The evolution of virulence in wild populations of the plant pathogen *Melampsora lini*

Luke Barrett

A thesis submitted for the degree of Doctor of Philosophy

The Australian National University

School of Botany and Zoology
Faculty of Science
The Australian National University
Canberra ACT 0200
Australia

August 2008
Declaration

The research presented in this thesis comprises my own original work, except where indicated in the preface. Assistance from other workers, organisations, and funding sources is referred to in the acknowledgments.

Luke Barrett
August 2008
Preface

The work contained in the following 6 chapters was predominantly undertaken by me, expressly for the purpose of this thesis. Each of the four main chapters are all formatted for publication.

Chapter 2 is based on a published paper.


Contributions: All authors contributed conceptually. Pete Thrall helped with fieldwork. I carried out all lab and glasshouse experiments, analyses and write-up.

Chapter 3 is based on a manuscript for submission to *Molecular Biology and Evolution*.

Barrett, L.G., Thrall, P.H., Burdon, J.J., Van der Merwe, M., Linde, C.C., and Dodds, P.N. Adaptive evolution at avirulence loci across wild populations of the flax rust *Melampsora lini*.

Contributions: All authors contributed conceptually. Peter Dodds developed the gene constructs used for transient transformations and assisted with glasshouse experiments. I carried out all remaining lab and glasshouse experiments, analyses and write-up.

Chapter 4 is based on a published paper.


Contributions: All authors contributed conceptually. Pete Thrall helped with collections. I carried out all lab and glasshouse experiments, analyses and write-up.

Chapter 5 is based on a paper in press.


Contributions: All authors contributed conceptually. I carried out literature reviews and write-up.
As an appendix, I include a published paper describing the development of microsatellite primers for *M. lini*. This work was carried out during my candidature, and has direct relevance to the work described herein.


**Contributions:** Curt Brubaker provided guidance and expertise. I carried out experiments, analyses and the write-up.
Abstract

Genetic variation for virulence is widespread in both agricultural and wild pathogen populations. Despite the importance of such variation for determining disease outcomes, there is only a limited understanding of how various evolutionary and ecological processes interact to generate and maintain genetic polymorphisms underlying such pathogenic variability. In this thesis the roles of environmental, spatial and life-history heterogeneity in the generation and maintenance of pathogenic diversity in the interaction between the rust pathogen *Melampsora lini*, and its native Australian host, *Linum marginale* are examined. As demonstrated by a combination of molecular, experimental and field approaches, two genetically and geographically divergent pathogen lineages dominate interactions with the host across Australia (lineage AA and AB), one of which (AB) is of hybrid origin. It is further shown that these lineages differ in a range of key reproductive, life-history, and epidemiological traits that generate striking broad scale and regional divergence in the diversity and identity of pathogenic variation within local populations. Importantly, the key processes involved in the generation and maintenance of these among-population and regional differences involve gene flow and subsequent hybridisation, together with distinct spatial patterns of environmental heterogeneity. To further investigate how genetic polymorphisms underlying phenotypic variation in pathogen virulence are generated and maintained across broad spatial scales, patterns of nucleotide polymorphism were examined at two genes underlying pathogen virulence, the *AvrP123* and *AvrP4* avirulence loci. Results from these experiments provide strong evidence for diversifying selection and local adaptation in generating and maintaining virulence diversity in *M. lini*. Finally, it is argued that better integration of molecular, empirical and theoretical studies aimed at understanding interactions between host and pathogen life history are fundamental to understanding the genetic and demographic dynamics driving the evolution of disease. Together, these findings demonstrate the general importance of environmental, spatial and life-history heterogeneity in generating and maintaining genetic variation for pathogen virulence across broad spatial scales.
Acknowledgements

I was very fortunate to have been involved with an excellent panel of supervisors who have all contributed in many ways to this body of work and to my development as a scientist. This work would not have been possible without Jeremy Burdon’s hard work over more than 20 years in generating the enormous background of work on the Linum-Melampsora interaction. Throughout my PhD he has freely shared his wisdom and knowledge, and despite a demanding management position, has provided excellent supervision. Adrienne Nicotra has provided unwavering support and supervision, even though the direction of my research veered from hers early on in my PhD. Celeste Linde freely gave valuable advice, support and friendship. Pete Thrall, as my primary supervisor, has provided guidance, support and inspiration. He is also a good mate. We have shared many bushwalks, beers and books, and he, his wife Cindy, and his son William have been like family to me while I have been in Canberra.

This work would not have been possible without financial support. An Australian Postgraduate Award and CSIRO top-up scholarship provided a stipend for three and a half years. CSIRO provided the large proportion of the funding for the research.

CSIRO Plant Industry has proven a bountiful environment for research. Without the contribution of Caritta Eliasson, it is doubtful that this PhD would have been possible. Her hard work, friendship and evil sense of humour have all been invaluable. I truly hope that after therapy and rehabilitation she will be able to live a productive life without having to indulge her addiction to rust. I have spent countless hours in the company of David Field and Melinda Pickup, who have been good friends in both good times and bad. Peter Dodds has patiently and freely provided knowledge and advice regarding the molecular basis of pathogen virulence, and without his contribution one of the most exciting aspects of my research would not have been possible. Greg Lawrence has made a strong contribution to this project, through many stimulating conversations and the provision of numerous isolates of flax rust collected from wide-ranging areas of Australia. Curt Brubaker has been endlessly patient and forthcoming with advice on all things molecular. I have partaken in numerous
conversations regarding my project and population genetics more generally with Tony Brown. I received ongoing technical advice and had numerous stimulating discussions regarding molecular evolution with Marlien van der Merwe. Linda Broadhurst, Augusto Becerra Lopez-Lavalle, Bronwyn Matheson, Andrew Young and Brian Murray have all provided valuable help and advice at various stages of this project.

Finally, I would like to thank my family, who have never wavered in their support from afar, and my partner Jody Higgins, who has been a true champion.
Table of Contents

Declaration .................................................................................................................. i
Preface ........................................................................................................................ ii
Abstract ....................................................................................................................... iv
Acknowledgements ..................................................................................................... v
Table of Contents ......................................................................................................... vii

CHAPTER 1
Introduction: The evolutionary and ecological determinants of variation in pathogen virulence ................................................................................................................................. 1
1.1 Genetic variation for pathogen virulence: Overview and relevance ................... 2
1.2 The generation and maintenance of genetic variation for pathogen virulence .... 4
1.3 Study species .......................................................................................................... 8
1.4 Previous studies ..................................................................................................... 9
   1.4.1. The L. marginale-M. lini association .......................................................... 9
   1.4.2. The L. usitatissimum-M. lini association .................................................... 10
1.5 Overview of this research ..................................................................................... 12
1.6 Literature cited .................................................................................................... 15

CHAPTER 2
Evolutionary diversification through hybridisation in a wild host-pathogen interaction ................................................................................................................................. 22
2.1 Introduction .......................................................................................................... 23
2.2 Methods and Results ............................................................................................ 24
   2.2.1 The host-pathogen association ................................................................. 25
   2.2.2. Genetic and geographic structure of M. lini ......................................... 25
   2.2.3 Patterns of variation in virulence .............................................................. 28
   2.2.4. Patterns of variation in life history .......................................................... 30
2.3 Discussion ............................................................................................................ 34
2.4 Literature Cited ................................................................................................... 40

CHAPTER 3
Adaptive evolution of avirulence loci in wild populations of the flax rust Melampsora lini ................................................................................................................................. 43
3.1 Introduction .......................................................................................................... 44
3.2 Methods ................................................................................................................. 49
   3.2.1 Sampling .................................................................................................... 50
CHAPTER 4
Population structure and diversity across sexual and asexual populations of the pathogenic fungus Melampsora lini

4.1 Introduction
4.2 Materials and Methods
4.3 Results

4.2.1 Study sites
4.2.2 Sample collection
4.2.3 Pathotype identification
4.2.4 DNA extraction
4.2.5 AFLP
4.2.6 AvrP123 and AvrP4 amplification and sequencing
4.2.7 Data analyses

4.3.1 Variation across all isolates
4.3.2 Variation between lineages
4.3.3 Variation among populations

4.5 Literature Cited
4.4 Discussion.................................................................................................................. 105
  4.4.1. The geographic distribution of lineage AA and AB................................. 105
  4.4.2. Genetic and pathogenic structure within lineages................................. 106
  4.4.3. Genetic structuring among populations............................................... 109
  4.4.4. Conclusions............................................................................................... 112
4.5 Literature Cited..................................................................................................... 113

CHAPTER 5
Life-history determines genetic structure and evolutionary potential of host-parasite interactions.................................................................................................................................................. 117
5.1 Introduction........................................................................................................... 118
5.2 Impacts of parasite life-history on population demography and genetic structure........................................................................................................................................... 119
  5.2.1. Host specificity......................................................................................... 119
  5.2.2. Mode of reproduction............................................................................. 120
  5.2.3. Transmission and dispersal..................................................................... 121
  5.2.4. Life-cycle complexity and epidemiology............................................. 123
  5.2.5. Parasite infection strategies................................................................. 124
5.3 Impact of host life-history and spatial structure on parasite dynamics and evolution........................................................................................................................................... 125
  5.3.1. Spatial Structure.................................................................................... 126
  5.3.2. Host longevity and phenology.............................................................. 127
  5.3.3 Host resistance........................................................................................ 128
5.4 Genetic expectations for host-pathogen interactions........................................... 130
5.5 Inferring demographic and evolutionary dynamics......................................... 132
5.6 Conclusions....................................................................................................... 134
5.7 Literature cited................................................................................................... 135

CHAPTER 6
Synthesis and Conclusions......................................................................................... 140
6.1 Overview: *M. lini* in the geographic mosaic.................................................... 141
6.2 The ecological and evolutionary determinants of virulence polymorphisms in *M. lini*.................................................................................................................................................. 142
  6.2.1 Conclusions.............................................................................................. 147
6.3 Literature cited................................................................................................... 148

Appendix...................................................................................................................... 150
CHAPTER 1

Introduction: The evolutionary and ecological determinants of variation in pathogen virulence
1.1. Genetic variation for pathogen virulence: Overview and relevance

*Infectious disease has a major influence on the demography of human, plant and animal populations. Understanding how genetic variation in pathogen virulence and aggressiveness is generated and maintained is of central importance for understanding patterns of disease emergence (Friesen et al. 2006; Brault et al. 2007) and epidemiology (Real et al. 2005). Despite this, few empirical studies have examined how population level ecological and evolutionary processes influence the generation and maintenance of pathogenic variation. The general aim of this thesis is to investigate the roles of environmental, spatial and life-history heterogeneity in the generation and maintenance of polymorphisms underlying phenotypic variability in pathogen virulence.*

Pathogens – here broadly defined as disease causing organisms – are ubiquitous and influence either directly or indirectly almost every possible level of biological organisation (Thompson 2005). The evolutionary and ecological impacts that pathogens have on their hosts is strongly dependent on a range of factors that influence patterns of disease incidence and prevalence. This includes environmental factors, host and pathogen life history (e.g. longevity, mating system, transmission), as well as genetic variation for traits that determine infection – the forces that drive local and regional patterns of coevolution. It is generally accepted that characterising patterns of genetic variation in host resistance and pathogen virulence or aggressiveness\(^1\) is of central importance to understanding the dynamics and distribution of disease (Johnson 1961; Person 1966; Hill 1998; Ellis *et al.* 2000; de Meaux and Mitchell-Olds 2003). Thus, correlations between individual host and/or pathogen genotypes and disease outcomes have been demonstrated for a number of host-pathogen interactions (Flor 1955; Chin and Wolfe 1984a; Buitkamp *et al.* 1996; Paterson *et al.* 1998; Stahl *et al.* 1999; Lockett *et al.* 2001; Williams-Blangero *et al.* 2002; Ong and Innes 2006). Furthermore, high levels of variation have been directly documented for a range of host resistance (Bergelson *et al.* 2001; Piertney and Oliver 2006) and pathogen virulence loci (Hacker and Kaper 2000; Catanzariti *et al.* 2007)

\(^1\) Throughout this thesis I refer to virulence and aggressiveness in the standard plant pathology context. Virulence is a measure of infectivity, and refers to the ability of a pathogen to overcome a given host resistance gene. Aggressiveness refers to the severity of disease symptoms caused by infection (equivalent to the term ‘virulence’ as used by researchers working on human and animal diseases).

