shared by KWS and Warrior.

These results are in accordance with the markers selected to which the primers were designed. For example this particular primer was devised around SNP markers present in KWS and Warrior, but absent in Solstice. Moreover, it can be suggested that the markers are linked to *AvrRob* and *AvrSol* as Solstice does not carry the virulence on *Yr* genes '*Rob*' and '*Sol*' that are present in both KWS and Warrior.

4.4.5.5 Primer pair P5_Pst43_19594_KS_F and P6_Pst43_19594_KS_R

Primer pair P5_*Pst*43_19594_KS_F (Table 4-5) and P6_Pst43_19594_KS_R (Table 4-5) were similarly designed around SNP markers unique to KWS and Solstice, but absent in Warrior. It was important to develop different primers around similar regions in order to differentiate between closely related isolates, but also to identify which *Avrs* the markers are potentially linked to. This primer set was also designed to target gene *Pst*43_19594, with an annealing temperature of 60°C.

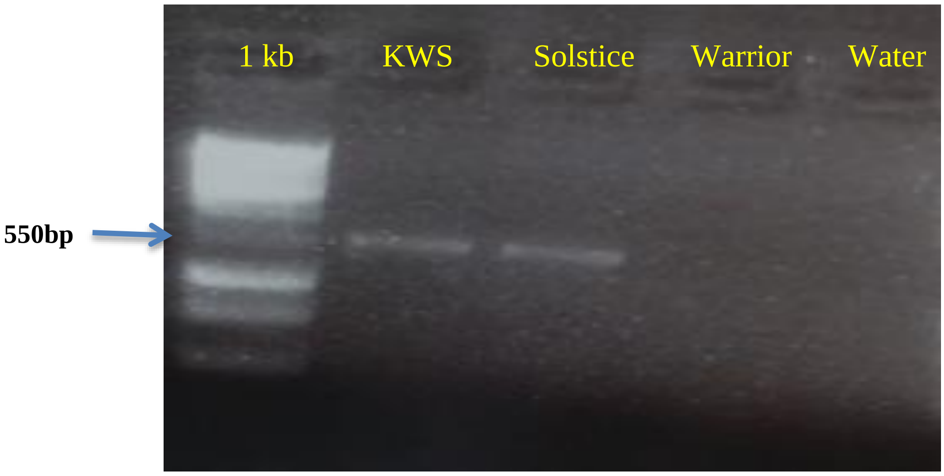


Figure 4-5 Gel electrophoresis of PCR-amplified products using DNA from isolates KWS, Solstice and Warrior at an annealing temperature of 60°C. Primer pair used for amplification was P3 *Pst*43_19594_KS _F and P6 P3 *Pst*43_19594_KS _R. Primers amplified from sample KWS and Solstice. No detection of Warrior agrees with the primer design, based on SNP markers shared by KWS and Solstice.

By analysing the products amplified from this particular primer set (P5_*Pst*43_19594_KS_F and P6_Pst43_19594_KS_R) it can be implied that these amplifications are suggestive of markers unique to only KWS Sterling and Solstice, as established in chapter 3. We can relate these markers to specific *Avrs* identified in Table 4-1. For example, Table 4-1 displays the differences in *Avrs* present in the three new isolates. Warrior distinctively carries *Yr* genes '*Sp*' '*Tim*' and '*War*' that are not shared by KWS Sterling or Solstice, which conversely comprise of *AvrSp*, *AvrTim,* and *AvrWar*. It can therefore be proposed that the markers may possibly be linked to *AvrSp*, *AvrTim*, and *AvrWar* as it amplified from pathotypes KWS and Warrior.

4.4.5.6 Primer Pair P1_Pst21_17389_Sol_F and P4_Pst21_17389_Sol_R

Primer Pair P1_*Pst*21_17389_Sol_F (Table 4-5) and P4_*Pst*21_17389_Sol_R (Table 4-5) were designed to target gene *Pst*21_17389. Primers amplified products with an annealing temperature of 64°C. This primer pair focuses on discriminating between the older isolates (234E139, 169E136, 237E141 and 109E9). Primers were designed around SNPs unique to isolate 234E139 and 237E141.

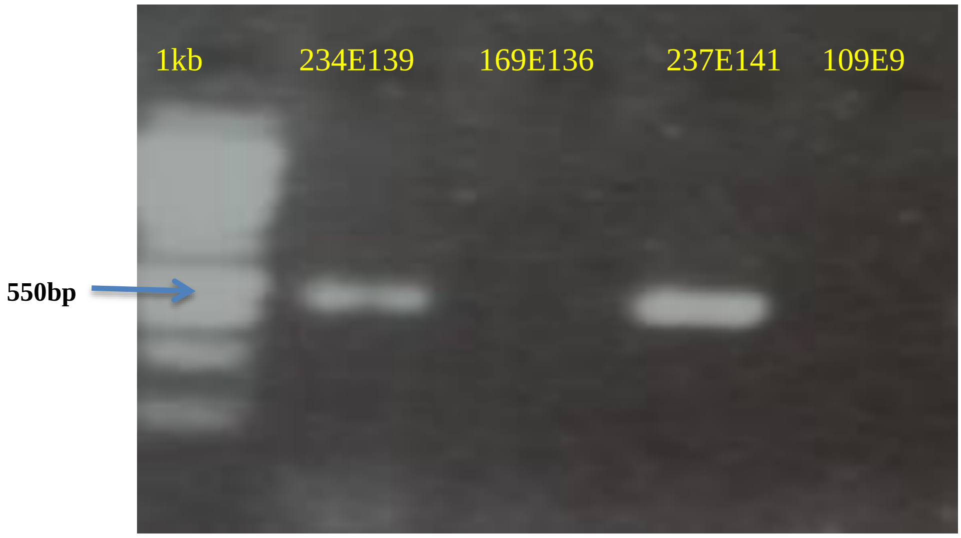


Figure 4-6 Amplification of the older isolates with primer pair P1_*Pst*21_17389_Sol_F and P4_*Pst*21_17389_Sol_R. Amplification of isolate 234E139 and 237E141, but no amplification product for sample 169E136 and 109E9, both of which are absent of *Yr* ‘7’.

Analysing products amplified from gene *Pst*21_17389 gives insight into the markers, which distinguish them from the other isolates. P1_*Pst*21_17389_Sol_F and P4_*Pst*21_17389_Sol_R failed to amplify isolates 169E136 and 109E9, and interestingly these two isolates are absent of virulence on *Yr* gene ‘*7*’ (but is present in the other isolates), which in turn implies that the *Avr*7 is present amongst 169E136 and

109E9. Lack of amplification of these two particular pathotypes from primer set P1_*Pst*21_17389_Sol_F and P4_*Pst*21_17389_Sol_R encourages the notion that there is a possibility of markers linked to *Avr7*. These findings will be important for developing resistant wheat varieties but more importantly in identifying which virulent strains are prevailing presently.

4.4.5.7 Primer pair P4_Pst43_10686_F and a P5_Pst43_10686_R

Primer pair P4_*Pst*43_10686_F and a P5_*Pst*43_10686_R were similarly designed around SNPs shared by the new set of isolates, KWS, Solstice, and Warrior. This primer pair was generated to target gene *Pst*43_10686 and had an annealing temperature of 60°C.

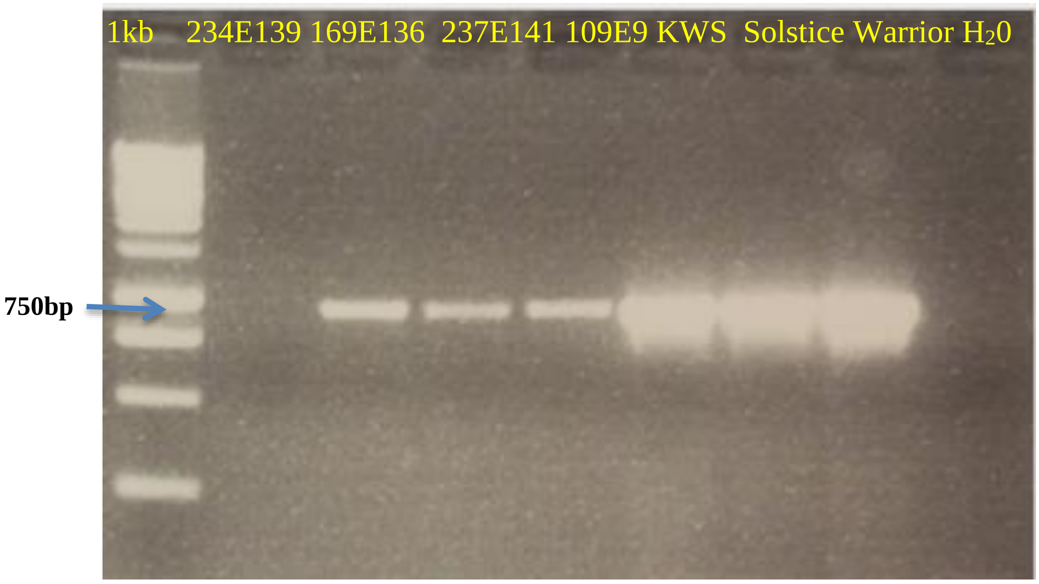


Figure 4-7 Primer P4_*Pst*43_10686_F and a P5_*Pst*43_10686_R detect all isolates, except isolate 234E139. Stronger bands are apparent for KWS, Solstice, and Warrior. 1kb ladder, and water control as a negative control is also displayed. Annealing temperature of 60°C was used.

P5_*Pst*43_10686_R did not amplify isolate 234E139. As previously described, this primer was designed around bands present in only the new set of three isolates (KWS Sterling, Warrior and Solstice). It is interesting to see that isolate 234E139; an older pathotype, was not detected. It would initially be assumed that there is a marker correlating to one of the *Avrs* in pathotype 234E139. However, based on Table 4-5, pathotypes 234E139 and 237E141 share the same avirulence profile. Therefore, if the marker were associated to one of the *Avrs* missing in pathotype 234E139 (i.e. possibly *Avr*1 to *Avr*6), then it should also fail to amplify from pathotype 237E141. The fact that this didn't happen would suggest that the marker is not associated to either of these *Avrs*. Therefore, whilst this primer pair might be useful for discriminating between pathotype 234E139 and the other pathotypes used in this study, it does not appear to be correlating to any of the *Avrs* identified in these pathotypes being studied.

4.5 Discussion

A number of studies have been published which look at the genetic relationships between *Pst* isolates (Ali *et al.*, 2010; Hubbard *et al.*, 2015). Development of PCR for detection of *Pst* has become extremely important with new varieties rapidly emerging. However, there is considerable variation in the type of DNA assays and efficiency. Designing primers to amplify specific markers present in some isolates gives an insight in to the virulence profile. This aim of this chapter was to develop the knowledge gained from Chapter 2 and Chapter 3, and use the data on identified markers and correlate them with specific virulences. The continuing aim of this study is to demonstrate a method by which the presence of several genes associated with virulence in particular strains can be used to identify emerging pathotypes in a field with the help of specified DNA assays.

Molecular markers are useful tools for genetic studies, especially genome comparisons and to understand genetic differences. Gaining further understanding in to markers that differentiate isolates is valuable information to provide a specific and sensitive DNA amplification method. This is advantageous due to its time efficiency but also its specificity. Designing primers around specific SNPs has been developed as a method by which isolates can be differentiated on this basis, and this can further be establish assays to identify virulence markers for rapid and reliable detection.

Further developing the AFLP data obtained from Chapter 2, has provided valuable information to link AFLP bands to any of the *Avrs* identified in the studied pathotypes. AFLP markers that were looked at with two different primer combinations *EcoR*I AC + *Mse*I AA and *EcoR*I AC + *Mse*I AC displayed some similarities in the markers that were identified. For example, the marker bandsACAA117 and ACAC117 were both identified in the isolate Warrior but absent in the other isolates, further supporting it as an obvious possibility that these markers are associated to *Avr32, AvrRob, AvrSol, AvrSp, AvrTim, AvrWar*. Similarly, markers ACAA403 and ACAC404 are very closely associated and therefore strongly indicate these also as markers linked to *Avr32, AvrRob, AvrSol, AvrSp, AvrTim, AvrWar*. Another similarity between the same primers were observed in the AFLP marker, ACAA167 and ACAC166. It was evident that markers associated to the above *Avrs* (*Avr32, AvrRob, AvrSol, AvrSp, AvrTim, AvrWar*) were predominantly identified through AFLP analysis using different primer combinations, therefore, consistencies within the AFLP analyses can be appreciated.