Chapter 1: Introduction
important in determining the outcomes of disease. Nevertheless, there is only a limited understanding of how these polymorphisms are generated and maintained within natural communities.

Genetic variation for host resistance and pathogen virulence are likely to be critical factors influencing disease epidemiology and the emergence and spread of new diseases. For example, at the species level, the inter-specific acquisition of novel genes determining virulence has been implicated in the emergence and spread of recently evolved pathogen species (Prentice et al. 2001; Friesen et al. 2006). At the population level, some work has shown negative relationships between the overall diversity of hosts and levels of parasitism (Black et al. 1995; Meagher 1999; Coltman et al. 2001; Acevedo-Whitehouse et al. 2003). Chin and Wolfe (1984b) demonstrated that increased host resistance diversity in crop situations is correlated with lower rates of disease spread and overall prevalence within experimental host plant populations. Recent work on a native plant-pathogen interaction demonstrated that increased resistance diversity as well as overall levels of resistance within host populations was correlated with lower disease prevalence during epidemics (Thrall and Burdon 2000). However, it is only in a very few cases that pathogen population genetic variation has been implicated in rates of epidemic spread of disease (e.g. the relatively slow spread of HIV in Cameroon may be related to levels of genetic variability in particular viral groups; Heyndrickx et al. 2000). Overall, there has been a paucity of studies that have directly examined causal links between pathogen population genetic structure and disease incidence and prevalence.

The best characterised model of the genetic interaction between host and pathogens is defined by the gene-for-gene framework (Flor 1956), under which major genes for host resistance ($R$) recognise specific pathogen effector genes, known as avirulence ($Avr$) genes. While originally defined using Linum usitatissimum (cultivated flax) and its rust pathogen Melampsora lini, gene-for-gene interactions control the resistance of many plants to a wide variety of pathogens including fungi, bacteria, viruses, nematodes and some insects (Thompson and Burdon 1992; Parker 1994; Clay and Kover 1996). The gene-for-gene paradigm has shaped much of our current thinking about host-pathogen coevolution in both animal and plant systems. In particular the presence of corresponding $R$ and $Avr$ genes in host and pathogen populations implies
the possibility of coevolution driven by selection pressure on the pathogen to escape recognition by host R gene products, and concomitant pressure on the host to respond to new virulent strains of the pathogen. However, we have little knowledge of how population level ecological processes influence the generation and maintenance of variation at host resistance and pathogen avirulence loci. Likewise, we do not know how variation at the molecular level influences the population and genetic dynamics of disease in nature. Increasingly, this lack of knowledge has led to calls for an integrated approach to the study of pathogenic organisms, incorporating both ecological and evolutionary processes (Ewald 1994; Schrag and Wiener 1995; Tibayrenc 1998; Real et al. 2005; Mitchell-Olds et al. 2007).

1.2. The generation and maintenance of genetic variation for pathogen virulence

Generally, variability at resistance and virulence loci within species and populations is considered to be generated and maintained by selective forces associated with antagonistic interactions between hosts and pathogens (Jeffery and Bangham 2000; Stahl and Bishop 2000; Bergelson et al. 2001; Mackey et al. 2003; Borghans et al. 2004; Dodds et al. 2006). Different hypotheses have been proposed to explain how selection might promote the maintenance of polymorphisms at host R and pathogen Avr loci, ranging from equilibrium (balancing selection), to non equilibrium (transient polymorphism, past selection) conditions and involving direct or indirect selection (multilocus associations) (reviewed in Parker 1992). Negative frequency dependent selection, where rare pathogen (or host) genotypes have higher fitness when rare, is a common assumption in models of host-parasite interactions, and can help to explain the maintenance of resistance and virulence polymorphisms within populations (Gillespie 1975; Clarke 1976; Anderson and May 1982). However, empirical data demonstrating co-adaptive changes in resistance and virulence structure with respect to disease dynamics are rare. Instead, results from the only well studied example of a natural gene for gene system, the Linum-Melampsora interaction (Burdon and Jarosz 1991; Burdon and Jarosz 1992; Burdon and Thrall 2000), do not match the classical scenario predicted by theory based on single populations. Rather than tight frequency-dependent couplings of resistance and virulence, real-world systems show
considerable potential for stochastic and non-equilibrium dynamics, emphasizing the need for studies that encompass underlying ecological and demographic dynamics across multiple host and pathogen populations.

Many host species, particularly in natural ecosystems, exist as interacting groups of small, geographically and genetically differentiated populations, subdividing pathogen populations into relatively small, discrete units (Burdon 1992). Such spatial and environmental heterogeneity among pathogen populations is likely to strongly influence spatial patterns of disease incidence and persistence (Thrall and Burdon 1997), with populations of many species shown to undergo frequent local extinctions and recolonisations (Antonovics et al. 1994; Ericson et al. 1999; Thrall et al. 2001; Smith et al. 2003; Laine and Hanski 2006). Thus, advances in the understanding of disease dynamics in pathogenic species require a metapopulation approach that accounts for the genetic and demographic consequences of significant levels of local extinction and among-population recolonisation processes.

Within such a metapopulation framework, variation in different micro-evolutionary forces among demes may act to generate and maintain resistance and virulence polymorphisms. Local host populations, particularly in natural ecosystems, often vary spatially in the identity and diversity of resistance genotypes present (Jarosz and Burdon 1991; Bevan et al. 1993; Laine 2004; Laine and Hanski 2006), and local adaptation of pathogens to their hosts has been demonstrated as a strong driver of pathogen population genetic structure in a number of host-pathogen interactions (Gandon et al. 1996; Gandon 1998; Lively and Dybdahl 2000; Greischar and Koskella 2007). It has also been suggested that there is strong potential for non-selective factors, such as random genetic drift, founder events and selection on linked traits, to influence host and pathogen evolution and drive genetic divergence among demes (Parker 1991; Burdon and Thompson 1995; Salathé et al. 2005). Within this framework, rates of genetic recombination (Maynard Smith et al. 1993; Milgroom 1996), and the degree to which pathogen dispersal occurs at local scales relative to the metapopulation as a whole (Thrall and Burdon 1999; Thrall and Burdon 2002; Greischar and Koskella 2007), will further influence how variation is maintained and distributed within the metapopulation.
Sexual reproduction is generally thought to play a significant role in the generation of variability in pathogen species because it results in new gene combinations. Indeed, several studies demonstrate that pathogen populations that undergo sex generally exhibit higher levels of genotypic and pathogenic diversity than species that rely exclusively on asexual reproduction (Burdon and Roelfs 1985; Milgroom 1996; Liu and Kolmer 1998; Bourassa et al. 2007). However, reproductive modes (and the degree of outcrossing) are complex traits that may reflect interactions among a range of environmental, genetic, structural and demographic factors (Barrett and Eckert 1990; Kohn 1995; Milgroom 1996). Thus, while sexual reproduction is common among pathogen species, many species and/or populations self-fertilise, can reproduce only clonally, or exhibit both clonal and sexual reproductive modes that vary in time and space.

The reproductive mode of pathogens can have a strong influence on the evolutionary outcomes of interactions between pathogens and their hosts (Fox et al. 1996; Lively et al. 2004). The Red Queen Hypothesis (RQH) predicts that frequency dependent selection imposed on hosts by co-adapted pathogens favours rare host genotypes, because local parasites become adapted to the most common genotypes within pathogen populations. If infection reduces host fitness, then common host genotypes should decrease in frequency and be replaced by previously rare and less infected host genotypes. Such a change in the host population should result in selection on the pathogen to infect the newly common host genotypes. Under this model, sexual reproduction in the host is hypothesised to be an adaptation against rapidly evolving parasites because it enables hosts to respond quickly to the appearance of novel pathotypes through recombination of resistance genes. (Hamilton 1980; Hamilton et al. 1990; Clay and Kover 1996). By extension the RQH also predicts that parasites infecting sexual hosts should be genetically more variable than those infecting asexual hosts because recombination of resistance genes will drive frequency-dependent selection on the corresponding virulence genes of parasites (Ooi and Yahara 1999). Consistent with the RQH, some studies suggest there may be broad relationships between sexual reproduction and parasitism. For example, there is a positive correlation between the number of fungal pathogen species known to infect a host plant and the outcrossing rate of the host (Busch et al. 2004), suggesting that host and/or pathogen mating systems have the potential to influence disease dynamics in...
natural populations. Similarly, biological control is generally more effective on asexually reproducing weed species than on sexually reproducing weeds (Thrall and Burdon 2004). At a finer scale, Ooi and Yahara (1999) found higher genetic diversity in populations of viruses infecting sexual compared to asexual host plant populations.

The geographic mosaic theory of coevolution (Thompson 1999b; Thompson 2005) predicts that, depending on local conditions, the intensity and direction of reciprocal selection can vary geographically, resulting in different evolutionary trajectories in different locations. Geographic variation in selection trajectories has been demonstrated for some antagonistic species interactions (Benkman 1999; Lively 1999; Brodie et al. 2002; Zangerl and Berenbaum 2003), but the scale at which the selection mosaic is formed will vary both within and among different systems. Characterizing interactions at a local scale may highlight immediate evolutionary processes, such as local adaptation (Thrall et al. 2002; Laine 2006) and frequency dependent selection (Dybdahl and Lively 1998). In contrast, studying interactions across broad geographic and species ranges should provide an understanding of higher level macro-evolutionary processes operating in host and pathogen species as they respond to each other across diverse environmental and ecological conditions (Burdon et al. 1999; Thompson 1999a).

Plant-pathogen associations are among the most amenable of biotic interactions in which coevolution can be studied, and in which genetic and molecular components important to the interaction can be characterised. Sampling can readily be done across a range of defined spatial scales (from individuals through populations to geographical regions), making it possible to address major issues relating to host pathogen coevolution. In contrast to human diseases or many animal systems, experimental determination of the genetics of pathogen virulence and aggressiveness (and host resistance) is relatively straightforward. Phenotypic variation within natural populations for both host resistance (Burdon 1987; Laine 2004) and pathogen virulence (Lebeda and Petrzelova 2004; Meyer et al. 2005) is common and well documented. Furthermore, plant pathogens show the ability to rapidly evolve new virulence (Steele et al. 2001; Abu-El Samen et al. 2003; Thrall and Burdon 2003; Jimenez-Gasco et al. 2004) with pathogens occurring on crop species often quickly overcoming novel host resistance genes introduced through plant breeding (Steele et
al. 2001; McDonald and Linde 2002; Sprague et al. 2006). Importantly, the underlying genetic models governing the specificity of plant host-pathogen interactions are often well characterized, having been shown to be both polygenic (Geiger and Heun 1989; Wilson et al. 2001; Kliebenstein et al. 2002; Thrall et al. 2005), or influenced by single genes with major effects (Flor 1956; Thompson and Burdon 1992; Burdon 1994; Rossi et al. 1998).

### 1.3. Study species

*Melampsora lini* is a hemibasidiomycete, order Uredinales, with a probable chromosome number of \( n = 5 \) or \( 6 \) (Kapooria 1973). It is a biotrophic, macrocyclic species with all five spore stages occurring on the same host species (autoecious). The dikaryotic mycelium growing in susceptible host plants produce dikaryotic urediospores which form the asexual repeating stage of the life-cycle responsible for disease epidemics during host growing seasons. Under the right conditions these can lead to the formation of telia, the precursor stage for going through sexual recombination. Subsequently, meiosis takes place during the dormant teliospore phase, giving rise to haploid basidiospores. The basidiospores re-infect living host tissue, producing haploid pycniospores. The dikaryophase is re-initiated by the transfer of haploid pycniospores between pycnia of different mating types. F1 urediospores from a cross between two divergent pathotypes therefore carry one haploid nucleus from each parent (Fig. 2.1).

*M. lini* has a wide host range in the genus *Linum*, being reported to occur on numerous species occurring in Europe, North America and New Zealand. In Australia, it infects the sole species in the genus *Linum marginale* (Lawrence et al. 2007). The native Australian flax species *L. marginale* is a perennial herb endemic across a broad swathe of southern Australia. Populations may vary distinctly with regard to phenological patterns. In dry, inland areas which experience a strong summer drought, plants either die during the summer months or survive as underground rootstock. Rainfall during the autumn induces new growth, and plants grow throughout the winter and flower in spring. In cooler, mountainous areas, plants are dormant during the winter months, surviving with a few short shoots. The coming
of spring induces new growth, with plants flowering in mid to late summer (Burdon et al. 1999).

1.4. Previous studies

1.4.1. The L. marginale-M. lini association

The interaction between the fungal rust pathogen *Melampsora lini* and its wild host plant *Linum marginale* has been extensively utilised to investigate the epidemiology and evolutionary dynamics in a wild host-pathogen system. Lawrence and Burdon (1989) examined the pathogenic association occurring between *L. marginale* and *M. lini* and found it to be a longstanding, naturally occurring interaction related to, but evolutionarily isolated from that between the pathogen and its agriculturally cultivated host (*L. usitatissimum*). The interaction between *M. lini* and *L. marginale* follows a gene-for-gene model, where major genes for pathogen virulence interact with major genes for plant resistance (Burdon 1994). Host and pathogen populations are frequently composed of many different resistance and virulence phenotypes (Burdon and Jarosz 1991), and *M. lini* has the potential to impose strong selection on host resistance, causing 60-80% reductions in population size during severe epidemics (Jarosz and Burdon 1992). These studies provide clear evidence of the potential for coevolution between *M. lini* and *L. marginale* (i.e. there is a genetic basis for resistance, there is variation among host and pathogen individuals for resistance and virulence respectively, and pathogen attack negatively affects host fitness). However, studies of temporal variation in patterns of virulence across multiple natural populations (Burdon and Jarosz 1991; Burdon and Jarosz 1992) show considerable departure from frequency-dependent cycling of resistance and virulence predicted under classical theory (Person 1966; Jayakar 1970; Hamilton 1980), demonstrating strong potential for stochastic processes to influence the evolution of pathogen virulence.

Further experimentation and demographic disease monitoring across multiple host and pathogen populations has revealed marked differences in the virulence structure of pathogen populations, and strong evidence for local adaptation of *M. lini* to *L. marginale* host populations (Thrall et al. 2002). Evidence across multiple populations
also indicates a strong positive relationship between average pathogen virulence and average host resistance, with more susceptible host populations harbouring pathotypes of lower virulence but high aggressiveness (Thrall and Burdon 2003). In contrast, resistant host populations are dominated by pathogens able to overcome a broader range of resistance genes – this correlation is at least partly driven by what appears to be a trade-off between virulence and aggressiveness. Overall, this work has demonstrated considerable variability in the resistance and virulence structure of host and pathogen populations and provides clear evidence for the selective forces that host and pathogen may exert on each other.

Much of this earlier work has focused on within and among population processes within a single metapopulation. At larger geographic scales, an understanding of the potential for factors such as genetic divergence, regional patterns of population structure, gene flow and environmental heterogeneity to influence coevolutionary dynamics remains limited. Previous work revealed significant differences in the partitioning of host resistance within and among populations between adjacent biogeographic regions (Burdon et al. 1999). The two regions differ also markedly in environment, host phenology, host outcrossing rates and disease epidemiology (Burdon et al. 1999). The consequences of such differences in genetic structure and disease dynamics have yet to be assessed but may well reflect variation in intensity and persistence of selective pressures leading to distinct hot and cold spots of coevolutionary activity (Thompson 1994; Thompson 1999b; Thompson 2005).

Examining the potential for these broad, regional scale dynamics to influence coevolutionary interactions in the *L. marginale-M. lini* interaction has been one of the main goals of this PhD.

1.4.2. The *L. usitatissimum*-M. *lini* association

One of the key strengths of the *L. marginale-M. lini* interaction as a model system is its links to the genetically well characterised *L. usitatissimum – M. lini* interaction. The interaction between cultivated flax (*Linum usitatissimum*) and its rust pathogen (*Melampsora lini*) is a powerful genetic model for understanding race-specific disease resistance in plants. At least 30 interacting *R-Avr* gene pairs have been distinguished at the genetic level (Flor 1971; Lawrence et al. 1981; Islam and Mayo 1990). For the pathogen alone, this involves a minimum of 19 independently segregating *Avr* loci.
(Lawrence et al. 2007). Recently, several families of flax rust Avr genes have been identified at the molecular level in the interaction with *L. usitatissimum*, either through map based cloning, or searching for secreted proteins in rust haustoria (specialised infection structures that penetrate the plant cell wall and are the primary site of pathogen infection). These include the *AvrL567* gene family, whose members correspond to the *L5, L6* and *L7* resistance genes (Dodds *et al.* 2004), the *AvrP123* and *AvrP4* genes, which correspond to the *P, P1, P2, P3* and *P4* resistance genes respectively, and *AvrM*, corresponding to the *M* resistance gene (Catanzariti *et al.* 2006). The avirulence function of these alleles has been confirmed through transient genetic transformation experiments (within host leaves), whereby *in planta* expression of individual *Avr* alleles in a host carrying the specific corresponding *R* gene or allele resulted in a hypersensitive response and cell death on the transformed leaf. In contrast, *Avr* gene alleles expressed in host plants not carrying the specific corresponding *R* gene or allele displayed no reaction (Dodds *et al.* 2004; Catanzariti *et al.* 2006; Dodds *et al.* 2006). Two of these loci (*AvrL567* and *AvrM*) maintain multiple linked gene copies, whereas the other two (*AvrP123* and *AvrP4*) are single copy loci. While the exact function of these genes is not known, all encode novel, small secreted proteins that are generally thought to function as pathogenicity effectors that play a role in establishing infection of the host (Catanzariti *et al.* 2007).

Although only a limited number of isolates have been sequenced, the *Avr* loci in *M. lini* are highly polymorphic, and multiple virulence polymorphisms are maintained via amino acid differences in the expressed *Avr* protein. Functional *Avr* genes are typically conserved, and all allelic copies expressed (Dodds *et al.* 2006). The highly polymorphic nature of these genes raises the question of how these genes have evolved different recognition capabilities, and how the variants are maintained in populations. Analysis of nucleotide variation at the *AvrL567* and *AvrP123* loci shows a significant excess of non-synonymous to synonymous changes, indicating that diversifying selection is likely acting on these genes (Dodds *et al.* 2004; Catanzariti *et al.* 2006). Evolution at *Avr* loci in *M. lini* may thus have been driven by the need for the pathogen to escape recognition by host *R* genes, but at the same time retain the pathogenicity function of the effector proteins.
This previous work on *M. lini* and its host species provides a solid foundation on which to study the evolution of pathogen virulence. The availability of well-characterised host differential lines (facilitating the reliable identification of pathogen virulence phenotypes), and cloned *Avr* genes, allowing detailed analyses of the pathogenic structure of naturally diseased populations, makes *Melampsora lini* a good study species for developing a clear understanding of the role of spatial and environmental variation in influencing the evolution of pathogen virulence and host-pathogen coevolution. However, these observations of *Avr* gene evolution in *M. lini* are based thus far on rust isolates collected from cultivated flax. The role of these genes as targets for *L. marginale* resistance proteins is unknown. Therefore, one of the goals of this current work will be to assess whether *Avr* gene sequence variation in the rust or recognition in *L. marginale* has the potential to contribute to coevolution, and disease outcomes in natural populations.

### 1.5. Overview of this research

The focus of this thesis work is to understand how genetic variation for pathogen virulence is generated and maintained across broad spatial scales. Genetic variation for the ability to overcome host resistance is widespread in populations of microbial pathogens, yet the ecological and evolutionary forces involved in generating and maintaining this variation are not well understood. Through the application and integration of a range of molecular, experimental and field approaches, I have examined how broad scale patterns of environmental heterogeneity, variation in host and pathogen life-history, and local patterns of genetic drift, sexual recombination, and natural selection might influence the evolution of genetic polymorphisms for virulence across natural populations of the plant pathogen *Melampsora lini*.