A genetic difference between the old and new pathotypes has been established in Chapters 2 and 3. This chapter further established that markers specific to older pathotypes have the ability to detect isolates through a PCR assay. Such assays potentially correlate to the virulence profiles.

The ability of a virulent pathotype to replace an existing dominant strain is becoming increasingly common. Therefore early detection through such molecular methods is an important strategy for controlling spread of disease. Repeated cycles of denaturation, polymerisation, and elongation increase the reliability of PCR techniques and accurate detection in short periods (Mullis and Faloona, 1987). However, the possibility of false negative detections requires the frequent use of controls, mainly due to DNA or PCR contaminants (Capote *et al.*, 2012). This study used water as a negative control and the previously developed 72 primers from Cantu *et al.*, 2011 as positive controls, as such controls reduce the possibilities of false positive results.

The amplifications detected in this chapter were all based upon the same total concentration of DNA template, which otherwise would have a direct effect on DNA amplifications. PCR conditions differed according to primer properties i.e. annealing temperatures. A longer primer length (26bp), inclusive of more SNPs would demand higher annealing temperatures, in comparison to smaller primer lengths (18bp) with fewer SNPs present. Reaction volumes were limited to 25ul, as a larger reaction volume would demand higher costs and increase sensitivity of the assay.

Specific identification of fungal species has been popular through molecular tools (Mirmajlessi *et al.*, 2015), who produced a systematic review that looked into PCR-based methods for detection and quantification of various strawberry pathogens (for example *Fusarium oxysporum* f.sp. *fragariae* and *Botrytis cinerea*). They analysed widely available published literature focusing on such methods, and concluded that effective detection methods for strawberry pathogens have been widely developed, reinstating confidence for *Pst* studies focusing on similar aims.

In order to design specific primers, prior knowledge of the targeted sequences is required for specific detection (Williams *et al.,* 1990). Previous studies on fungal pathogens have often focused on genes such as the internal transcribed spacer (ITS) regions, the most widely sequenced DNA region in fungi (Peay *et al.,* 2008). Specific detection with primers specifically designed around the ITS region has proved to be useful for identifying fungal species (Bonants *et al.,* 1997), but cannot be used to identify specific pathotypes of a fungus, since the ITS is not linked to avirulence traits. ITS are not likely to be linked to pathotypes is because the sequences corresponding to the rRNA of the small ribosomal subunit is highly conserved in the genome, and therefore not useful for strain discrimination (Cubero and Graham, 2002). Chapter 3 emphasised the importance on focusing sequencing of targeting regions to potentially identify more specific markers. These regions were potential avirulence gene sequences from which gene-specific primers were developed in this (Chapter 4).

Importance was placed on effector proteins in Chapter 1, and the focus of this thesis has been to analyse these disease-causing regions. Comparing random (AFLP) primers to gene specific primers was the experimental approach taken in this thesis, and therefore targeted sequencing on these specified areas enabled us to analyse genetic differences between the pathotypes. Choosing markers in candidate effectors instead of markers based on random sequences proved to be the likely choice for designing PCR primers, as we looked at genomic targets with enough sequence variation between species. Prior knowledge from bioinformatics analyses informed the targeting of specific SNP markers for developing primers sets. Depending on the DNA sequences selected, highly specific primer sets (five) were designed.

Previously, PCR based method for detecting *Pst* in Chinese wheat leaves has established that PCR application is important for rapid detection (Wang *et al.,* 2008). Wang *et al.,* (2008) analysed that genome specific primers developed rapid PCR assays to detect infection in latent infected wheat plants. By contrast, this thesis used lower annealing temperatures (51°C and 54°C) during PCR, reflecting the lack of variation in the selected *Pst* sequences. The results from Wang *et al.,* (2008) suggest that effective primer sets to detect latent stage infections, allows promising applications for the development of highly specific primers in more variable gene sequences, as established in this thesis. The concepts behind both studies share the same foundations, but we further looked into the variable sequences and related them back to virulence profiles for potential, easier isolate identification.

Primers that amplified isolates that had specific SNP markers for a particular isolate but were absent in another (as seen in Figures 4-3, 4-4, 4-5 and 4-6, and 4-7) displayed the expected banding patterns. Discriminating samples through SNP markers is valuable tool to further understand their genetic diversity, but also to enhance detection mechanisms. As expected, a lot of primers were designed and tested (see Appendix II) with varying annealing temperature before the required conditions were achieved.

It was important to design primer assays on gene sequences that provide a large scope of variation. Particular focus was placed upon genes *Pst*21_17389, *Pst*43_19594 and *Pst*43_10686 due to their variable nature, The primers in this study were designed based on the most variable sequences were those that gave bands associated with avirulences. As a result, this increased the possibility of isolate differentiation and identification through PCR amplifications. Future screening over larger *Pst* isolate populations can further validate the primers developed in this study, unfortunately this was not possible due to the lack of Pst isolates available.

Chapter 5

General Discussion

Chapter 5: General Discussion

In the past decade, new *Pst* races have emerged that have over time changed in virulences and aggressiveness (Ali *et al.*, 2014, Hubbard *et al.*, 2015). Yellow rust urediospores can undergo long distance dispersal causing frequent new disease incursions. Therefore this disease has been reported as widespread globally, becoming a major threat to wheat production in areas of susceptible hosts and favourable environmental conditions (Wellings *et al.*, 2003, Hovmøller *et al.*, 2015).

Identifying population structures of *Pst* isolates through developing diagnostic tools can help alleviate their impact by understanding their genetic relatedness, which can in turn be used to drive appropriate use of resistance gene combinations in planting strategies. We report here on the previously recognised diversity between isolates that originated between 1988 and 2006 and those that have recently become prevalent from 2008. Through AFLP and DNA sequencing experiments, this study showed how the recent isolates KWS, Solstice and Warrior vary genetically from the historical isolates 234E139, 169E136, 237E141, 109E9, and 94/519. However, phylogenetic analyses from AFLP data also identified genetic similarities between old and new pathotypes, with Warrior sharing similarities with 169E136 using primer combination *EcoR*I AA + *Mse*I TG (Fig. 2-8). Though major peaks identified in the old isolates were distinct to the new isolates, genetic relatedness between Warrior with 234E139 were established in Figure 2-5, where 234E139 shared major peaks with all three new isolates (Warrior, KWS, and Solstice).

From the sequencing data in Chapter 3 (Fig 3-7 B) and AFLP data (Fig 2-7) in Chapter 2, it can be seen that the new isolates are more closely related to each other than they are to the old isolates, with the old isolates displaying more diversity.. Overlap between certain isolates, as discussed above, is mainly due to different primer combinations used in AFLP experiments. However, reliability of this data and confirmation of relatedness is shown in Figure 2-6 where Warrior, a representative of all new isolates (KWS and Solstice) is in one clade, separate to 169E136, 237E141, and 234E139. Isolate 109E9 and 94/519 also clustered separately. This particular tree included all primer combinations, which displayed likeness to Figure 3-7 B, where the same distinctions were also seen.

The introduction of *Pst* strains into areas in which it has previously been scarce has encouraged researchers to study population diversity between *Pst* isolates in larger geographical areas (Steele *et al.,* 2001, Villaréal *et al.,* 2002, Wellings *et al.,* 2003, Hovmøller *et al.,* 2011, Zhao *et al.,* 2011). Using molecular markers, studies documented clonal populations in Europe (Enjalbert *et al.,* 2005) and Australia (Steele *et al.,* 2001, Wellings, 2007). These studies centred their findings on the theory of yellow rust only having asexual reproduction. Recently however, it has been shown that *Berberis* spp. can serve as the alternate host to complete the sexual reproduction, so recombination and emergence of new strains is now known to also occur by these means (Jin, 2011). Pakistan and China have documented high genetic diversity amongst isolates (Mboup *et al.,* 2009, Bahri *et al.,* 2011), and it has recently been suggested that the centre of origin for *Pst* isolates is the Himalayan (Pakistan and Nepalese) and near-Himalayan (China) regions, in which recombinant populations are most frequently found (Ali *et al.,* 2010, Ali *et al.,* 2014), which has been used to suggest that the Himalayan region may be a hotspot of sexual reproduction and evolution of new pathotypes, which then migrate / are transferred to other regions of the world where they then evolve asexually.

Pst has become an interesting model to address the evolution of sex. New strains are a consequence of adaptation to the environment and not of selection pressure to overcome *R* genes. This theory was supported through experimental procedures, in which the relationship between telial production (structures important for sexual reproduction) and genetic diversity of *Pst* populations was tested (Ali *et al.,* 2010). Ali *et al.,* (2010) found high telial production for genetically diverse populations and lower telial production within clonal populations, supporting the notion that sexual reproduction occurs in Himalayan and near-Himalayan areas.

Findings from this study are consistent with the above research (Ali *et al.,* 2010), as AFLP and sequencing data support the expectation of the differences between historical and new isolates. Our work shows that diversity identified between old and new pathotypes is also due to the introduction of a new strain and not resultant of a clonal production. The evidence supports the theory that the historical isolates used in this study were typical of clonal populations identified in NW European populations, which have been largely replaced by the more recent introduction of a new pathotype that has evolved sexually and been replaced by KWS, Solstice and Warrior.

Emergence of new *Pst* strains is consequential of their adaptation to different environments, for example cooler environments (Hovmøller *et al.,* 2011). This as a result generates more aggressive strains, overcoming major resistance genes and thus increasing their virulence profiles. For example, since 2011, the studied Warrior race has become prevalent in the UK replacing older strains. This strain presents the most aggressive spread of yellow rust thus far. In 2014, the ‘Kranich’ race was identified in the UK, but was not as prevalent as the Warrior race because no further ‘Kranich’ isolates (which overcame resistance to a Scandinavian variety, and thus detected within) were collected in 2015 (AHDB, 2016). This observation adds to the rust complexity identified in UK samples, suggesting Warrior, KWS and Solstice were new introductions. Prevalence of the Warrior race can be further confirmed as the UKCPVS survey found that most isolates collected in 2015 were dominated with Warrior-type isolates. Similarly, Solstice race was also dominant in the UK before the Warrior race. Recently, Warrior has been divided into 4 categories, with one group resembling the Solstice strain, a likeness found in this study through NGS technology. A recent shift in UK isolates has been acknowledged from the emergence of Warrior race behaving differently to previous strains adapted in Europe. For instance a larger number of adult plants were being infected which carried adult plant resistance. Hubbard *et al.* (2015) recognised a major shift in *Pst* population structures with high genetic diversity within isolates, a differentiation witnessed between the historical and new isolates of this study. An understanding of the changes in population structure through analysing the levels of diversity and mode of reproduction is relevant for accurately distinguishing groups of isolates and identifying emerging pathotypes.

5.1 Analysis of avirulence genes in other fungal pathosystems

The problem of disease resistance is being addressed on many fronts. Much effort has been invested in identifying new sources of resistance to yellow rust isolates with major focus on the development of resistant wheat cultivars (Chen and Line, 1992, Sui *et al.*, 2009, Powell *et al.*, 2013, Calvo-Salazar *et al.*, 2015). The use of resistant wheat varieties is the best way to control wheat losses to yellow rust. However, in order to achieve this goal, a better understanding of the host-pathogen relationship is essential.

Plant-pathogen interactions have been studied for many years, and Flor's gene-for-gene theory has been fundamental to rust research focusing on the genetic relationships between plant and pathogens (Flor, 1971, Cai *et al.*, 2002, McDowell and Woffenden, 2003, Knepper and Day, 2010, Gassmann and Bhattacharjee, 2012) . The gene-for gene theory has increasingly offered a conceptual framework for studies concentrating on protein interactions. As described in Chapter 1, Flor (1971) introduced the concept that an *R* gene product recognises a corresponding *Avr* protein (See section 1.2.8). A large number of studies resulted in the cloning and characterisation of many *R* and *Avr* genes.