**Chapter 2** seeks to establish an underlying genetic and phylogeographic framework for understanding broad-scale evolutionary dynamics in *M. lini*. Using a sample of isolates representing the geographic range of *M. lini* in Australia, the genetic and pathogenic structure of the pathogen is characterised, revealing two genetically, pathogenically and geographically divergent *M. lini* lineages (lineage AA and AB), that dominate interactions with the host across Australia. In addition, patterns of genetic polymorphism suggest a hybrid origin for lineage AB, a lack of sexual
reproduction in lineage AB, and a mixed sexual/clonal lineage AA. Further experiments demonstrate that the geographic divergence of the two lineages of *M. lini* in Australia is related to variation among lineages in key life-history life traits and response to environmental conditions.

**Chapter 3**, examines how polymorphisms are generated and maintained at gene loci (*AvrP123* and *AvrP4*) underlying virulence determination in *M. lini*. Evidence for adaptive evolution is often reported in comparative studies of *R* and *Avr* gene sequences. However, the evolutionary dynamics that determine how variation is generated and maintained remain little studied, particularly in natural populations. In this chapter patterns of sequence polymorphism at *AvrP123* and *AvrP4* are characterised across a range of spatial scales, with discussion focusing on how variation at these loci might be generated and maintained with respect to gene flow and patterns of selection among local populations. In addition, experiments that determine whether the *AvrP123* and *AvrP4* genes are recognised by the native Australian host *L. marginale* are described. Strong evidence for positive diversifying selection driving the generation and maintenance of overall patterns of allelic variation was found at broad spatial scales. Among local populations, strong genetic differentiation among populations and regions is shown for both *AvrP123* and *AvrP4*, in addition to considerable variation in haplotype diversity within populations. These results suggest that positive diversifying selection acting across a broad range of scales may generate and maintain virulence diversity in pathogen populations.

In **chapter 4** the population genetic structure of *M. lini* across two biogeographic regions is described. The two regions differ markedly in environment, host resistance structure, host phenology and pathogen epidemiology, and results from chapter 2 suggest that the distributions of lineage AA and AB are likely to be disjunct among these two regions. This scenario offers an excellent opportunity to examine how life-history differences among the two pathogen lineages, particularly in their mode of reproduction, might influence the generation and maintenance of genotypic diversity and the potential for divergent coevolutionary dynamics among two geographic regions. Through the application of AFLP markers, allelic data from two avirulence genes, and information on phenotypic variation in virulence, it is shown that lineages AA and AB are have strongly disjunct distributions, with genotypic and phenotypic...
diversity significantly higher in (AA) than in asexual (AB) pathogen populations. Within this framework, an attempt is also made to evaluate the relative importance of local directional selection, fluctuating selection and genetic drift to the evolution of genetic and phenotypic divergence among populations, by comparing estimates of genetic structure among AFLP, *Avr* and phenotypic virulence markers.

**Chapter 5** is an essay review examining how heterogeneity in parasite and host life-history traits may affect the genetic structure of parasite populations. Information concerning the genetic structure and diversity of parasite populations is commonly used to draw inferences regarding the evolutionary history and potential of those species. However, relatively little consideration has been given to how heterogeneity in parasite and host life-history traits may affect the pathogenic or genetic structure of parasite populations, and thus our ability to correctly interpret population genetic signatures. The aim here is to demonstrate that recognition of the importance of life-history and spatial structure in influencing the demography of populations of parasitic organisms is necessary in order to gain a clear understanding of the evolutionary potential of parasite populations and their demographic and evolutionary histories.

**Chapter 6** serves as an overall summary and places the results within a general evolutionary and coevolutionary framework. It is argued that this work has yielded several key results with important implications for understanding how spatial, environmental, reproductive and life-history variation influence the evolution and maintenance of pathogen virulence across geographic landscapes.
1.6. Literature cited


Chapter 1: Introduction


CHAPTER 2

Evolutionary diversification through hybridisation in a wild host-pathogen interaction
2.1 Introduction

Antagonistic coevolution between species is a dynamic and powerful process with the potential for organizing patterns of diversity across broad geographic landscapes (Benkman 1999; Burdon et al. 1999; Brodie et al. 2002; Burdon et al. 2002; Zangerl and Berenbaum 2003; Enjalbert et al. 2005; Thompson 2005). Such species interactions have the potential to shape community dynamics (Thrall et al. 2007) drive the evolution of diversity in host resistance and pathogen virulence (Frank 1993) and promote the evolution and maintenance of sexual reproduction (Hamilton 1980; Lively and Jokela 2002). Coevolving interactions typically occur across heterogeneous environments and encompass a range of different spatial scales, within which species exist as groups of genetically differentiated populations (Thompson 2005). Characterizing interactions at a local scale may highlight immediate evolutionary processes, such as local adaptation (Thrall et al. 2002; Laine 2006) and frequency dependent selection (Dybdahl and Lively 1998). In contrast, studying interactions across broad geographic and species ranges should provide an understanding of more macro-evolutionary processes operating in host and pathogen species as they respond to each other across diverse environmental and ecological conditions (Burdon et al. 1999; Thompson 1999a).

Interactions between host plants and their pathogens are ideal systems in which to investigate coevolutionary scenarios. Phenotypic variation within natural populations for both host resistance (Burdon 1987b; Laine 2004) and pathogen virulence (Lebeda and Petzelova 2004; Meyer et al. 2005) is common and well documented (where virulence is defined as the ability of a pathogen to infect any given host). Furthermore, plant pathogens show the ability to rapidly evolve new virulence (Steele et al. 2001; Abu-El Samen et al. 2003; Thrall and Burdon 2003; Jimenez-Gasco et al. 2004) with pathogens occurring on crop species often quickly overcoming novel resistance genes introduced into their host through plant breeding (Steele et al. 2001; McDonald and Linde 2002; Sprague et al. 2006). Importantly, the underlying genetic models governing the specificity of plant host-pathogen interactions are often well characterized, having been shown to be both polygenic (Geiger and Heun 1989;
Wilson et al. 2001), or influenced by single genes with major effects (Flor 1956; Thompson and Burdon 1992; Burdon 1994).

In this study the geographic structure of the rust pathogen *Melampsora lini* in Australia is examined for a range of genetic, pathogenic and life history traits over much of its hosts' natural range in Australia. The interaction between the wild host plant *Linum marginale*, widely distributed throughout southern, temperate areas of Australia, and *M. lini* is well suited to investigating processes influencing disease dynamics and the evolution of host resistance and pathogen virulence. The interaction between these two species follows the 'gene-for-gene' model (Flor 1956), where the genetic determination of host resistance (Burdon 1994) and pathogen virulence (Flor 1955) is the best understood of any natural plant-pathogen interaction, and where variation for both host resistance and pathogen virulence is widespread and well documented in interacting populations (Burdon et al. 1999; Thrall et al. 2002). A number of earlier investigations have revealed spatially structured variation in host population genetic structure, host resistance and pathogen virulence at scales ranging from individual populations to the entire range of the interaction in Australia (Burdon et al. 1999; Burdon et al. 2002; Thrall et al. 2002; Thrall and Burdon 2003).

Together, these studies suggest that broad scale variation in environmental conditions, spatial population structure and host life history may be important in generating regional structure in genetic and pathogenic variation in populations of *M. lini*. However, little was known about how underlying genetic variability was distributed across the broader landscape in *M. lini*, and how this might potentially influence coevolutionary dynamics within the *L. marginale-M. lini* interaction. Using a combination of molecular and experimental approaches to characterise pathogen isolates that represent the scope of the distribution of *M. lini* in Australia, our goal for the research described here was to examine how environmental and spatial heterogeneity influences the development of genetic structure and regional adaptation in *M. lini*.

### 2.2. Methods and Results
2.2.1 *The host-pathogen association*

Linum marginale is a perennial herb endemic to southern, temperate areas in Australia. Populations may vary distinctly with regard to phenological patterns. In dry, inland areas which experience a strong summer drought, plants either die during the summer months or survive as underground rootstock. Rainfall during the autumn induces new growth, and plants grow throughout the winter and flower in spring. In cooler, mountainous areas, plants are dormant during the winter months, surviving with a few short shoots. The coming of spring induces new growth, with plants flowering in mid to late summer.

Melampsora lini is an autoecious, macrocyclic rust species that clonally produces large numbers of dikaryotic urediospores during epidemics. Meiosis and haploidisation take place during the dormant teliospore phase. The dikaryophase is initiated by the transfer of haploid pycniospores between pycnia of different mating types. F1 urediospores from a cross between two divergent pathotypes therefore carry one haploid nucleus from each parent (Fig. 2.1). Like most other rust pathogens, *M. lini* has the potential to be aerially dispersed over long distances.

2.2.2 *Genetic and geographic structure of M. lini*

Previous studies have successfully used a combination of AFLP and microsatellite markers to investigate population genetic structure and infer the demographic and evolutionary history of pathogenic rust fungi (Steele et al. 2001; Enjalbert et al. 2005). Here, a combination of nine microsatellite loci and four AFLP primer combinations are applied to 39 isolates of *M. lini* that were collected from *L. marginale* at known locations across the geographical range of the species in Australia. Methods whereby these isolates where collected, purified and bulked up are provided in Burdon et al. (2002). A further 4 isolates collected from cultivated flax (*Linum usitatissimum*) in North America (3 isolates) and South America (1 isolate) were also used to provide a reference in the cluster analysis. These isolates were imported into Australia as urediospores suspended in ethanol. Total genomic DNA was extracted from 100 mg of urediospores using a DNeasy plant mini kit (Qiagen), following standard protocols. The genotype of each isolate was determined using 9 microsatellite loci (Barrett and Brubaker 2006). AFLP genotyping was performed as described by Becerra Lopez-Lavalle and Brubaker (Becerra Lopez-Lavalle and Brubaker In press) except that 4
selective primer combinations were used (E-TA/M-GTC, E-AAC/M-GC, E-GC/M-GC, E-GC/M-GT). Cluster analysis was performed for both AFLP and microsatellite data with WINDIST (Yap and Nelson 1996) using Dice's similarity coefficient and UPGMA clustering methods. The UPGMA tree was constructed using NT-SYSpc v 2.11 (Rohlf 1993). Bootstrap values (10 000 replicates) for each branch were calculated using WINBOOT (International Rice Research Institute). AMOVA and Mantel tests were conducted on the combined molecular datasets using GenAlEx (v. 6).

Figure 2.1. The full life-cycle of *Melampsora lini* on *Linum marginale*. Multiple asexual cycles (shaded portion of the diagram) can occur during a single disease epidemic. The seasonality of this life-cycle is variable in Australia according to local environmental conditions. In areas where there is a strong summer drought, epidemics are likely to occur during the winter and early spring. In alpine areas, epidemics occur during the late spring and summer.