Exploring *Avr* genes from broad pathosystems gives an insight in to how plant-pathogen interactions can develop a strong immune response, following recognition. *Magnaporthe oryzae*, one of the most devastating diseases of rice, is an interesting example to explore the relationship between avirulence genes and resistance. To date, 85 rice blast *R* genes have been identified; however, the variable nature of this pathogen results in resistance genes being quickly overcome. Huang *et al.* (2014) explored the evolutionary patterns of *Avr* genes in the pathogen *M. oryzae.* This study looked at six characterised *Avr* genes, which were used to examine genetic differences in 62 Chinese rice blast strains. Thus far, 10 *Avr* genes have been cloned from *M. oryzae*, and this study found that presence/absence of polymorphisms were very common amongst the *Avr* genes; for instance, *Avr-Pia* and *Avr-Pii* are present in one particular strain i.e. Ina-168 but absent in strain 70-15.

This information relates with the high sequence variation also found in *Pst* potential effector proteins. This current study observed that the majority of the potential avirulence genes thought to be highly ranked candidate effector proteins also displayed high sequence variation. This sequence variation has been highly informative in distinguishing genetic groups between different strains.

Exploring plant-pathogen interactions in different pathogen systems, particularly obligate biotrophs, presents an opportunity to better understand the role of *Avr* genes in pathogenicity and thus improve our efforts to prevent establishment of disease. For example, in the obligate biotrophic oomycete *Hyaloperonospora parasitica*, which causes downy mildew on *Arabidopsis thaliana,* Rehmany *et al.* (2005) identified an avirulence gene (ATR^{NdWsB}) which is recognised by *Arabidopsis thaliana* resistance gene product *RPP1*. Cloning six divergent alleles of the avirulence gene from different downy mildew isolates reveals that resistance gene *RPP1* differentially recognises each ATR^{NdWsB} allele. This study revealed that plant resistant proteins have the ability to recognise either single or multiple forms of a particular avirulence gene (different alleles of the same avirulence gene). Furthermore, identification of a highly conserved motif (RXLR) within ATR^{NdWsB} suggested that it might play an important role in pathogenicity by translocating effector proteins in to host plant cell, a characteristic widely used to characterise potential effector proteins (Stergiopoulos and de Wit, 2009, Petre and Kamoun, 2014, Petre *et al.*, 2014, Sperschneider *et al.*, 2015).

Examples of avirulence genes fitting Flor's gene-for-gene concept, where one *Avr* gene is recognised by the corresponding *R* gene, have been identified in *C. fulvum* (Wulff *et al.,* 2009) and *L. maculans* (Hayward *et al.,* 2012). But the complexities of these genes are further understood through examples where single *Avrs* are recognised by multiple *R* genes, as in the case of *M. lini* (Barrett *et al.,* 2009) and *M. oryzae* (Yoshida *et al.,* 2009). Intricacies of avirulence genes can be further understood as their role as both triggers and suppresses of resistance genes. Houterman *et al.* (2008) identified *Avr1* in the fungal pathogen *Fusarium oxysporum*, corresponding to an *R* genes in tomato. Although this particular avirulence gene triggers disease resistance when the host tomato carries the corresponding *R* gene, it simultaneously supresses the effect of two other resistant genes (*I* or *I-1*). It is becoming clear that the 'arm's race' between pathogen and plants vary markedly between different pathosystems. The ability of a pathogen to infect, and the hosts ability to counteract this invasion varies in different populations. A trade-off must be achieved to improve durable resistance, because selective forces being placed upon hosts i.e. pathogens overcoming *R* genes, an increase in epidemics is likely. Therefore understanding avirulence genes at the molecular level can help counteract such issues.

5.2 Development of molecular markers to identify avirulence genes

In the second and third Chapters of this thesis, we focused on identifying molecular markers using two different approaches, a random approach (using AFLPs) to identify markers linked to specific pathotypes of yellow rust, and a targeted approach (Illumina Miseq DNA sequencing) to identify markers such as SNPs in potential effector genes linked to avirulence genes. In the fourth chapter we explored the association of both types of markers with virulence/avirulence phenotypes of isolates. The aim of developing markers was that they could be used to test collections of spore samples using rapid molecular techniques to identify their pathotypes.

From the AFLP data it was possible to correlate specific AFLP markers to virulence/avirulence traits. The sufficiency of markers to distinguish between the two distinct populations was reflected through the correlation with virulences. For example

Table 4-2 highlighted that the majority of the markers were correlating to Warrior pathotype from primer combination *EcoR*1 AC and *Mse*I AA. Markers linked to newer virulences (*Yr32, YrRob, YrSol, YrSte, YrSp, YrTim, YrWar)* were not as common as finding markers linked to older virulences (*Yr1, Yr4, Yr6, Yr7, Yr17).* This supports the key virulence changes over the years, which have been monitored by the UKCPVS, underlining the need of such disease management programmes to monitor such changes. Virulence for the resistant genes *Yr1, Yr2, Yr3, Yr4, Yr9, and Yr17* was identified in most isolates suggesting these virulences are now fixed in the population. Virulence to *Yr2, Yr3,* and *Yr9* are virulent for R genes that have been overcome. If they are present in all samples markers are not likely to detect polymorphism associated with them, and therefore were not included for marker identification, .

Yellow rust pathogens are always evolving and this has been associated with the increase in virulences for cognate *R* genes. With the Warrior race emerging most recently, along with Solstice and KWS, it is anticipated that there is variation in the population for virulence to *Yr32, YrRob, YrSol, YrSte, YrSp, YrTim, YrWar*. The distinction between the markers identified in the old and new pathotypes, with fewer markers correlating to new virulences, agrees with the notion that there was a new introduction, separate from the previous clonal population, relates with these findings. Using a targeted NGS approach by sequencing potential effector protein regions, identifying presence/absence of genes correlated with presence/absence of avirulence genes using data obtained from Chapter 3 proved to be a successful method. As discussed above, this study found that many of the genes correlated with the older virulences, and only one gene (*Pst*130_4892) shared any association with a new virulence (*Avr*Ste). In summary, it can be concluded that genes *Pst*43_3149, *Pst*877_18061, *Pst*0821_12746, *Pst*877_3038, *Pst*43_20465, *Pst*43_2877, *Pst*21_14230, *Pst*21_18506, *Pst*877_4028, *Pst*130_4892 from the 147 studied potential effector proteins, could be known avirulence genes, since they show correlation with particular *Avrs*.

Developing markers in fungal pathosystems is a common method in pathology studies. Many studies have linked markers to resistance genes, for example, Mahe *et al*. (2008) developed markers for coffee leaf rust resistance. As seen with *Pst* studies, AFLP markers were developed into SCAR markers, linking the markers with resistance. Similarly, SSR markers have commonly been used in fungal pathogens to identify variation between different populations. Stroud *et al.* (2015) used SSR markers developed from isolates infecting commercial potato crops, to analyse *P. infestans* from British potatoes and found no variance in the clonal population due to overwintering.

Global food security is the major challenge for plant pathologists, as pathogens continue to spread, new virulent races overcome resistance, and this is placing pressures on breeders. Wheat varieties must be developed with several resistance genes, so that if fungal pathogen overcomes one particular *R* gene, there are still other *R* genes in the host plant to prevent disease. Therefore, breeders putting in a single resistant gene at a time is a problem that needs to be overcome and combining resistance genes for yellow rust resistance may lead to more durable resistance. However, in order to achieve this goal with traditional breeding methods is time consuming, thus development of long-lasting resistance should be explored through other methods such as Genetic Modification (GM), although such transgenic approaches to potential durable resistance come with strong public oppositions.

5.3 Conclusions and Future work

Yoshida *et al.* (2009) used whole genome sequencing to identify *M. oryzae Avr* genes, a technique that was adapted in this current study which, by contrast, used a targeted sequencing approach. Analysing the previously published draft sequence of *M. oryzae* isolate 70-15, (Dean *et al.*, 2005), this approach, however, failed to identify candidate effectors by associating polymorphisms in secreted protein genes to avirulence phenotypes. This was achieved by designing primers to potentially amplify DNA associated with presence/absence of polymorphisms. The lack of *Avr* genes identified in 70-15 strain initiated the idea to sequence a different strain Ina68, which as a result enabled the identification of 316 candidate effector proteins.

Identification of *Avr* genes in pathogenic fungi has been slowly accelerating over the years due to advancements in molecular techniques. Future work focused on whole genome sequencing, rather than targeting sequencing can reduce discrepancies between incomplete sequences, as seen in Yoshida *et al.* (2009). The absence of 1.68 Mb of genomic DNA in strain 70-15 of *M. oryzae* is a good example to appreciate that limiting a study to targeted sequencing may miss genomic regions specific to pathogenicity. Moreover, basing a study on a published draft genome may possibly limit identification of *Avr* genes due to incompleteness of the draft genome. As seen in Yoshida *et al.* (2009) large proportions of closely related strains may not be shared between isolates, therefore this further emphasises the need for sequencing of whole genomes, rather than targeted sequencing. This will however demand extra time, money and expertise in comparison to targeted regions.

Another study supports this idea; whole genome sequencing was used to identify genetic diversity in avirulence genes in the fungal plant pathogen *Leptosphaeria* *maculans*. SNP markers present a good resource to assist in developing resistant cultivars to prevent the spread of blackleg disease on Brassica species (Zander *et al.*, 2013). As seen in this thesis, genome sequencing has proved to be a great development to identify markers in a cost-effective manner to study genome-wide diversity. Other studies have also found that a reference genome provides a good opportunity to which molecular markers can be mapped to as seen with other fungal pathogens such as *Stagonospora nodorum* (Hane *et al.*, 2007) and *Fusarium graminearum* (Cuomo et al., 2007). Despite major research based on markers such as AFLP and SSRs, genome sequencing is becoming increasingly popular to assess SNPs in diverse genomes and thus linking potential markers. This is due to the NGS technology becoming progressively cost-effective and methods are becoming simpler. Despite the progress in characterising molecular markers for *Pst* described above, there are improvements that can be made to further increase our knowledge and to develop diagnostic tools for identification.

Future work must also include looking at larger number of isolates. This research although limited to eight isolates, was valuable because the isolates came from two separate introductions (the old and the new), but shared common avirulences. Therefore potentially linked markers could be identified with more confidence than if the population was a single clonal population. However, if more *Pst* isolates were included/available, and in particular from different regions of the World, it would expand the genetic pool being studied, and thus present an opportunity to look into larger sets of markers that would potentially correlate to avirulence/virulence properties.

To further classify them, it would be important to look at RNA expressions of these genes. To strengthen the findings from this study, and if more time had been available, RNA expression would have been examined in this thesis. We focused on genes in clusters that had been shown to exhibit high gene expression, and those expressing moderate gene expression (Cantu *et al.*, 2013, Hubbard *et al.*, 2015). For RNA expression experiments, this approach would enable us to identify whether expression of any of the potential effector genes identified through the marker analysis correlated with the presence / absence of the avirulence trait, thus providing more confidence that it was an avirulence gene. Expression of these genes would enable us to confirm if a gene is virulent or avirulent and further advance the findings from this thesis.

Overall, this thesis has identified some markers potentially linked to specific virulence phenotypes. Future studies, using more isolates and gene expression studies should focus on confirming which markers to develop further so that a toolkit of primers can be developed for rapid identification of pathotypes in the field.

References

References

AIRD, D., ROSS, M. G., CHEN, W.-S., DANIELSSON, M., FENNELL, T., RUSS, C., JAFFE, D. B., NUSBAUM, C. & GNIRKE, A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome biology,* 12**,** 1.

AKFIRAT, F., AYDIN, Y., ERTUGRUL, F., HASANCEBI, S., BUDAK, H., AKAN, K., MERT, Z., BOLAT, N. & UNCUOGLU, A. 2010. A microsatellite marker for yellow rust resistance in wheat. *Cereal Research Communications,* 38**,** 203-210.