The microsatellite and AFLP datasets reveal that two genetically distinct lineages of *M. lini* occur on *L. marginale* (hereafter referred to as AA and AB). Both lineages are clearly distinct from the samples collected from *L. usitatissimum* (Fig. 2.2a). Given the broad geographic scale of sampling, microsatellite and AFLP variation across isolates.
collected from *L. marginale* was generally low. Across lineages, 43% (124 of 291) of AFLP marker loci were variable. For the 9 microsatellite loci, a total of 20 alleles (mean number of alleles per locus = 2.2) were recorded. Most of this variation described differences between lineages AA and AB, with AMOVA partitioning 88% of all genetic variation among lineages, and only 12% within. Mantel testing revealed no relationship between genetic and geographic distance within lineages (AA: *p*=0.38; AB: *p*=0.41).

![Figure 2.2](image)

**Figure 2.2.** *M. lini* lineages and their spatial distribution. (A) UPGMA cluster analysis of combined AFLP and microsatellite genetic similarity of 39 Australian isolates of *M. lini* collected from *L. marginale* and four isolates collected from cultivated flax, *L. usitatissimum* (uppermost cluster). Bootstrap values (>50%) from analysis of 10,000 replicates are shown above nodes (B) Geographic distribution of isolates of *M. lini* assigned to lineage AA (open circles) and lineage AB (closed triangles).

Genetic variation within the two pathogen lineages was partitioned differently, with the pattern observed for lineage AB being indicative of a hybrid origin. Observed microsatellite heterozygosity within lineage AA was very low, reflecting the low levels of polymorphism and allelic richness in this lineage (mean observed heterozygosity = 0.01; mean number of alleles per locus = 1.2; 2 of 9 loci...
polymorphic). In contrast, observed heterozygosity within lineage AB was high, with 7 microsatellite loci fixed for heterozygosity (mean observed heterozygosity = 0.88, mean number of alleles per locus = 2; 8 of 9 loci polymorphic). For all microsatellite loci polymorphic in lineage AB, the two lineages shared one common allele, for which lineage AA was typically homozygous. Importantly, lineage AB isolates were consistently heterozygous for a second allele not found in lineage AA. AFLP fingerprints exposed analogous patterns, with nearly all markers found in lineage AA also found in lineage AB, while more than one third of all AFLP markers were unique to lineage AB (Table 2.1). Nuclear staining confirmed that both lineages were dikaryotic. These results are consistent with lineage AB having been formed by genetic exchange between individuals belonging to lineage AA and a second unidentified lineage carrying the putative B genome (termed BB).

The geographic locations for the *M. lini* collections were mapped using ARCVIEW vers. 9.1. Environmental variables (annual mean precipitation (mm); annual mean temperature (°C); annual mean radiation (Mj/m²/day) for each of the sites were collected using the climatic software ANUCLIM (see Bickford and Laffin (2006) for details of methodology) and tested for significant differences between the two lineages using non-parametric, one-way analyses of variance (SAS 2000, vers. 8.2, proc npar1way). The results indicated that the two lineages have geographically distinct distributions, with lineage AB more likely to occur in southerly and coastal areas (Fig. 2.2b), reflecting an apparent preference for significantly cooler and wetter locations than lineage AA (Table 2.1). Intensive sampling across New South Wales confirms this general distribution at regional and local population levels (LG Barrett, unpublished). However, the occurrence of both lineages in a single South Australian population (this study), and the overlapping ranges of lineages in transitional environments (Fig. 2.2b) demonstrates the potential for interaction and subsequent competition between lineages within local populations.

### 2.2.3 Patterns of variation in virulence

Potential differences between lineages in their ability to overcome host resistance were evaluated by assessing 35 isolates (19 AA, 18 AB) for infection type responses induced on a differential set of 44 *L. marginale* lines. The sampling distribution of the host lines is representative of the range of the species in Australia, and approximates
the sampling distribution of the *M. lini* isolates used. The sampling locations of the host and pathogen lines are described in Burdon et al (2002). Details of the procedures for inoculation and assessment of infection type responses are given elsewhere (Burdon and Jarosz 1991; Jarosz and Burdon 1992).

**Table 2.1.** Summary of differentiation between lineage AA and AB in molecular genetic traits, physiological traits, and environmental variables for the locations from which individual isolates were collected. *P* values give the significance of differences among lineages based on ANOVA for physiological traits, and non-parametric ANOVA for environmental variables.

<table>
<thead>
<tr>
<th>Lineage</th>
<th>AA</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Molecular genetic traits</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed microsatellite heterozygosity $^$</td>
<td>0.01</td>
<td>0.88</td>
</tr>
<tr>
<td>Microsatellite loci fixed for heterozygosity $^$</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Private microsatellite alleles $^\ddagger$</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Private AFLP bands $^\ddagger$</td>
<td>3</td>
<td>106</td>
</tr>
<tr>
<td><strong>Virulence and life-history traits</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean virulence ***</td>
<td>0.36</td>
<td>0.55</td>
</tr>
<tr>
<td>Mean no. days to formation of uredia $^{NS}$</td>
<td>7.5</td>
<td>7.3</td>
</tr>
<tr>
<td>Mean no. days to formation of telia **</td>
<td>14.2</td>
<td>17.8</td>
</tr>
<tr>
<td>Mean proportion of telia formed ****</td>
<td>0.65</td>
<td>0.26</td>
</tr>
<tr>
<td><strong>Environmental variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual mean precipitation (mm)*</td>
<td>639.9</td>
<td>882.8</td>
</tr>
<tr>
<td>Annual mean temperature (°C)**</td>
<td>14.7</td>
<td>11.3</td>
</tr>
<tr>
<td>Annual mean radiation (Mj/m²/day)***</td>
<td>174.8</td>
<td>153.0</td>
</tr>
</tbody>
</table>

$^\$Based on 9 polymorphic loci. $^\ddagger$From a total of 20 alleles

To examine whether pathogen lineages differed in their overall virulence (presence/absence of infection) against the differential set of host lines, generalised linear model (Genstat 9) with a Bernoulli distribution and a logit link function was used. Pathogen isolates were nested within pathogen lineage. The main effect of host line and the interaction with pathogen lineage was included to control for the possibility that host lines might be differentially susceptible to one pathogen lineage or the other.

The inoculation results showed that all pathogen isolates had unique virulence profiles when assessed on the full set of 44 host testing lines. For lineage AA and AB isolates....
respectively, 292 of a total 836, versus 409 of 792 individual inoculations, resulted in a susceptible host response. Thus, on average, lineage AB isolates were able to overcome the resistance in 18% (22.7 v 15.4) more host lines than lineage AA isolates. The overall difference in virulence between the two lineages was highly significant (Table 2.2). While, there was also a significant interaction between pathogen lineage and host line (Table 2.2), the magnitude of this effect was much smaller than the main effect of pathogen lineage.

**Table 2.2.** Effects of pathogen lineage and host line on the virulence of isolates of the rust fungus, *M. lini*. Results are from an analysis of deviance using a generalized linear model.

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>mean deviance</th>
<th>deviance ratio</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>pathogen lineage*</td>
<td>1</td>
<td>30.26</td>
<td>40.55</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>pathogen isolate (lineage)</td>
<td>33</td>
<td>2.52</td>
<td>3.36</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>host line</td>
<td>43</td>
<td>17.10</td>
<td>22.91</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>lineage x host line</td>
<td>43</td>
<td>4.70</td>
<td>6.30</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>residual error</td>
<td>1462</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>1539</td>
<td>1.37</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note that if the effect of pathogen lineage is tested using the pathogen isolate (lineage) mean deviance as the error term, the effect is still highly significant (P<0.005).

2.2.4. **Patterns of variation in life history**

To quantify potential epidemiological and life-history differences between lineage AA and AB, host and pathogen life-history and disease epidemiology were monitored in single populations at three sites in each of two regions of NSW that correspond with the distribution of the two lineages (Table 2.3). The plains populations (where lineage AA isolates have been predominantly sampled) are situated in the western plains region of New South Wales, which is characterised by cool, moist winters and hot dry summers. The mountains populations (where lineage AB isolates have been predominantly sampled) are characterised by cool summers and cold wet winters.

At 6-weekly intervals throughout 2004/2005, host growth stage (i.e. vegetative, flowering, senescent) was recorded, and presence/absence of pathogen uredia and telia for 100 haphazardly sampled host plants. Population monitoring was not possible for the mountains region during winters, due to extensive snowfall. The end of epidemics in the plains populations were defined as occurring when all infected plants
maintained only telia, and for the mountains region as the last survey before winter conditions, by which time plants had largely died back to small above-ground shoots.

Distinct differences were found between lineages in their mode of survival between epidemics and tendency to form the resting teliospore stage. At the end of epidemics in the three plains (lineage AA) populations, all infections visible on *L. marginale* existed as telia. No uredia were observed, and all host plants in these populations were either dead or senescent. In the mountains (lineage AB) populations, the proportion of infected plants harboring telia at the end of epidemics ranged between 0 and 15.4%, with a mean of 7.4%. All infected plants maintained uredia. The proportion of plants with green shoots in the three mountains populations ranged between 29.2 and 81.3 %, with a mean of 54.6% (Table 2.3; Fig. 2.3).

However, the field data alone do not conclusively demonstrate that the observed differences in telial formation under natural conditions are under genetic control. For example, differences in the environmental conditions experienced by mountains vs. plains populations may influence the degree to which teliospores are produced. Therefore, to test the potential for genetic control of these life history differences, plants were experimentally inoculated to examine the tendency of isolates from both lineages to form telia under controlled environmental conditions. Fifteen isolates of each of lineage AA and AB were randomly selected from the pool of isolates used for molecular genotyping and virulence testing.

The experiment was initially conducted using a fully susceptible *L. marginale* inbred line (‘G’) as the inoculation target. Because infections were difficult to score for some isolates, the experiment was repeated using the more experimentally pliable and universally susceptible *L. usitatissimum* cultivar Hoshangabad. Data from this second experiment is reported here. Each of these isolates was inoculated onto the main stems of individual host plants using a standardised technique, and replicated on 6 plants for each isolate. Plants were monitored daily for the appearance of uredia and telia. The experiment was concluded after 60 days. At that time each host stem was measured to determine the total area of uredial vs. telial infection. Differences in the time to formation of telia, as well as the overall propensity to form telia (% infected area)
were analyzed using a fully factorial, mixed model analysis of variance (SAS 2000, vers. 8.2, proc mixed), with isolates nested within pathogen lineage.

**Figure 2.3.** Tendency of *M. lini* to form telia under field and glasshouse conditions. The closed bars represent the proportion of diseased host plants harboring *M. lini* telia in field surveys at the end of epidemics. The exclusive presence of AA and AB genotypes in these populations is inferred from the identity of isolates previously collected at these sites. The open bars represent the proportion of telia vs uredia formed on host stems in a controlled glasshouse experiment.

There was no significant difference between lineages in the mean number of days from inoculation to the formation of uredia (*F*₁,₂₈ = 2.79; *P* = 0.106). However, lineage AA isolates formed telia significantly faster than lineage AB isolates (*F*₁,₂₈ = 11.70; *P* < 0.01), and lineage AA isolates formed a higher proportion of teliospores than lineage AB isolates (*F*₁,₂₈ = 43.01; *P* < 0.001) (Fig. 2.3). There was a strong positive correlation in the tendency to form telia for isolates across the two inoculation experiments (*r* = 0.792; *P* < 0.0001), providing support for a genetic determination of isolate level differences in this life-history trait. Furthermore, field data showing qualitatively similar patterns of telial formation under local conditions (Fig. 2.3) suggest that adaptation of the separate lineages to differential environmental cues are unlikely to be responsible for the observed experimental differences.
Table 2.3. Host and pathogen life-history data for 6 interacting populations of *L. marginale* and *M. lini* in two separate geographic and climactic regions. Data were recorded at the end of epidemics, and used to infer the mode of pathogen survival between epidemics. N = total number of host plants sampled. Green hosts = number of plants with green shoots at time of sampling. * Telia and uredia co-occurred on these hosts. Uredia were found on green plant material only.

<table>
<thead>
<tr>
<th>Region</th>
<th>Population</th>
<th>Population size (approx)</th>
<th>Sample size</th>
<th>Green hosts</th>
<th>Host plants with uredia</th>
<th>Host plants with telia</th>
<th>% infected hosts harbouring telia</th>
<th>location</th>
<th>Nearest population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mountains</td>
<td>Kiandra</td>
<td>5000</td>
<td>137</td>
<td>40</td>
<td>33</td>
<td>5 *</td>
<td>15.2</td>
<td>S 35°48', E 148°30'</td>
<td>M3 (500 m)</td>
</tr>
<tr>
<td></td>
<td>M3</td>
<td>200</td>
<td>55</td>
<td>32</td>
<td>23</td>
<td>1 *</td>
<td>4.3</td>
<td>S 35°48', E 148°30'</td>
<td>Kiandra (500 m)</td>
</tr>
<tr>
<td></td>
<td>SH2</td>
<td>2000</td>
<td>123</td>
<td>100</td>
<td>97</td>
<td>0</td>
<td>0</td>
<td>S 35°53', E 148°31'</td>
<td>Kiandra (14 km)</td>
</tr>
<tr>
<td>Plains</td>
<td>Gundagai</td>
<td>1000</td>
<td>156</td>
<td>1</td>
<td>0</td>
<td>68</td>
<td>100</td>
<td>S 35°03', E 148°06'</td>
<td>M3 (89 km)</td>
</tr>
<tr>
<td></td>
<td>Canowindra</td>
<td>1000</td>
<td>171</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>100</td>
<td>S 33°33', E 148°06'</td>
<td>Garra (48 km)</td>
</tr>
<tr>
<td></td>
<td>Garra</td>
<td>2000</td>
<td>123</td>
<td>0</td>
<td>0</td>
<td>107</td>
<td>100</td>
<td>S 33°07', E 148°45'</td>
<td>Canowindra (48 km)</td>
</tr>
</tbody>
</table>
2.3. Discussion

Overall, our results reveal that two genetically divergent lineages of *M. lini* (lineage AA and AB) dominate interactions with the native Australian host, *L. marginale*. Importantly, the data also support a hybrid origin for lineage AB, which appears to directly influence critical life-history traits influencing pathogen distribution, virulence and disease development. The patterns of fixed heterozygosity and apparent lack of introgression between the putative A and B genomes further suggest a lack of sexual recombination in lineage AB individuals, a result that agrees with a high level of linkage disequilibrium detected among isolates from AB dominated mountains populations (Burdon and Roberts 1995). Most importantly, the geographic distribution of this variation is consistent with the generation of a geographically divergent mosaic of natural selection (Thompson 1999b; Thompson 2005) for resistance in the host, *L. marginale*.

Gene flow via hybridisation is a potential means of quickly introducing new genetic material into populations and generating rapid evolutionary change (Brasier 2000). Reports describing the ecological and evolutionary consequences of natural hybridisation events in fungal pathogens are rare. However, it is becoming increasingly clear that there may be profound biological consequences associated with pathogen hybridisation events. For example, in Europe, hybridisation between the introduced *Phytophthora cambivora*, and a second *Phytophthora sp.* has created a new pathogen via a host shift, with the hybrid able to attack a novel host species (Brasier et al. 1999). Similarly, hybrids between the poplar rust species *Melampsora occidentalis* and *M. medusae* combine the host ranges of both species, where hybrid rust pathogens are able to exclusively exploit hybrid host cultivars (Newcombe et al. 2000). Hybridisation is also facilitating rapid evolution of the Dutch elm disease pathogen, where sudden changes in the genetic structure of *O. novo-ulmi* populations can be attributed to gene transfer from *O. ulmi* of vegetative compatibility genes. This restricts the spread of deleterious viruses within the population, presumably resulting in an overall fitness advantage and increasing the chances of host infection (Brasier 2001).
To date, it has not proven possible to identify the putative BB parental lineage, despite sampling at wide geographic scales (this study), as well as more intensively in local populations (chapter 4). However, the extensive genetic differences between the A and B genomes suggest that the two parental lineages had evolved in isolation for a long period of time before the formation of hybrid lineage AB. The formation of this lineage is therefore likely to have involved the migration of one of the parents beyond its normal geographic range, and suggests that the potential for repeated hybridisation events may be geographically constrained. Given that lineage AA isolates have been recovered within the study area extensively, and that there are no other native hosts, it is suggested that the parental BB lineage typically occurs outside of the sampling area, where it infects a different host species. Long distance dispersal (via intercontinental migration), followed by inter-lineage somatic hybridisation has been previously demonstrated to be responsible for changes in pathogenicity of wheat stem rust (Puccinia graminis f.sp tritici) populations in Australia (Burdon and Silk 1997).

Despite some controversy around the long term maintenance of heterokaryosis in fungi (Pawlowska and Taylor 2004; Bever and Wang 2005), the consistent patterns of heterozygosity observed in lineage AB urediospores, and an absence of any evidence for introgression of the putative A and B genomes, are best explained by a lack of meiotic recombination and the maintenance of two divergent nuclei within each cell throughout its life-history. This conclusion is also consistent with lower levels of telial formation in lineage AB, the resting spore stage during which meiosis takes place, and a likely dead end in the absence of successful meiotic recombination.

The success of many fungal pathogens is attributed to the capacity to reproduce both sexually and asexually (Burdon 1987a). This facilitates periodic bursts of rapid population growth, with regular generation of new genotypes. In contrast, the inability of asexual lineages to generate variation through recombination is thought to be a disadvantage in temporally dynamic environments (Peters and Otto 2003). The ability to reproduce sexually may be expected to be especially important in a highly dynamic system like the Linum marginale-Melampsora lini interaction, where selection for variation in virulence is likely to be periodically intense (Burdon and Thompson 35).
1995). Thus, how an obligate asexual lineage AB maintains such a strong presence in Australia is of considerable interest.

Importantly, analysis of environmental variables for each of the sampled *M. lini* populations suggests that the two lineages have geographically distinct distributions, with lineage AB more likely to occur in southerly and coastal areas, reflecting an apparent preference for significantly cooler and wetter locations than lineage AA. Data describing the mode of pathogen survival between epidemics, and the tendency to form telia, suggest that the restricted distribution of lineage AB isolates to cool wetter areas is related to the inability of the pathogen to either initiate or survive the sexual cycle. Thus, in hot, dry areas, host senescence selects for the ability to survive in the dormant teliospore stage, during which meiosis takes place. This favours lineage AA isolates, despite their decreased virulence compared to lineage AB. In contrast, in wetter areas, where year-round availability of at least some living host tissue means that survival as teliospores is not necessary, lineage AB isolates appear to be more successful. The decreased tendency of lineage AB isolates to form telia may even provide a fitness advantage in areas where living host tissue is available throughout the year, through more constant urediospore production and less lag time in producing inoculum to initiate epidemics.

Striking differences in overall levels of virulence between the two lineages were also found, with AB isolates virulent on an average of 18% more host lines than AA isolates. However, it remains unclear to what extent increased levels of virulence in lineage AB might influence the broad scale geographic divergence of the two lineages. Isolates from both lineages were able to infect at least some host lines collected from non-corresponding geographic ranges (data not shown), indicating that the opportunity for lineage AA isolates to occur within the current range of lineage AB is not restricted by a lack of potential hosts. In addition, it appears unlikely that direct competition between the lineages strongly influences their broad-scale geographic distribution. Previous studies have demonstrated that at local scales, increased virulence can provide an advantage in highly resistant host populations. However, at larger, regional scales, highly virulent pathotypes are not universally dominant, and host populations with low levels of resistance generally harbour less...
virulent pathogens (Thrall and Burdon 2003). Together these data suggest that the geographic separation of the two lineages of *M. lini* in Australia is most likely a function of variation in life history traits and response to environmental conditions.

Increased virulence in the highly heterozygous lineage AB also raises an interesting dilemma. Theoretical studies suggest that selection should favour pathogen individuals that express the fewest potential elicitors of resistance in the host. Hence, for example, haploid parasite species should hold an advantage over diploids (Nuismer and Otto 2004). By extension, unless mechanisms exist to limit gene expression, increasing heterozygosity in polyploid species should equate to an increasing probability of being recognised by any given host. In *M. lini* (where virulence is a recessive trait), classic studies on the genetics of pathogenicity (Flor 1955), and recent studies exploring the molecular basis of avirulence (Dodds et al. 2004; Dodds et al. 2006) indicate that all allelic copies of genes determining virulence are expressed. Furthermore, DNA sequence data from the *AvrP123* and *AvrP4* loci, flax rust avirulence genes (Catanzariti et al. 2006), revealed the fixed presence of two allelic variants in each of the lineage AB isolates. In contrast, lineage AA isolates were on average heterozygous in only 19% of cases (chapter 3). This raises an important question: why do lineage AB isolates have a lower probability of being recognised by any given host?