ALI, S., GLADIEUX, P., LECONTE, M., GAUTIER, A., JUSTESEN, A. F., HOVMØLLER, M. S., ENJALBERT, J. & DE VALLAVIEILLE-POPE, C. 2014. Origin, migration routes and worldwide population genetic structure of the wheat yellow rust pathogen Puccinia striiformis f. sp. tritici. *PLoS Pathog,* 10**,** e1003903.

ALI, S., RAHMAN, H., JAWAD, S., SHAH, A. & ALI SHAH, S. 2010. Assessment of field resistance using host-pathogen interaction phenotype for wheat yellow rust.

AMINHA, M. 1998. *Comparison of molecular markers for study of DNA*

polymorphism in Chickpea (Cicer arietinum L.) Jawaharlal Nehru Technological University

ARIF, I., BAKIR, M., KHAN, H., AL FARHAN, A., AL HOMAIDAN, A., BAHKALI, A., AL SADOON, M. & SHOBRAK, M. 2010. Application of RAPD for molecular characterization of plant species of medicinal value from an arid environment. *Genetics and Molecular research,* 9**,** 2191-2198.

BAHRI, B., LECONTE, M., OUFFROUKH, A., DE VALLAVIEILLE - POPE, C. & ENJALBERT, J. 2009. Geographic limits of a clonal population of wheat yellow rust in the Mediterranean region. *Molecular Ecology,* 18**,** 4165-4179.

BAHRI, B., SHAH, S., HUSSAIN, S., LECONTE, M., ENJALBERT, J. & DE VALLAVIEILLE - POPE, C. 2011. Genetic diversity of the wheat yellow rust population in Pakistan and its relationship with host resistance. *Plant Pathology,* 60**,** 649-660.

BALTA, H., METİN, Ö. K., AKFIRAT, F. Ş., ERTUĞRUL, F., HASANCEBI, S., AYDIN, Y., AKAN, K., MERT, Z., TÜRET, M. & UNCUOĞLU, A. A. 2014. Identification of an AFLP marker linked with yellow rust resistance in wheat (Triticum aestivum L.). *Turkish Journal of Biology,* 38**,** 371-379.

BARRETT, L. G., THRALL, P. H., DODDS, P. N., VAN DER MERWE, M., LINDE, C. C., LAWRENCE, G. J. & BURDON, J. J. 2009. Diversity and evolution of effector loci in natural populations of the plant pathogen Melampsora lini. *Molecular biology and evolution,* 26**,** 2499-2513.

BARRUS, M. F. 1911. Variation of varieties of beans in their susceptibility to anthracnose. *Phytopathology,* 1**,** 190-195.

BECERRA, V., PAREDES, M., MADARIAGA, R., BARIANA, H., MELLADO, M. & ROJO, C. 2007. High genetic diversity in Chilean populations of wheat yellow rust (Puccinia striiformis f. sp. tritici West.) assessed by RAPD and AFLP. *Crop and Pasture Science,* 58**,** 525-531.

BERNOUX, M., MONCUQUET, P., KROJ, T. & DODDS, P. N. 2014. A novel conserved mechanism for plant NLR protein pairs: the 'integrated decoy'hypothesis. *Frontiers in plant science,* 5**,** 606.

BOLTON, M. D., VAN ESSE, H. P., VOSSEN, J. H., DE JONGE, R., STERGIOPOULOS, I., STULEMEIJER, I. J., VAN DEN BERG, G., BORRÁS - HIDALGO, O., DEKKER, H. L. & DE KOSTER, C. G. 2008. The novel Cladosporium fulvum lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species. *Molecular microbiology,* 69**,** 119-136.

BOMBARELY, A., ROSLI, H. G., VREBALOV, J., MOFFETT, P., MUELLER, L. A. & MARTIN, G. B. 2012. A draft genome sequence of Nicotiana benthamiana to enhance molecular plant-microbe biology research. *Molecular Plant-Microbe Interactions,* 25**,** 1523-1530.

BONANTS, P., HAGENAAR-DE WEERDT, M., VAN GENT-PELZER, M., LACOURT, I., COOKE, D. & DUNCAN, J. 1997. Detection and identification of Phytophthora fragariae Hickman by the polymerase chain reaction. *European Journal of Plant Pathology,* 103**,** 345-355.

BOYD, L. A., RIDOUT, C., O'SULLIVAN, D. M., LEACH, J. E. & LEUNG, H. 2013. Plant–pathogen interactions: disease resistance in modern agriculture. *Trends in genetics,* 29**,** 233-240.

BROWN, J. K. & HOVMØLLER, M. S. 2002. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science,* 297**,** 537-541.

BRYANT, R. R., MCGRANN, G. R., MITCHELL, A. R., SCHOONBEEK, H.-J., BOYD, L. A., UAUY, C., DORLING, S. & RIDOUT, C. J. 2014. A change in temperature modulates defence to yellow (stripe) rust in wheat line UC1041 independently of resistance gene Yr36. *BMC plant biology,* 14**,** 1.

BUERMANS, H. & DEN DUNNEN, J. 2014. Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease,* 1842**,** 1932-1941.

BUX, H., ASHRAF, M., CHEN, X. & MUMTAZ, S. 2011. Effective genes for resistance to stripe rust and virulence of Puccinia striiformis f. sp. tritici in Pakistan. *African Journal of Biotechnology,* 10**,** 5489-5495.

CAI, X., XU, Y. & ZHENG, Z. 2002. [Avirulence genes of plant pathogens]. *Sheng wu gong cheng xue bao= Chinese journal of biotechnology,* 18**,** 5-9.

CALVO-SALAZAR, V., SINGH, R., HUERTA-ESPINO, J., CRUZ-IZQUIERDO, S., LOBATO-ORTIZ, R., SANDOVAL-ISLAS, S., VARGAS-HERNÁNDEZ, M., GERMAN, S., SILVA, P. & BASNET, B. 2015. Genetic Analysis of Resistance to Leaf Rust and Yellow Rust in Spring Wheat Cultivar Kenya Kongoni. *Plant Disease,* 99**,** 1153-1160.

CANTU, D., GOVINDARAJULU, M., KOZIK, A., WANG, M., CHEN, X., KOJIMA, K. K., JURKA, J., MICHELMORE, R. W. & DUBCOVSKY, J. 2011. Next generation sequencing provides rapid access to the genome of Puccinia striiformis f. sp. tritici, the causal agent of wheat stripe rust. *PLoS One,* 6**,** e24230.

CANTU, D., SEGOVIA, V., MACLEAN, D., BAYLES, R., CHEN, X., KAMOUN, S., DUBCOVSKY, J., SAUNDERS, D. G. & UAUY, C. 2013. Genome analyses of the wheat yellow (stripe) rust pathogen Puccinia striiformis f. sp. tritici reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC genomics,* 14**,** 1.

CAPOTE, N., AGUADO, A., PASTRANA, A. M. & SÁNCHEZ-TORRES, P. 2012. *Molecular tools for detection of plant pathogenic fungi and fungicide resistance*, INTECH Open Access Publisher.

CATANZARITI, A.-M., DODDS, P. N., LAWRENCE, G. J., AYLIFFE, M. A. & ELLIS, J. G. 2006. Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *The Plant Cell,* 18**,** 243-256.

CHAI, Y., PHILIP PARDEY, J. B., TERRY HURLEY, & JOACHIM-BRAUN, D. K. A. H. 2014. Global Occurrence and Economic

Consequences of Stripe Rust in Wheat. *Second international wheat stripe rust symposium.* Turkey, Izmir, Regional cereal rust research center.

CHEN, W., WELLINGS, C., CHEN, X., KANG, Z. & LIU, T. 2014. Wheat stripe (yellow) rust caused by Puccinia striiformis f. sp. tritici. *Molecular plant pathology,* 15**,** 433-446.

CHEN, W., WU, L., LIU, T., XU, S., JIN, S., PENG, Y. & WANG, B. 2009. Race dynamics, diversity, and virulence evolution in Puccinia striiformis f. sp. tritici, the causal agent of wheat stripe rust in China from 2003 to 2007. *Plant Disease,* 93**,** 1093-1101.

CHEN, X. 2005. Epidemiology and control of stripe rust [Puccinia striiformis f. sp. tritici] on wheat. *Canadian Journal of Plant Pathology,* 27**,** 314-337.

CHEN, X. & LINE, R. F. 1992. Inheritance of stripe rust resistance in wheat cultivars used to differentiate races of Puccinia striiformis in North America. *Phytopathology,* 82**,** 633-637.

CHEN, X., LINE, R. F. & LEUNG, H. 1993. Relationship between virulence variation and DNA polymorphism in Puccinia striiformis. *Phytopathology,* 83**,** 1489-1497.

CHEN, X., MOORE, M., MILUS, E. A., LONG, D. L., LINE, R. F., MARSHALL, D. & JACKSON, L. 2002. Wheat stripe rust epidemics and races of Puccinia striiformis f. sp. tritici in the United States in 2000. *Plant Disease,* 86**,** 39-46.

CHENG, P. 2012. *Development and use of cDNA-derived SSR markers for studying Puccinia striiformis populations and molecular mapping of new genes for effective resistance to stripe rust in durum wheat*, Washington State University.

CLARK, B. 2015. Yellow rust. *Farmers weekly.*

COCK, P. J., FIELDS, C. J., GOTO, N., HEUER, M. L. & RICE, P. M. 2010. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic acids research,* 38**,** 1767-1771.

COLLARD, B., JAHUFER, M., BROUWER, J. & PANG, E. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica,* 142**,** 169-196.

CUOMO, C. A., GÜLDENER, U., XU, J.-R., TRAIL, F., TURGEON, B. G., DI PIETRO, A., WALTON, J. D., MA, L.-J., BAKER, S. E. & REP, M. 2007. The Fusarium graminearum genome reveals a link between localized polymorphism and pathogen specialization. *Science,* 317**,** 1400-1402.

DADREZAIE, S. T., LABABIDI, S., NAZARI, K., GOLTAPEH, E. M., AFSHARI, F., ALO, F., SHAMS-BAKHSH, M. & SAFAIE, N. 2013. Molecular Genetic Diversity in Iranian Populations of Puccinia triticina, the Causal Agent of Wheat Leaf Rust. *American Journal of Plant Sciences,* 4**,** 1375.

DE WIT, P. J., MEHRABI, R., VAN DEN BURG, H. A. & STERGIOPOULOS, I. 2009. Fungal effector proteins: past, present and future. *Molecular Plant Pathology,* 10**,** 735-747.

DEAN, R., VAN KAN, J. A., PRETORIUS, Z. A., HAMMOND - KOSACK, K. E., DI PIETRO, A., SPANU, P. D., RUDD, J. J., DICKMAN, M., KAHMANN, R. & ELLIS, J. 2012. The Top 10 fungal pathogens in molecular plant pathology. *Molecular plant pathology,* 13**,** 414-430.

DEAN, R. A., TALBOT, N. J., EBBOLE, D. J., FARMAN, M. L., MITCHELL, T. K., ORBACH, M. J., THON, M., KULKARNI, R., XU, J.-R. & PAN, H. 2005. The genome sequence of the rice blast fungus Magnaporthe grisea. *Nature,* 434**,** 980-986.

DEL FABBRO, C., SCALABRIN, S., MORGANTE, M. & GIORGI, F. M. 2013. An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS One,* 8**,** e85024.

DICKINSON, M. 2003. *Molecular plant pathology*, Garland Science.

DODDS, P. N., LAWRENCE, G. J., CATANZARITI, A.-M., AYLIFFE, M. A. & ELLIS, J. G. 2004. The Melampsora lini AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *The Plant Cell,* 16**,** 755-768.

DODDS, P. N., RAFIQI, M., GAN, P. H., HARDHAM, A. R., JONES, D. A. & ELLIS, J. G. 2009. Effectors of biotrophic fungi and oomycetes: pathogenicity factors and triggers of host resistance. *New Phytologist,* 183**,** 993-1000.