One possible explanation is that lineage AB may be functionally haploid with regard to interactions with *L. marginale*. If, as suggested previously, the parental lineage BB naturally occurs on a different host species, genetic divergence of avirulence elicitors in the B genome may mean they are not recognised by *L. marginale*. This would decrease the mean likelihood of detection of lineage AB genotypes by *L. marginale*, as potential avirulence elicitors would only be present in the ‘A’ nucleus. This hypothesis is supported by patterns of variation in *AvrP123* and *AvrP4* alleles associated with the ‘A’ and ‘B’ genomes in isolates infecting *L. marginale*. For *AvrP123* and *AvrP4*, 5 and 10 different protein coding variants were recovered from the ‘A’ genome respectively. In contrast, for *AvrP123* the highly divergent ‘B’ type alleles were invariant, and for *AvrP4* only two alleles were recovered (with the variant a singleton) (chapter 3). Such a pattern may be expected if diversifying or balancing
selection maintains genetic variation within ‘A’ type avirulence elicitors, and where there is little selective advantage to maintaining variation in potential ‘B’ type avirulence elicitors that are not recognised by *L. marginale*.

The geographic and evolutionary separation of these pathogen lineages may also have important consequences for the evolution of resistance in the host, *L. marginale*. Our observations of contrasting life-histories and levels of virulence in *M. lini* are particularly intriguing when examined in the context of a previous study (Burdon et al. 1999) that examined patterns of mating and resistance in two metapopulations of *L. marginale*. In the mountains metapopulation, which is dominated almost exclusively by lineage AB isolates (chapter 4), the host was found to be almost completely selfing. In contrast, in the hotter, drier plains metapopulation, where lineage AA isolates have been predominantly sampled (chapter 4), the host showed significant levels of outcrossing. Differences between metapopulations in host outcrossing rates were also reflected in significant variation in overall levels and phenotypic diversity of resistance, with the plains region showing consistently higher values. This situation, where clonal pathogens interact with inbreeding hosts and sexual pathogens associate with outcrossing hosts, is consistent with the Red Queen Hypothesis (Hamilton 1980). The RQH predicts that sexual reproduction provides an evolutionary advantage in host-pathogen interactions through frequency dependent selection causing different gene combinations to be favoured in different generations.

Overall, the results presented here demonstrate not only the importance of investigating coevolutionary interactions across a broad range of spatial scales, but also the value of integrating studies using molecular markers with experimental epidemiology. A fundamental pattern of genetic divergence in *M. lini* at both regional and continental scales has been uncovered, revealing the potential for accelerated pathogen evolution through gene flow and hybridization. Moreover, geographic and regional patterns of environmental heterogeneity appear to play an important role in maintaining these broad differences in life-history and virulence among major pathogen lineages, and may have further impact on variation in host mating system and phenology. Together, these results highlight the potential for geographic selection
mosaics to generate and maintain coevolutionary diversification in long-standing host-pathogen systems.
2.4 Literature Cited


CHAPTER 3

Adaptive evolution of avirulence loci in wild populations of the flax rust Melampsora lini
3.1 Introduction

Genetic variation for the ability to overcome host resistance is widespread in populations of microbial pathogens. Understanding how genetic variation in pathogen virulence and aggressiveness is generated and maintained is of central importance to understanding the dynamics and distribution of disease. Correlations between individual host and/or pathogen genotypes and disease outcomes have been demonstrated for a number of host-pathogen interactions (Buitkamp et al. 1996; Paterson et al. 1998; Lockett et al. 2001; Williams-Blangero et al. 2002; Dodds et al. 2006; Ong and Innes 2006). Furthermore, genetic variation for host resistance and pathogen virulence are likely to be critical factors influencing disease epidemiology (Thrall and Burdon 2000) and the emergence and spread of new diseases (Prentice et al. 2001; Friesen et al. 2006; Piertney and Oliver 2006). However, there is limited knowledge of how population level ecological and evolutionary processes influence the generation and maintenance of pathogenic variation. Increasingly, this lack of knowledge has led to calls for an integrated approach to the study of pathogenic organisms, incorporating ecology, evolution and molecular biology (Tibayrenc 1998; Real et al. 2005; Mitchell-Olds et al. 2007).

The gene-for-gene paradigm provides a powerful framework for understanding coevolutionary dynamics between hosts and their pathogens. In antagonistic gene-for-gene interactions, first described by Flor (1955), resistance to disease is governed by the recognition of pathogen avirulence (Avr) genes (avirulence genes are factors which the plant host has evolved to recognise as an indicator of invasion) by corresponding host resistance (R) genes. The gene-for-gene concept has shaped much of our current thinking about host-pathogen coevolution in both animal and plant systems. In particular the presence of corresponding R and Avr genes in host and pathogen populations implies the possibility of coevolution driven by selection pressure on the pathogen to escape recognition by host R gene products, and concomitant pressure on the host to respond to new virulent strains of the pathogen. Such antagonistic evolutionary pressure should leave detectable signals at the molecular level in loci that determine host resistance and pathogen virulence.
Natural selection in host-pathogen interactions is typically envisaged to be either directional or balancing. One example of when directional selection may occur is in situations that favour hosts that are resistant to all pathogen genotypes and pathogens that can infect all host genotypes. Under such an extreme dynamic, continual selective sweeps and selective turnover of alleles may be expected to result in \( R \) and \( Avr \) loci that are relatively young and lacking in diversity (Bergelson et al. 2001; Ford 2002; Tiffin and Moeller 2006). Evidence for the coevolution of both pathogen virulence and host resistance via strong, species level directional selection is limited. However, there are some examples where either \( R \) or \( Avr \) loci display a molecular signature consistent with this scenario (Bergelson et al. 2001; de Groot et al. 2002).

Significant intraspecific polymorphisms for disease resistance and pathogen virulence (Flor 1956; Harry and Clarke 1987; Jarosz and Burdon 1990; Burdon 1994) are generally more consistent with a model where polymorphisms are maintained, rather than eliminated, by selection. Many host species, particularly in natural ecosystems, exist as interacting groups of small, geographically and genetically differentiated populations, subdividing pathogen populations into relatively small, discrete units (Burdon 1992). Within such a framework, spatially heterogeneous patterns of directional selection may occur in different populations, causing genetic divergence and local adaptation (Gandon et al. 1996; Gandon 1998; Hedrick 2006). Alternatively, coevolutionary dynamics can maintain polymorphisms within populations via balancing selection (i.e. frequency dependent or overdominant selection). Negative frequency dependent selection, where rare pathogen (or host) genotypes have higher fitness, is a common assumption in models of host-parasite interactions (Gillespie 1975; Clarke 1976; Anderson and May 1982) and helps explain the population level maintenance of resistance and virulence polymorphisms (Jayakar 1970; Hedrick 1974; Antonovics and Thrall 1994; Stahl et al. 1999). Furthermore, spatially asynchronous frequency dependent dynamics can maintain polymorphisms across wider metapopulation scales (Thrall and Burdon 2002), while directional and balancing selection can also act simultaneously, with positive selection generating allelic variants on which balancing selection acts (Thompson 2005).
The nature of the interaction between host R and pathogen Avr genes is likely to strongly influence the intensity and direction of evolutionary selection, and thus levels and patterns of diversity, at host R and pathogen Avr loci. Currently, two alternative pathways have been proposed for the recognition event between the host R and pathogen Avr proteins. In one case (the guard hypothesis: van der Biezen and Jones, 1998), R proteins confer an indirect mode of resistance by detecting changes in host proteins that are modified by the effector function of corresponding Avr proteins. For instance, the Arabidopsis thaliana Rpm1 protein detects changes induced in the host protein RIN4 by the corresponding bacterial Avr proteins (Mackey et al. 2002; Axtell and Staskawicz 2003; Mackey et al. 2003). In this case, simple balanced polymorphisms for resistance and virulence have been maintained over long evolutionary time scales in the host and pathogen (Bergelson et al. 2001). Other R loci involved in indirect recognition events are also characterised by low allelic diversity, and functional alleles are of relatively ancient origin and subject to purifying selection (Stahl et al. 1999). Large fitness penalties associated with carrying resistance alleles in the absence of disease pressure may explain why these alleles are not fixed in the host population (Tian et al. 2003).

The second R-Avr interaction pathway involves direct protein-protein interaction between corresponding R and Avr proteins. In this case, Avr genes may overcome resistance via sequence diversification rather than loss of function, so that the Avr protein retains its effector function, but escapes recognition. Recent molecular studies indicate that this mode of interaction is likely to result in high levels of polymorphism and nucleotide diversity at R and Avr loci. For example, in the interaction between flax (Linum usitatissimum) and flax rust (Melampsora lini), direct protein interactions underlie specificity between alleles at the L resistance locus and corresponding Avr alleles in flax rust (Dodds et al. 2006). The interacting loci are characterised by high levels of functional polymorphism and strong diversifying selection.

Given the diversity of pathogen effector proteins, the generality of the role of demonstrated Avr genes and their homologs as recognition targets in different host taxa is not well understood. However, the potential for the long term maintenance of an interaction between specific host R and pathogen Avr genes has been demonstrated.
in the interaction between the pathogen *Cladosporium fulvum* and several different host species in the genus *Lycopersicon*. Differential recognition of the *C. fulvum* Avr elicitors Avr4 and Avr9 occurs across multiple *Lycopersicon* spp. Several functional homologs of the cloned R genes Cf-4 and Cf-9 originally identified as interacting with Avr4 and Avr9 elicitors were also identified in divergent host species, suggesting that the interaction between *C. fulvum* and host species in the genus *Lycopersicon* pre-dates *Lycopersicon* speciation (Kruijt et al. 2005). In addition, the AvrB protein from the plant pathogen *P. syringae* pv. *glycinea* can act as an avirulence factor in both *Arabidopsis* and *Glycine max*, interacting with the R genes Rpm1 and Rpg1b respectively (Ong and Innes 2006). However, while the two R genes share a recognition specificity, phylogenetic analyses suggest they have evolved independently (Ashfield et al. 2004). In the genus *Melampsora*, homologs of the AvrP4 gene (originally identified in *M. lini*) are conserved across several species. Although positive selection appears to have played an important role in their divergence (Marlien van der Merwe unpublished manuscript), whether these genes are targets for resistance in their various host species remains unknown.

The interaction between the rust pathogen *M. lini* and two host species, *Linum marginale* and *L. usitatissimum* serve as model systems for understanding coevolutionary dynamics in natural systems, and for understanding the genetic basis of host resistance and pathogen virulence respectively. The interaction between *M. lini* and *L. usitatissimum* has been a model system for understanding the genetic basis of pathogenicity since 1942 (Flor 1942). In *L. usitatissimum*, resistance to *M. lini* is controlled by genes segregating at five loci. In contrast, *M. lini* exhibits complex polymorphisms for virulence controlled by avirulence genes that are dispersed throughout the genome (Lawrence et al. 2007). To date, *M. lini* avirulence genes have been cloned from four loci. Two of these loci (*AvrL567* and *AvrM*) maintain multiple linked gene copies, whereas the other two (*AvrP123* and *AvrP4*) are single copy loci. While the exact function of these genes is not known, all encode novel, small secreted proteins that are generally thought to function as pathogenicity effectors that play a role in establishing infection (Catanzariti et al. 2007). Multiple virulence polymorphisms are maintained via amino acid differences in the expressed *Avr* proteins (Dodds et al. 2006). Typically, these *Avr* loci have high allelic diversity, and
show a significant excess of non-synonymous to synonymous changes, indicating that diversifying selection is likely acting on these genes (Dodds et al. 2004; Catanzariti et al. 2006).