DRACATOS, P. M., DUMSDAY, J. L., OLLE, R. S., COGAN, N. O., DOBROWOLSKI, M. P., FUJIMORI, M., RODERICK, H., STEWART, A. V., SMITH, K. F. & FORSTER, J. W. 2006. Development and characterization of EST-SSR markers for the crown rust pathogen of ryegrass (Puccinia coronata f. sp. lolii). *Genome,* 49**,** 572-583.

DUAN, X., TELLIER, A., WAN, A., LECONTE, M., DE VALLAVIEILLE-POPE, C. & ENJALBERT, J. 2010. Puccinia striiformis f. sp. tritici presents high diversity and recombination in the over-summering zone of Gansu, China. *Mycologia,* 102**,** 44-53.

DUPLESSIS, S., CUOMO, C. A., LIN, Y.-C., AERTS, A., TISSERANT, E., VENEAULT-FOURREY, C., JOLY, D. L., HACQUARD, S., AMSELEM, J. & CANTAREL, B. L. 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences,* 108**,** 9166-9171.

ELLIS, J. G., LAGUDAH, E. S., SPIELMEYER, W. & DODDS, P. N. 2014. The past, present and future of breeding rust resistant wheat. *Frontiers in plant science,* 5**,** 641.

ENJALBERT, J., DUAN, X., GIRAUD, T., VAUTRIN, D., VALLAVIEILLE - POPE, D. & SOLIGNAC, M. 2002. Isolation of twelve microsatellite loci, using an enrichment protocol, in the phytopathogenic fungus Puccinia striiformis f. sp. tritici. *Molecular Ecology Notes,* 2**,** 563-565.

ENJALBERT, J., DUAN, X., LECONTE, M., HOVMØLLER, M. S. & DE VALLAVIEILLE - POPE, C. 2005. Genetic evidence of local adaptation of wheat yellow rust (Puccinia striiformis f. sp. tritici) within France. *Molecular ecology,* 14**,** 2065-2073.

EWING, B., HILLIER, L., WENDL, M. C. & GREEN, P. 1998. Base-calling of automated sequencer traces usingPhred. I. Accuracy assessment. *Genome research,* 8**,** 175-185.

FAKRUDDIN, M. & CHOWDHURY, A. 2012. Pyrosequencing an alternative to traditional Sanger sequencing. *Am J Biochem Biotechnol,* 8**,** 14-20.

FLOR, H. H. 1971. Current status of the gene-for-gene concept. *Annual review of phytopathology,* 9**,** 275-296.

GARNICA, D. P., NEMRI, A., UPADHYAYA, N. M., RATHJEN, J. P. & DODDS, P. N. 2014. The ins and outs of rust haustoria. *PLoS Pathog,* 10**,** e1004329.

GARNICA, D. P., UPADHYAYA, N. M., DODDS, P. N. & RATHJEN, J. P. 2013. Strategies for wheat stripe rust pathogenicity identified by transcriptome sequencing. *PLoS One,* 8**,** e67150.

GASSMANN, W. & BHATTACHARJEE, S. 2012. Effector-triggered immunity signaling: from gene-for-gene pathways to protein-protein interaction networks. *Molecular Plant-Microbe Interactions,* 25**,** 862-868.

GODFREY, D., BÖHLENIUS, H., PEDERSEN, C., ZHANG, Z., EMMERSEN, J. & THORDAL-CHRISTENSEN, H. 2010. Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. *Bmc Genomics,* 11**,** 1.

GOUT, L., FUDAL, I., KUHN, M. L., BLAISE, F., ECKERT, M., CATTOLICO, L., BALESDENT, M. H. & ROUXEL, T. 2006. Lost in the middle of nowhere: the AvrLm1 avirulence gene of the Dothideomycete Leptosphaeria maculans. *Molecular microbiology,* 60**,** 67-80.

GURURANI, M. A., VENKATESH, J., UPADHYAYA, C. P., NOOKARAJU, A., PANDEY, S. K. & PARK, S. W. 2012. Plant disease resistance genes: current status and future directions. *Physiological and Molecular Plant Pathology,* 78**,** 51-65.

GUYON, K., BALAGUÉ, C., ROBY, D. & RAFFAELE, S. 2014. Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen Sclerotinia sclerotiorum. *BMC genomics,* 15**,** 1.

HAMMOND-KOSACK, K. E. & JONES, J. D. 1997. Plant disease resistance genes. *Annual review of plant biology,* 48**,** 575-607.

HAMMOND - KOSACK, K. E. & KANYUKA, K. 2007. Resistance genes (R genes) in plants. *eLS*.

HANE, J. K., LOWE, R. G., SOLOMON, P. S., TAN, K.-C., SCHOCH, C. L., SPATAFORA, J. W., CROUS, P. W., KODIRA, C., BIRREN, B. W. & GALAGAN, J. E. 2007. Dothideomycete–plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen Stagonospora nodorum. *The Plant Cell,* 19**,** 3347-3368.

HARDER, D. & CHONG, J. 1984. Structure and physiology of haustoria. *The cereal rusts: origins, specificity, structure, and physiology,* 1**,** 431-476.

HAYWARD, A., MCLANDERS, J., CAMPBELL, E., EDWARDS, D. & BATLEY, J. 2012. Genomic advances will herald new insights into the Brassica: Leptosphaeria maculans pathosystem. *Plant Biology,* 14**,** 1-10.

HOLDGATE, S. 2016. New yellow rust race detected. *ADHB cereals and diseases.*

HOLDGATE, S. & WATTS, J. 2015. Yellow rust in wheat. HGCA.

HOLT, R. A. & JONES, S. J. 2008. The new paradigm of flow cell sequencing. *Genome research,* 18**,** 839-846.

HOUTERMAN, P. M., CORNELISSEN, B. J. & REP, M. 2008. Suppression of plant resistance gene-based immunity by a fungal effector. *PLoS Pathog,* 4**,** e1000061.

HOVMØLLER, M., JUSTESEN, A. & BROWN, J. 2002. Clonality and long - distance migration of Puccinia striiformis f. sp. tritici in north - west Europe. *Plant Pathology,* 51**,** 24-32.

HOVMØLLER, M. S., SØRENSEN, C. K., WALTER, S. & JUSTESEN, A. F. 2011. Diversity of Puccinia striiformis on cereals and grasses. *Annual review of phytopathology,* 49**,** 197-217.

HOVMØLLER, M. S., WALTER, S., BAYLES, R., HUBBARD, A. A., FLATH, K., SOMMERFELDT, N., LECONTE, M., ALGABA, J. R., HANSEN, J. G. & LASSEN, P. Replacement of the European wheat yellow rust population. IPM Innovation in Europe, 2015.

HOVMØLLER, M. S., YAHYAOUI, A. H., MILUS, E. A. & JUSTESEN, A. F. 2008. Rapid global spread of two aggressive strains of a wheat rust fungus. *Molecular Ecology,* 17**,** 3818-3826.

HUANG, J., SI, W., DENG, Q., LI, P. & YANG, S. 2014. Rapid evolution of avirulence genes in rice blast fungus Magnaporthe oryzae. *BMC genetics,* 15**,** 1.

HUBBARD, A., LEWIS, C. M., YOSHIDA, K., RAMIREZ-GONZALEZ, R. H., DE VALLAVIEILLE-POPE, C., THOMAS, J., KAMOUN, S., BAYLES, R., UAUY, C. & SAUNDERS, D. G. 2015. Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome biology,* 16**,** 1.

IJAZ, S. 2011. Microsatellite markers: An important fingerprinting tool for characterization of crop plants. *African Journal of Biotechnology,* 10**,** 7723-7726.

JIN, Y. 2011. Role of Berberis spp. as alternate hosts in generating new races of Puccinia graminis and P. striiformis. *Euphytica,* 179**,** 105-108.

JOHNSEN, J. M., NICKERSON, D. A. & REINER, A. P. 2013. Massively parallel sequencing: the new frontier of hematologic genomics. *Blood,* 122**,** 3268-3275.

JOHNSON, R. 1984. A critical analysis of durable resistance. *Annual review of phytopathology,* 22**,** 309-330.

JUSTESEN, A., RIDOUT, C. & HOVMØLLER, M. 2002. The recent history of Puccinia striiformis f. sp. tritici in Denmark as revealed by disease incidence and AFLP markers. *Plant pathology,* 51**,** 13-23.

KEIPER, F. J., HAYDEN, M. J. & WELLINGS, C. R. 2003. Molecular genetic variability of Australian isolates of five cereal rust pathogens. *Mycological research,* 107**,** 545-556.

KEMEN, E., KEMEN, A., EHLERS, A., VOEGELE, R. & MENDGEN, K. 2013. A novel structural effector from rust fungi is capable of fibril formation. *The Plant Journal,* 75**,** 767-780.

KEMEN, E., KEMEN, A. C., RAFIQI, M., HEMPEL, U., MENDGEN, K., HAHN, M. & VOEGELE, R. T. 2005. Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Molecular Plant-Microbe Interactions,* 18**,** 1130-1139.

KNEPPER, C. & DAY, B. 2010. From perception to activation: the molecular-genetic and biochemical landscape of disease resistance signaling in plants. *The Arabidopsis Book***,** e012.

KULKARNI, M., OZGUR, S. & STOECKLIN, G. 2010. On track with P-bodies. *Biochemical Society Transactions,* 38**,** 242-251.

KUMAR, J., SCHÄFER, P., HÜCKELHOVEN, R., LANGEN, G., BALTRUSCHAT, H., STEIN, E., NAGARAJAN, S. & KOGEL, K. H. 2002. Bipolaris sorokiniana, a cereal pathogen of global concern: cytological and molecular approaches towards better control‡. *Molecular Plant Pathology,* 3**,** 185-195.

KUMAR, P., GUPTA, V., MISRA, A., MODI, D. & PANDEY, B. 2009. Potential of molecular markers in plant biotechnology. *Plant Omics,* 2**,** 141.

LAWRENCE, G. J., DODDS, P. N. & ELLIS, J. G. 2007. Rust of flax and linseed caused by Melampsora lini. *Molecular plant pathology,* 8**,** 349-364.

LINK, T. I., LANG, P., SCHEFFLER, B. E., DUKE, M. V., GRAHAM, M. A., COOPER, B., TUCKER, M. L., MORTEL, M., VOEGELE, R. T. & MENDGEN, K. 2014. The haustorial transcriptomes of Uromyces appendiculatus and Phakopsora pachyrhizi and their candidate effector families. *Molecular plant pathology,* 15**,** 379-393.

LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L. & LAW, M. 2012. Comparison of next-generation sequencing systems. *BioMed Research International*, 2012.

LIU, X., HUANG, C., SUN, Z., LIANG, J., LUO, Y. & MA, Z. 2011. Analysis of population structure of Puccinia striiformis in Yunnan Province of China by using AFLP. *European journal of plant pathology,* 129**,** 43-55.

MAJER, D., LEWIS, B. & MITHEN, R. 1998. Genetic variation among field isolates of Pyrenopeziza brassicae. *Plant Pathology,* 47**,** 22-28.

MAJER, D., MITHEN, R., LEWIS, B. G., VOS, P. & OLIVER, R. P. 1996. The use of AFLP fingerprinting for the detection of genetic variation in fungi. *Mycological Research,* 100**,** 1107-1111.

MARKELL, S. & MILUS, E. 2008. Emergence of a novel population of Puccinia striiformis f. sp. tritici in eastern United States. *Phytopathology,* 98**,** 632-639.

MBOUP, M., LECONTE, M., GAUTIER, A., WAN, A., CHEN, W., DE VALLAVIEILLE-POPE, C. & ENJALBERT, J. 2009. Evidence of genetic recombination in wheat yellow rust populations of a Chinese oversummering area. *Fungal Genetics and Biology,* 46**,** 299-307.

MCDOWELL, J. M. & WOFFENDEN, B. J. 2003. Plant disease resistance genes: recent insights and potential applications. *TRENDS in Biotechnology,* 21**,** 178-183.

MCINTOSH, R., DEVOS, K., DUBCOVSKY, J., ROGERS, W., MORRIS, C., APPELS, R., SOMERS, D. & ANDERSON, O. 2007. Catalogue of gene symbols for wheat: 2007 supplement. *Annu Wheat Newsl,* 53**,** 159-180.

MCINTOSH, R. A., WELLINGS, C. R. & PARK, R. F. 1995. *Wheat rusts: an atlas of resistance genes*, Csiro Publishing.

METZKER, M. L. 2010. Sequencing technologies—the next generation. *Nature reviews genetics,* 11**,** 31-46.

MILNE, I., STEPHEN, G., BAYER, M., COCK, P. J., PRITCHARD, L., CARDLE, L., SHAW, P. D. & MARSHALL, D. 2012. Using Tablet for visual exploration of second-generation sequencing data. *Briefings in bioinformatics***,** bbs012.

MILUS, E., SEYRAN, E. & MCNEW, R. 2006. Aggressiveness of Puccinia striiformis f. sp. tritici isolates in the south-central United States. *Plant Disease,* 90**,** 847-852.

MIRMAJLESSI, S. M., DESTEFANIS, M., GOTTSBERGER, R. A., MÄND, M. & LOIT, E. 2015. PCR-based specific techniques used for detecting the most important pathogens on strawberry: a systematic review. *Systematic reviews,* 4**,** 1.

MORGANTE, M. & OLIVIERI, A. 1993. PCR - amplified microsatellites as markers in plant genetics. *The plant journal,* 3**,** 175-182.

MULLIS, K. B. & FALOONA, F. A. 1987. [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in enzymology,* 155**,** 335-350.

MUNDT, C. & BROWNING, J. 1985. Genetic diversity and cereal rust management. *The cereal rusts,* 2**,** 527-560.

NEI, M. & LI, W.-H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences,* 76**,** 5269-5273.

ORDOÑEZ, M. & KOLMER, J. 2009. Differentiation of molecular genotypes and virulence phenotypes of Puccinia triticina from common wheat in North America. *Phytopathology,* 99**,** 750-758.

PAIS, M., WIN, J., YOSHIDA, K., ETHERINGTON, G. J., CANO, L. M., RAFFAELE, S., BANFIELD, M. J., JONES, A., KAMOUN, S. & SAUNDERS, D. G. 2013. From pathogen genomes to host plant processes: the power of plant parasitic oomycetes. *Genome biology,* 14**,** 1.

PARK, R., FETCH, T., HODSON, D., JIN, Y., NAZARI, K., PRASHAR, M. & PRETORIUS, Z. 2011. International surveillance of wheat rust pathogens: progress and challenges. *Euphytica,* 179**,** 109-117.

PARK, R. & WELLINGS, C. 1992. Pathogenic specialisation of wheat rusts in Australia and New Zealand in 1988 and 1989. *Australasian plant pathology,* 21**,** 61-69.

PAUN, O. & SCHÖNSWETTER, P. 2012. Amplified fragment length polymorphism: an invaluable fingerprinting technique for genomic, transcriptomic, and epigenetic studies. *Plant DNA Fingerprinting and Barcoding: Methods and Protocols***,** 75-87.

PAVLICEK, A., HRDA, S. & FLEGR, J. 1999. Short Communication FreeTree-Freeware Program for Construction of Phylogenetic Trees on the Basis of Distance Data and Bootstrap/Jackknife Analysis of the Tree Robustness. Application in the RAPD Analysis of Genus Frenkelia. *Folia Biologica (Praba),* 45**,** 97-99.

PEAY, K. G., KENNEDY, P. G. & BRUNS, T. D. 2008. Fungal community ecology: a hybrid beast with a molecular master. *Bioscience,* 58**,** 799-810.

PETRE, B., JOLY, D. L. & DUPLESSIS, S. 2014. Effector proteins of rust fungi. *Frontiers in plant science,* 5**,** 416.

PETRE, B. & KAMOUN, S. 2014. How do filamentous pathogens deliver effector proteins into plant cells? *PLoS Biol,* 12**,** e1001801.

POWELL, N., LEWIS, C., BERRY, S., MACCORMACK, R. & BOYD, L. 2013. Stripe rust resistance genes in the UK winter wheat cultivar Claire. *Theoretical and Applied Genetics,* 126**,** 1599-1612.

PRETSCH, K., KEMEN, A., KEMEN, E., GEIGER, M., MENDGEN, K. & VOEGELE, R. 2013. The rust transferred proteins—a new family of effector proteins exhibiting protease inhibitor function. *Molecular plant pathology,* 14**,** 96-107.

REHMANY, A. P., GORDON, A., ROSE, L. E., ALLEN, R. L., ARMSTRONG, M. R., WHISSON, S. C., KAMOUN, S., TYLER, B. M., BIRCH, P. R. & BEYNON, J. L. 2005. Differential recognition of highly divergent downy mildew avirulence gene alleles by RPP1 resistance genes from two Arabidopsis lines. *The Plant Cell,* 17**,** 1839-1850.

ROELFS, A. P. 1992. *Rust diseases of wheat: concepts and methods of disease management,* Cimmyt.

SAJID, A. 2013. *Population biology and invasion history of Puccinia striiformis f. sp. tritici at worldwide and local scale.* Citeseer.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences,* 74**,** 5463-5467.

SARAÇLI, S., DOĞAN, N. & DOĞAN, İ. 2013. Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications,* 2013**,** 1-8.

SAUNDERS, D. G., WIN, J., CANO, L. M., SZABO, L. J., KAMOUN, S. & RAFFAELE, S. 2012. Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS One,* 7**,** e29847.

SCHUSTER, S. C. 2007. Next-generation sequencing transforms today's biology. *Nature,* 200**,** 16-18.

SHARMA-POUDYAL, D., CHEN, X., WAN, A., ZHAN, G., KANG, Z., CAO, S., JIN, S., MORGOUNOV, A., AKIN, B. & MERT, Z. 2013. Virulence characterization of international collections of the wheat stripe rust pathogen, Puccinia striiformis f. sp. tritici. *Plant Disease,* 97**,** 379-386.

SPERSCHNEIDER, J., DODDS, P. N., GARDINER, D. M., MANNERS, J. M., SINGH, K. B. & TAYLOR, J. M. 2015. Advances and challenges in computational prediction of effectors from plant pathogenic fungi. *PLoS Pathog,* 11**,** e1004806.

STEELE, K. A., HUMPHREYS, E., WELLINGS, C. & DICKINSON, M. 2001. Support for a stepwise mutation model for pathogen evolution in Australasian Puccinia striiformis f. sp. tritici by use of molecular markers. *Plant Pathology,* 50**,** 174-180.

STEFFENS, A., BRÄUTIGAM, A., JAKOBY, M. & HÜLSKAMP, M. 2015. The BEACH domain protein SPIRRIG is essential for Arabidopsis salt stress tolerance and functions as a regulator of transcript stabilization and localization. *PLoS Biol,* 13**,** e1002188.

STERGIOPOULOS, I. & DE WIT, P. J. 2009. Fungal effector proteins. *Annual review of phytopathology,* 47**,** 233-263.

STOTZ, H. U., MITROUSIA, G. K., DE WIT, P. J. & FITT, B. D. 2014. Effector-triggered defence against apoplastic fungal pathogens. *Trends in plant science,* 19**,** 491-500.

STRACHAN, T., GOODSHIP, J. & CHINNERY, P. 2014. *Genetics and Genomics in Medicine,* Taylor & Francis.

STROUD, J., SHAW, D., HALE, M. & STEELE, K. 2016. SSR assessment of Phytophthora infestans populations on tomato and potato in British gardens demonstrates high diversity but no evidence for host specialization. *Plant Pathology,* 65**,** 334-341.

STUBBS, R. 1985. Stripe rust: the cereal rusts II: diseases, distribution, epidemiology, and control. Academic Press, New York.

STUBBS, R. 1988. Pathogenicity analysis of yellow (stripe) rust of wheat and its significance in a global context. *Breeding strategies for resistance to the rusts of wheat. CIMMYT, Mexico***,** 23-38.

SUI, X., WANG, M. & CHEN, X. 2009. Molecular mapping of a stripe rust resistance gene in spring wheat cultivar Zak. *Phytopathology,* 99**,** 1209-1215.

THOMMA, B. P., VAN ESSE, H. P., CROUS, P. W. & DE WIT, P. J. 2005. Cladosporium fulvum (syn. Passalora fulva), a highly specialized plant pathogen as a model for functional studies on plant pathogenic Mycosphaerellaceae. *Molecular Plant Pathology,* 6**,** 379-393.

UPADHYAYA, N. M., MAGO, R., STASKAWICZ, B. J., AYLIFFE, M. A., ELLIS, J. G. & DODDS, P. N. 2014. A bacterial type III secretion assay for delivery of fungal effector proteins into wheat. *Molecular Plant-Microbe Interactions,* 27**,** 255-264.

VILLARÉAL, L. M., LANNOU, C., DE VALLAVIEILLE-POPE, C. & NEEMA, C. 2002. Genetic variability in Puccinia striiformis f. sp. tritici populations sampled on a local scale during natural epidemics. *Applied and environmental microbiology,* 68**,** 6138-6145.

VOEGELE, R. T. & MENDGEN, K. 2003. Rust haustoria: nutrient uptake and beyond. *New Phytologist,* 159**,** 93-100.

VUYLSTEKE, M., PELEMAN, J. D. & VAN EIJK, M. J. 2007. AFLP technology for DNA fingerprinting. *Nature protocols,* 2**,** 1387-1398.

WALTER, S., ALI, S., KEMEN, E., NAZARI, K., BAHRI, B. A., ENJALBERT, J., HANSEN, J. G., BROWN, J. K., SICHERITZ - PONTÉN, T. & JONES, J. 2016. Molecular markers for tracking the origin and worldwide distribution of invasive strains of Puccinia striiformis. *Ecology and evolution,* 6**,** 2790-2804.

WANG, M. & CHEN, X. 2015. Barberry does not function as an alternate host for Puccinia striiformis f. sp. tritici in the US Pacific Northwest due to teliospore degradation and barberry phenology. *Plant Disease,* 99**,** 1500-1506.

WANG, X., ZHENG, W., BUCHENAUER, H., ZHAO, J., HAN, Q., HUANG, L. & KANG, Z. 2008. The development of a PCR-based method for detecting Puccinia striiformis latent infections in wheat leaves. *European Journal of Plant Pathology,* 120**,** 241-247.

WEISING, K., NYBOM, H., PFENNINGER, M., WOLFF, K. & KAHL, G. 2005. *DNA fingerprinting in plants: principles, methods, and applications,* CRC press.

WELLINGS, C. 2007. Puccinia striiformis in Australia: a review of the incursion, evolution, and adaptation of stripe rust in the period 1979–2006. *Crop and Pasture Science,* 58**,** 567-575.

WELLINGS, C. & MCINTOSH, R. 1990. Puccinia striiformis f. sp. tritici in Australasia: pathogenic changes during the first 10 years. *Plant Pathology,* 39**,** 316-325.

WELLINGS, C., WRIGHT, D., KEIPER, F. & LOUGHMAN, R. 2003. First detection of wheat stripe rust in Western Australia: evidence for a foreign incursion. *Australasian Plant Pathology,* 32**,** 321-322.

WELLINGS, C. R. 2011. Global status of stripe rust: a review of historical and current threats. *Euphytica,* 179**,** 129-141.

WHISSON, S. C., BOEVINK, P. C., MOLELEKI, L., AVROVA, A. O., MORALES, J. G., GILROY, E. M., ARMSTRONG, M. R., GROUFFAUD, S., VAN WEST, P. & CHAPMAN, S. 2007. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature,* 450**,** 115-118.

WILKINSON, K., GRANT, W. P., GREEN, L. E., HUNTER, S., JEGER, M. J., LOWE, P., MEDLEY, G. F., MILLS, P., PHILLIPSON, J. & POPPY, G. M. 2011. Infectious diseases of animals and plants: an interdisciplinary approach. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 366**,** 1933-1942.

WILLIAMS, J. G., KUBELIK, A. R., LIVAK, K. J., RAFALSKI, J. A. & TINGEY, S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic acids research,* 18**,** 6531-6535.

WULFF, B., CHAKRABARTI, A. & JONES, D. 2009. Recognitional specificity and evolution in the tomato-Cladosporium fulvum pathosystem. *Molecular plant-microbe interactions,* 22**,** 1191-1202.

YANG, Z. & YODER, A. D. 1999. Estimation of the transition/transversion rate bias and species sampling. *Journal of Molecular Evolution,* 48**,** 274-283.

YOSHIDA, K., SAITOH, H., FUJISAWA, S., KANZAKI, H., MATSUMURA, H., YOSHIDA, K., TOSA, Y., CHUMA, I., TAKANO, Y. & WIN, J. 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen Magnaporthe oryzae. *The Plant Cell,* 21**,** 1573-1591.

ZADOKS, J. C. 1961. Yellow rust on wheat studies in epidemiology and physiologic specialization. *European Journal of Plant Pathology,* 67**,** 69-256.

ZANDER, M., PATEL, D. A., VAN DE WOUW, A., LAI, K., LORENC, M. T., CAMPBELL, E., HAYWARD, A., EDWARDS, D., RAMAN, H. & BATLEY, J. 2013. Identifying genetic diversity of avirulence genes in Leptosphaeria maculans using whole genome sequencing. *Functional & integrative genomics,* 13**,** 295-308.

ZHANG, Y., QU, Z., ZHENG, W., LIU, B., WANG, X., XUE, X., XU, L., HUANG, L., HAN, Q. & ZHAO, J. 2008. Stage-specific gene expression during urediniospore germination in Puccinia striiformis f. sp tritici. *BMC genomics,* 9**,** 1.

ZHAO, J., WANG, L., WANG, Z., CHEN, X., ZHANG, H., YAO, J., ZHAN, G., CHEN, W., HUANG, L. & KANG, Z. 2013. Identification of eighteen Berberis species as alternate hosts of Puccinia striiformis f. sp. tritici and virulence variation in the pathogen isolates from natural infection of barberry plants in China. *Phytopathology,* 103**,** 927-934.

ZHAO, J., ZHANG, H., YAO, J., HUANG, L. & KANG, Z. 2011. Confirmation of Berberis spp. as alternate hosts of Puccinia striiformis f. sp. tritici on wheat in China. *Mycosystema,* 30**,** 895-900.

ZHENG, W., HUANG, L., HUANG, J., WANG, X., CHEN, X., ZHAO, J., GUO, J., ZHUANG, H., QIU, C. & LIU, J. 2013. High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nature communications,* 4.

Appendix I

Gene present/absent table

	Mapped Reads							
Reference sequence name	**234E139**	**169E136**	**237E141**	**109E9**	**95/519**	**KWS**	**Solstice**	**Warrior**
PST43_18806_Tribe544_CLUSTER_VII	12046	69482	6722	2556	0	6848	8690	10404
PST877_15584_Tribe1118_CLUSTER_VII	3966	1702	1650	1032	0	2908	6192	5422
PST21_17701_Tribe948_CLUSTER_VII	126808	13698	4350	19364	0	6714	8212	10018
PST21_09249_Tribe1232_CLUSTER_VII	78076	35074	11026	78650	0	29046	20652	24820
PST21_05006_Tribe1217_CLUSTER_VII	3376	4492	2122	2792	0	6640	6716	4140
PST43_05641_Tribe1007_CLUSTER_VII	27000	6320	1036	27288	0	8674	5066	3318
PST877_04381_Tribe1176_CLUSTER_VII	3818	19304	5874	4964	2	20272	14322	25868
PST877_12370_Tribe1102_CLUSTER_VII	18026	3182	5058	6600	0	10576	11604	14046
PST0821_04858_Tribe826_CLUSTER_VII	22450	3446	19270	54564	10	72150	32258	33280
PST43_13507_Tribe467_CLUSTER_VII	31854	18450	17194	20226	6	50750	33520	39116
PST21_18447_Tribe318_CLUSTER_VII	5568	45618	3994	3550	0	19142	11498	26208
PST43_17740_Tribe318_CLUSTER_VII	1404	944	5582	326	0	14270	8360	24516
PST43_18126_Tribe354_CLUSTER_VII	70208	87618	84656	81922	4	60174	54616	85578
PST0821_12746_Tribe354_CLUSTER_VII	504	4818	666	402	2	2460	1328	2092
PST877_01814_Tribe354_CLUSTER_VII	428	2138	256	62	0	5172	1428	5550
PST130_07531_Tribe31_CLUSTER_VII	23710	8596	7566	1920	0	84896	28052	53174
PST21_12873_Tribe31_CLUSTER_VII	11030	15370	8420	18096	10	67610	33386	57010
PST0821_13334_Tribe31_CLUSTER_VII	13552	48440	6604	7628	2	17702	17886	18178
PST0821_00025_Tribe31_CLUSTER_VII	3264	1528	4574	43424	0	15108	8822	14672
PST43_14940_Tribe31_CLUSTER_VII	15168	19438	5110	5940	0	9920	9864	17710
PST43_04247_Tribe31_CLUSTER_VII	86	282	146	318	0	714	540	836
PST0821_03901_Tribe31_CLUSTER_VII	9862	16176	4510	9506	0	29456	17434	23140
PST21_18570_Tribe31_CLUSTER_VII	6486	3232	3030	38880	0	12632	9420	11782

PST877_16576_Tribe31_CLUSTER_VII	4564	2646	3512	10816	2	12572	9542	13422
PST43_13004_Tribe31_CLUSTER_VII	135406	428092	44742	6564	18	176814	110416	137274
PST877_15596_Tribe31_CLUSTER_VII	3348	720	246	102	0	1104	1228	1498
PST21_18360_Tribe437_CLUSTER_VII	129034	34078	50060	52372	10	107422	69100	72970
PST43_13577_Tribe437_CLUSTER_VII	22510	9422	6958	153204	6	48998	22028	35530
PST0821_07170_Tribe437_CLUSTER_VII	76	220	388	26	0	3136	1170	2934
PST21_16369_Tribe596_CLUSTER_VII	80752	38040	26196	7270	0	2204	31476	31590
PST130_14831_Tribe596_CLUSTER_VII	76	90	170	4	0	54	374	518
PST130_11467_Tribe678_CLUSTER_VII	70290	90624	23108	51202	8	43462	34704	31108
PST43_15964_Tribe678_CLUSTER_VII	21948	13962	14980	16436	4	33184	23446	40816
PST43_03149_Tribe446_CLUSTER_VII	11376	562	4856	2880	0	13770	6626	6676
PST130_14131_Tribe446_CLUSTER_VII	476	112	438	692	0	488	1086	1690
PST21_11846_Tribe409_CLUSTER_VII	23568	17368	10488	25244	4	39110	26142	46012
PST130_16778_Tribe409_CLUSTER_VII	5108	1984	2534	5978	0	5754	2522	2396
PST130_04892_Tribe409_CLUSTER_VII	24336	19530	10102	7738	2	13540	13832	14812
PST21_16433_Tribe429_CLUSTER_VII	26168	48668	99962	28512	2	46132	35092	18336
PST0821_09560_Tribe429_CLUSTER_VII	3336	4218	1984	5886	4	19948	5588	9126
PST21_11721_Tribe219_CLUSTER_VII	16690	7292	14316	712	2	34922	18888	65666
PST130_07070_Tribe219_CLUSTER_VII	450532	177146	213244	44114	36	480772	234832	300162
PST43_10655_Tribe219_CLUSTER_VII	4628	34840	7938	5374	2	52330	34078	46430
PST877_04136_Tribe219_CLUSTER_VII	167230	27530	46272	7074	14	135974	103808	95584
PST21_15642_Tribe308_CLUSTER_VII	2874	11558	2844	8086	0	13672	9584	17416
PST130_12532_Tribe308_CLUSTER_VII	86996	62882	58278	152782	16	140670	93694	128976
PST877_19843_Tribe308_CLUSTER_VII	40	6	84	8	0	4	430	1750
PST877_04484_Tribe308_CLUSTER_VII	42838	18202	14852	914	8	93574	32636	64458
PST43_10686_Tribe238_CLUSTER_VII	144692	6834	55186	157920	4	76230	30972	33992
PST130_16051_Tribe238_CLUSTER_VII	2544	12532	3870	14600	0	1306	4992	13150
PST21_07457_Tribe351_CLUSTER_VII	61294	23160	29072	19768	30	305522	96402	191746
PST21_18431_Tribe317_CLUSTER_VII	690	744	866	174	2	1798	1728	6212

PST130_00285_Tribe317_CLUSTER_VII	376	618	328	2258	0	600	780	2198
PST43_03249_Tribe317_CLUSTER_VII	1608	458	2278	14	0	1172	1160	3204
PST21_17389_Tribe79_CLUSTER_VII	19280	10102	12658	29464	0	34170	21642	25186
PST43_19594_Tribe79_CLUSTER_VII	22626	1374	3994	4662	2	87746	15690	25986
PST0821_06060_Tribe79_CLUSTER_VII	2812	434	1362	250	0	5264	3478	5740
PST43_01402_Tribe79_CLUSTER_VII	206488	28604	19646	32732	4	63888	40888	22450
PST0821_06581_Tribe79_CLUSTER_VII	14310	3914	29472	550	2	18602	14472	11592
PST21_10879_Tribe79_CLUSTER_VII	64368	42480	27802	59596	14	192272	112710	130558
PST21_14230_Tribe54_CLUSTER_VII	5454	2438	2894	14848	0	15466	12674	15470
PST43_07857_Tribe54_CLUSTER_VII	424	172	172	3042	2	1210	742	764
PST877_15391_Tribe54_CLUSTER_VII	12874	1130	2816	610	0	7070	3622	6086
PST0821_02920_Tribe54_CLUSTER_VII	18944	2234	9332	21302	10	86480	33432	61114
PST21_15640_Tribe54_CLUSTER_VII	1142	3658	1558	1278	2	12436	5988	9106
PST0821_07254_Tribe54_CLUSTER_VII	2178	4220	5158	2238	6	26636	15290	14992
PST43_05258_Tribe54_CLUSTER_VII	1128	462	1832	402	0	384	5296	6268
PST43_09436_Tribe54_CLUSTER_VII	3668	2726	2684	5326	0	18578	8924	24172
PST43_20314_Tribe54_CLUSTER_VII	7236	2226	3928	740	2	46214	16702	24680
PST43_12897_Tribe54_CLUSTER_VII	158712	21162	52010	77828	6	68598	52170	58534
PST0821_18029_Tribe54_CLUSTER_VII	154288	109652	91202	147976	6	116448	112470	118338
PST21_12116_Tribe74_CLUSTER_VII	24504	14958	15598	7442	2	49896	24528	31310
PST130_00860_Tribe74_CLUSTER_VII	663562	283546	125948	223570	20	196298	179604	166784
PST43_11457_Tribe74_CLUSTER_VII	202	688	620	294	2	2024	1852	5620
PST130_03899_Tribe74_CLUSTER_VII	2920	262	322	6	0	2	1854	3720
PST43_08388_Tribe74_CLUSTER_VII	126	70	318	40	0	2218	788	3566
PST21_17139_Tribe74_CLUSTER_VII	35358	13316	7504	41798	2	44022	32642	46830
PST43_00698_Tribe74_CLUSTER_VII	62324	11834	13168	10294	0	92472	58884	108286
PST21_12834_Tribe74_CLUSTER_VII	12860	8016	6212	17648	0	14302	9728	14440
PST0821_09721_Tribe74_CLUSTER_VII	114	562	458	44	2	2782	1484	5932
PST0821_06756_Tribe74_CLUSTER_VII	814	296	1322	234	0	4096	3274	5906

PST877_15982_Tribe74_CLUSTER_VII	160	60	316	388	0	1130	876	2386
PST43_03362_Tribe394_CLUSTER_VII	1714	630	542	538	0	9846	3070	5466
PST130_09320_Tribe394_CLUSTER_VII	18508	3570	18372	20860	6	90384	45976	57558
PST0821_00728_Tribe394_CLUSTER_VII	718	290	732	11684	0	2682	1530	2404
PST21_17484_Tribe186_CLUSTER_VII	1654	1752	2194	8	0	6324	3830	5180
PST43_04708_Tribe186_CLUSTER_VII	27246	9874	2316	9542	0	5738	4816	5236
PST0821_09979_Tribe186_CLUSTER_VII	58	38	266	1316	2	2860	1136	1278
PST43_17354_Tribe186_CLUSTER_VII	33030	10184	15948	107626	4	22386	33192	18244
PST21_01358_Tribe186_CLUSTER_VII	83382	30078	15862	10492	2	40012	48652	42754
PST877_03038_Tribe186_CLUSTER_VII	2830	5350	1696	1734	2	11412	7138	11920
PST0821_05302_Tribe593_CLUSTER_VII	2684	12390	5014	1688	2	27500	16846	22908
PST43_01515_Tribe593_CLUSTER_VII	2460	5994	4636	13180	0	31376	18194	35262
PST43_20558_Tribe456_CLUSTER_VII	90954	27516	72948	93142	2	27702	27776	27848
PST21_10341_Tribe456_CLUSTER_VII	72298	61304	78388	68268	10	207118	115534	127320
PST877_10294_Tribe456_CLUSTER_VII	9086	2094	2138	2834	2	24376	7318	9116
PST0821_16435_Tribe456_CLUSTER_VII	674	646	640	88	0	3750	2366	3368
PST0821_05185_Tribe387_CLUSTER_VIII	10298	13666	7182	610	0	8684	8484	8510
PST130_02589_Tribe387_CLUSTER_VIII	62	36	128	32	0	156	110	420
PST21_15602_Tribe128_CLUSTER_VIII	41246	96924	12406	94898	2	16014	12610	17398
PST43_03536_Tribe128_CLUSTER_VIII	13060	10966	27244	1082	0	28352	17886	12022
PST130_12956_Tribe128_CLUSTER_VIII	35220	82878	13854	3486	6	41038	25748	30942
PST877_00017_Tribe128_CLUSTER_VIII	40	128	142	298	0	0	462	610
PST21_14919_Tribe226_CLUSTER_VIII	37238	16910	18384	16338	4	47292	33854	54662
PST0821_13679_Tribe226_CLUSTER_VIII	154	682	182	700	2	6610	1292	2712
PST43_16562_Tribe135_CLUSTER_VIII	80352	60538	26792	61752	0	45326	36016	45648
PST877_04028_Tribe237_CLUSTER_VIII	6082	5940	1654	1168	2	58528	11038	17206
PST43_20465_Tribe215_CLUSTER_VIII	5408	4288	3104	326	2	5724	6686	9094
PST21_19352_Tribe293_CLUSTER_VIII	130	42	410	64	0	698	528	1308
PST877_14212_Tribe293_CLUSTER_VIII	10944	11344	7536	3086	4	26446	22664	46322

PST0821_00217_Tribe40_CLUSTER_VIII	6348	2392	3292	886	0	4958	6546	3264
PST130_07695_Tribe40_CLUSTER_VIII	20282	26402	7894	32790	0	11934	13044	18632
PST43_16095_Tribe40_CLUSTER_VIII	19426	6888	3870	7630	0	17162	7982	10120
PST877_00240_Tribe40_CLUSTER_VIII	834	668	1454	2628	0	30	5718	3672
PST43_04767_Tribe40_CLUSTER_VIII	0	0	0	0	0	0	0	0
PST130_01536_Tribe40_CLUSTER_VIII	11952	1902	4642	17462	4	34254	18382	20446
PST21_18223_Tribe40_CLUSTER_VIII	108	48	362	50	0	678	486	1342
PST21_13429_Tribe40_CLUSTER_VIII	46944	20866	18176	25488	6	65978	33328	31880
PST0821_10041_Tribe40_CLUSTER_VIII	490	812	1172	2262	0	10412	4872	8296
PST21_15030_Tribe423_CLUSTER_VIII	2394	830	1292	1246	0	4430	3242	3128
PST877_09083_Tribe423_CLUSTER_VIII	9506	252	1116	316	0	544	4514	7176
PST43_18621_Tribe687_CLUSTER_VIII	14526	26874	12160	10666	4	18732	22250	22142
PST0821_03872_Tribe687_CLUSTER_VIII	2090	3982	2408	6220	2	21754	11438	17782
PST21_18221_tribe238	13628	3564	5870	33294	0	16068	10274	10626
PST21_18220_tribe238	1962	1340	3364	1408	2	16508	5048	4608
pst21_18506_tribe43	1974	21902	2304	11740	2	12686	7050	9806
pst21_10634_tribe43	35768	42830	32194	90182	0	33058	45838	50314
pst21_18784_tribe43	13078	2492	2928	38650	4	29096	10658	15384
pst43_08462_tribe43	12512	1530	3580	768	0	12332	9648	6546
pst0821_tribe43	1494	70	300	216	0	1640	1338	2740
pst130_14643_tribe9	3724	3094	2056	4576	2	7578	3884	6516
pst877_18061_tribe9	8662	2432	4244	1050	0	5832	5366	4158
pst21_10828_tribe9	93894	48646	48374	1038148	6	196230	150848	147896
pst43_12895_tribe9	40774	52024	33734	92314	16	238946	139392	245418
pst43_04004_tribe9	5362	7872	2540	560	4	25488	8034	6922
pst43_02877_tribe9	10328	4146	171610	9320	0	19322	11226	37682
pst130_06558_tribe9	148	38	264	120	0	28	786	1298
pst21_18427_tribe9	2782	1440	3434	3926	2	22686	10612	11108
pst21_18994_tribe9	3862	4204	2516	9020	0	13824	7940	7148
pst0821_20235_tribe9	8548	75016	11302	1570	0	64212	55732	68422
pst21_08753_tribe9	2306	72	442	6	0	680	1174	1260
pst0821_16446_tribe9	1182	798	964	4470	0	9934	4964	11000
pst0821_01106_tribe9	5646	832	1062	378	0	5062	3294	6250
Control_gene_beta_tubulin	434	5440	328	1740	0	4394	2604	5432
contol_gene_Elongation_factor	1226	598	21848	19970	0	16636	12788	10456

contol_gene_MAPK	12090	580	1924	806	0	11944	6920	15226

Appendix II

Primers designed but failed to discriminate between pathotypes

Name	Sequence (5' – 3')	Tm (°C)	Gene Targeted
Pst21_17389_War_For P1	CAAACTGCATACCTAATCCATTCCAC	66	Pst21_17389
P1_3_yRust_Pst21_17389	AATCATCAGTGGTCAATAGTCTGCTTGG	69	Pst21_17389
P2_3_yRust_Pst21_17389	TTGGAGATTTGAAAATTATCTCATGA	63	Pst21_17389
Pst21_17389_War_For P5	TGACCTCTCTCAAAAATTTATGATTT	59	Pst21_17389
Pst21_17389_War_For P6	ACTGCAAGCCAACTAGTGG	60	Pst21_17389
Pst21_17389_War_For P7	CCTAGCTGACCACACTTTG	58	Pst21_17389
Pst21_17389_War_For P8	TTTGTATGGTGGCATTTTGCATCTCT	69	Pst21_17389
Pst21_17389_War_For P9	TTTAAGGGTACAAACTTAACAT	53	Pst21_17389
Pst21_17389_War_For P10	ACAAACTTAACATAAGTCCTGATCCC	63	Pst21_17389
Pst21_17389_War_For P11	TAACACAAGTCCTGATCCCTTTGGGA	71	Pst21_17389
Pst21_17389_War_For P12	GATGAACCATTCACATCCGGCC	60	Pst21_17389
Pst21_17389_War_For P13	AAATGGCCTACTCCACGTGTTTTTTG	66	Pst21_17389
Pst21_17389_War_For P14	CCAAGCTGGCAAGGATGA	65	Pst21_17389
Pst21_17389_War_For P15	GCCTGGTTGGTTTGCATACAGTCTGA	72	Pst21_17389
Pst21_17389_War_Rev P1	GGAAATCTTTGAGAGAGGTCA	60	Pst21_17389
Pst21_17389_War_Rev P2	TCAAATCTCCAAGCTCCAAAGTAC	64	Pst21_17389
Pst21_17389_War_Rev P3	ACATGAGATAATTTTCAAATCTCCAA	63	Pst21_17389
Pst21_17389_War_Rev P4	AAAGGAGAGAGATGACAAAATTA	58	Pst21_17389
Pst21_17389_War_Rev P5	TGGGTACCCAAAGTTTGCTC	64	Pst21_17389
Pst21_17389_War_Rev P5_A	GAATCCGGGCCAAAATGTGTGTCT	72	Pst21_17389
Pst21_17389_War_Rev P5_B	CTTCCCTCCTTGCCAGCTTGGGCC	78	Pst21_17389
Pst21_17389_War_Rev P6	ACCAACCAGGCCTGTTTG	64	Pst21_17389
Pst21_17389_War_Rev P7	TGTGATAGGGAAATTAACTTG	56	Pst21_17389
Pst21_17389_War_Rev P8	GAGCTGAGGACAATGTCCAGAA	66	Pst21_17389
P1_Pst43_19594_For	ACACCGGTGATAGCCCAGAT	61	Pst43_19594
P1_Pst43_19594_Rev	TGCTGAGAGTGGAGTCTGTCG	62	Pst43_19594
P2_PST43_19594_Sol_For	TCCGACAGACTCCACTCTCA	60	Pst43_19594
P2_PST43_19594_Sol_Rev	GTGGGGTGTCGAGATTGACT	60	Pst43_19594
P3_PST43_19594_Sol_For	ACAGACCCCTCCCTCAGC	60	Pst43_19594
P3_PST43_19594_Sol_Rev	CAGGTGGGGTGTCAAGAGTT	60	Pst43_19594
P4_PST43_19594_Sol_For	GAAATATGGCCGTGACAAAGA	59	Pst43_19594
P4_PST43_19594_Sol_Rev	CTCGTGCCCAAACCTTCA	61	Pst43_19594
P5_PST43_19594_Sol_For	GAGATATGGCCGTGACAAAGA	60	Pst43_19594
P5_PST43_19594_Sol_Rev	CTCGTGCCCAAACCTTTAAC	60	Pst43_19594

PST43_12897_3yRust_For	TGAAACTGGGAACGCTACAG	59	Pst43_12897
PST43_12897_3yRust_Rev	TATTGGACGGGCATAATTCG	61	Pst43_12897
P1_ForPst21_17389_1yrust	CAAACTGCATACCTAATCCATTCCAT	67	Pst21_17389
P2_ForPst21_17389_1yrust	CATACCTAATCCATTCCATGG	63	Pst21_17389
P3_ForPst21_17389_1yrust	TCCTTTCCTATACCTTCTCATCCAAA	65	Pst21_17389
P4_ForPst21_17389_1yrust	TCCGGGGGCGCAAGCG	75	Pst21_17389
P5_ForPst21_17389_1yrust	ACACAAGTCCTGATCCATCCGG	72	Pst21_17389
P10_ForPst21_17389_1yrust	CAAGCCAAACACTCCACCCTTGG	72	Pst21_17389
P13_ForPst21_17389_1yrust	GGTTGGTTTGCATACAGTCTTG	63	Pst21_17389
P14_ForPst21_17389_1yrust	TGCATGAAACGCAGCTCGTTC	76	Pst21_17389
P2_RevPst21_17389_1yrust	CTATCAGAGTCAAAAAGCCATTT	60	Pst21_17389
P3_RevPst21_17389_1yrust	CCTCCCACCAAAATCTGGAAATTGC	72	Pst21_17389
P5_RevPst21_17389_1yrust	CCATTTTCAAGGCCGGATGTGA	71	Pst21_17389
P6_RevPst21_17389_1yrust	CCTTGCCAGCTTGGGCCA	72	Pst21_17389
P7_RevPst21_17389_1yrust	GTATGCAAACCAACCAGGCCTGTT	70	Pst21_17389
P8_RevPst21_17389_1yrust	ATCTGGGTCTTCCCAAAACAACA	67	Pst21_17389
P9_RevPst21_17389_1yrust	CATTATAAATGGAAGAATTGG	55	Pst21_17389