In Australia, *M. lini* infects *L. marginale*, an endemic wild, herbaceous plant species. *M. lini* occurs throughout the entire range of *L. marginale*, and Australian populations are specialised in terms of their virulence on the Australian host to an extent that suggests a sustained period of coevolution. Extensive phenotypic variation for host resistance and pathogen virulence both within and among populations has been demonstrated over multiple studies (Lawrence 1989; Lawrence and Burdon 1989; Thrall et al. 2002; Thrall and Burdon 2003). Although correlated genetic studies in both host and pathogen have not been carried out for this interaction, Burdon (1994) demonstrated that *L. marginale* possesses single, major genes for resistance, and that these genes are largely allelic or closely linked to each other. Recently, isolates of *M. lini* infecting *L. marginale* in Australia have been shown to fall into as two distinct genetic lineages, (AA and AB; chapter 2), with distinct genetic, life-history and pathogenic differences. Lineage AB isolates, which are thought to have a hybrid origin, are fixed for heterozygosity at multiple microsatellite loci. Lineage AA isolates in contrast have low heterozygosity, and consistently have an allele in common with lineage AB isolates. This suggests that lineage AB is a fixed F1 hybrid, and that sexual recombination is not occurring in this lineage. Furthermore, the two lineages have disjunct geographic distributions, and lineage AB isolates are on average nearly 20% more virulent than lineage AA isolates (chapter 2).

Theoretical studies suggest that selection should favour pathogen individuals that express the fewest potential elicitors of resistance in the host (Nuismer and Otto 2004). Hence, for example, haploid parasite species should hold an advantage over diploids. By extension, unless mechanisms exist to limit gene expression, increasing heterozygosity in polyploid species should equate to an increasing probability of being recognised by any given host. Given the high levels of microsatellite heterozygosity shown in lineage AB, this raises an important question: why are lineage AB isolates virulent on a wider range of host lines? This is particularly the case, given that an a priori expectation is that lineage AB isolates are likely to be
heterozygous at Avr loci, and thus should carry more rather than less potential elicitors of avirulence. Examination of patterns of polymorphism in the respective A and B genomes may give some insight into the increased virulence in lineage AB.

Here, the role of the \textit{AvrP4} and \textit{AvrP123} genes isolated in the \textit{L. usitatissimum-M. lini} interaction, are examined in determining avirulence of \textit{M. lini on L. marginale}. For isolates collected from both \textit{L. marginale} and \textit{L. usitatissimum}, polymorphisms at both the \textit{AvrP4} and \textit{AvrP123} loci are characterised and compared to levels of sequence diversity at the internal transcribed spacer region of the ribosomal DNA (ITS) and introns of the β-tubulin 1 gene. Of particular interest was the extent of haplotype and nucleotide polymorphism in the Australian isolates, and in determining if these genes were evolving under positive selection. In addition, patterns of polymorphism were closely examined within and among: (a) Australian lineage AB isolates, with an interest in explaining high levels of virulence in this lineage; and (b) isolates collected from the different host species.

3.2. Methods

3.2.1. Sampling

To investigate both local and broad scale patterns of polymorphism at \textit{AvrP123} and \textit{AvrP4}, isolates of \textit{M. lini} were sampled at two spatial scales (Fig. 3.1). To examine broad geographic and species level patterns of polymorphism, twenty-nine \textit{M. lini} isolates collected from \textit{L. marginale}, and four from \textit{L. usitatissimum} were selected. The \textit{L. marginale} isolates represented samples collected from across the range of the species in Australia. Twenty-six isolates had been previously shown to vary in virulence phenotype (Burdon \textit{et al.} (2002)), and included isolates from Western Australia, South Australia, Tasmania, New South Wales and Victoria. In addition, three further isolates collected from New South Wales (Marrar 9, 29) and South Australia (KP-15) were included. The additional isolates from New South Wales represent new variants identified in the population level survey. Overall, 15 \textit{L. marginale}-associated isolates belonged to lineage AA and 14 isolates belonged to lineage AB (chapter 2). Isolates from \textit{L. usitatissimum} (cultivated flax) were obtained from the US, South America and New Zealand. For the sake of simplicity, these
isolates are referred to as ‘flax isolates’, and the Australian isolates as lineage AA and lineage AB.

To examine patterns of polymorphism at local and regional spatial scales, 275 isolates from 10 Australian populations associated with *L. marginale* were sampled (Table 3.6). The pathogen populations occur in two geographically and environmentally distinct areas that closely correspond to the distributions of lineage AA and lineage AB (termed the mountains and plains regions) (chapter 4). The lineage identity of each of these isolates had been previously confirmed using AFLP markers (chapter 4). For the five plains populations, 123 isolates belonging to lineage AA were characterised. A small number of lineage AB isolates (18) recovered across the five populations were excluded from all analyses. For the mountains populations, 133 isolates belonging to lineage AB were characterised. A single lineage AA isolate recovered in the population N2 was excluded from all analyses.

**Figure 3.1.** Map with insert region showing locations of sampling sites and populations referred to in this study. Sampling locations for lineage AA and AB isolates are represented by open circles and closed triangles respectively. The insert region shows locations of local
populations sampled in New South Wales, Australia. Isolates sampled from the 5 populations in the plains region are exclusively lineage AA, and isolates from the 5 populations in the mountains region exclusively lineage AB.

3.2.2. DNA sequence determination

Total genomic DNA was extracted from 100 mg of urediospores using a DNeasy plant mini kit (Qiagen). A sequence dataset comprising 4 nuclear genes, including two Avr loci, an 810 bp fragment of the β-tubulin locus and the ITS region was generated. All sequence electropherograms were closely examined for double peaks, and cloned as required, using Promega pGEM-T vector system 1 cloning kits and following standard protocols. At least two identical copies of all cloned products were sequenced so as to avoid errors due to misincorporation by Taq polymerase. Sequence alignments were initially performed on the amino acid sequence, using conserved cystine motifs as reference points and then back-translated, using the software BioEdit (Hall 1999). Forward and reverse sequences were obtained for each locus in the species level sample. Only forward sequences were generated for the population level samples, however, any novel variants identified were confirmed via reverse sequences, and included in the species level analyses. Throughout the following chapter, for both AvrP123 and AvrP4, haplotypes recovered from isolates collected from L. marginale are prefixed with ‘Lm’, and alleles recovered from isolates collected from L. usitatissimum are prefixed with ‘Lu’.

AvrP4 encodes a 95 amino acid protein with a predicted 28 amino acid cleavable secretion signal peptide. PCR amplifications were performed on a Hybaid Express thermocycler under the following conditions: 95 °C for 3 min, 34 cycles at 94 °C for 30 s, 56 °C for 45 s, 72 °C for 90 s, followed by a 4 °C holding step. Two sets of primers were used for PCR amplification. For the continental and species level survey, primers were designed to amplify 180 bp of 5’ flanking sequence, 285 bp of open reading frame, and 103 bp of 3’ flanking sequence (AvrP4_F1 5’-CATCAAAATCTAACCCTGAT-3’ and AvrP4_R1 5’-GTAGCATTGAGATCCATGG-3’). All sequences were closely examined for double peaks, and cloned as required using Promega pGEM-T vector system 1 cloning kits and following standard protocols. This survey revealed fixed heterozygosity in lineage AB isolates, and a lack of any polymorphism in the ‘B’ type allele. Thus, to avoid
cloning all lineage AB isolates, for the population level survey an alternative reverse primer designed to amplify ‘A’ type alleles only (AvrP4_R2 5’-TTGTTACGGATAGATAGTGC-3’) was used. Any novel variants uncovered were re-amplified using the original primer set, sequences, and cloned as required.

AvrP123 encodes a 117 amino acid protein with a predicted 23 amino acid cleavable secretion signal peptide. Primers were designed to amplify 119 bp of 5’ flanking sequence, 351 bp of open reading frame, and 128 bp of 3’ flanking sequence (AvrP123_F 5’-ATTGTGAACCTTTTGAAGGAC-3’ and AvrP123_R 5’CGCCATGGTATTGTTCAGAC-3’). As for AvrP4, alternative protocols were used for the species level and population level surveys. For the species level surveys, PCR amplifications were performed as for AvrP4 except for a 54 °C annealing temperature. To eliminate the ‘B’ type allele (which had some mutations in the original priming sequence) for the population levels survey, the PCR annealing temperature was increased to 58 °C.

For β-tubulin and ITS genes, sequence polymorphism was characterised in the species level isolates only. Primers were designed for the amplification of an 810 bp fragment of the β-tubulin gene in *M. lini* (Ayliffe et al. 2001), incorporating 5’ flanking sequence and 5 introns and 6 exons (BT-793F 5’-AAAAACCCAAAAACCTAAATCAA-3’ and BT-793R 5’-GACCTTTGGCCCAGTTGTA-3’). The internal transcribed spacer (ITS) region was amplified using the primer pair ITS1 and ITS4 (White et al. 1990).

3.2.3. Transient in Planta Expression Assays
To assess if *Avr* haplotype variants could act as *Avr* factors in *L. marginale* Agrobacterium infiltration was used to transiently express these genes in leaves of a standard set of 12 *L. marginale* differential host lines (AA, CC, HH, II, RR, UU, B, C, G, T, U, V). Line ‘G’ has no known resistance genes and was included as a negative control. These lines have been used extensively in previous work assessing pathogen population structure (Burdon and Jarosz 1991; Jarosz and Burdon 1991; Burdon and Jarosz 1992; Thrall et al. 2002). Following sequencing, 6 unique alleles recovered from the *AvrP123* locus [Lm-1, Lm-3, Lm-4 (identical to Lu-5), Lm-5, Lm-7, Lu-1],
and 8 recovered from \textit{AvrP4} (Lm-1, Lm-5, Lm-7, Lm-8, Lm-10, Lm-12, Lu-1, Lu-2) were selected. Alleles were selected so as to ensure that the sequence diversity among the different isolates was well represented (Figs. 3.3, 3.4).

Transformation methods followed those implemented by Catanzariti et al (2006). Gene expression constructs contained the \textit{AvrP4} and \textit{AvrP123} coding sequences inserted between a 35S cauliflower mosaic virus promoter and a nopaline synthase terminator in the binary vector pTNotTReg (Anderson et al. 1997). The \textit{AvrP123} expression constructs were prepared by inserting a cloned PCR product as an \textit{EcoRI} fragment. This contains the entire coding sequences as well as 87 nucleotides 5' to the first ATG and 194 nucleotides of the 3' untranslated region. The \textit{AvrP4} expression constructs were constructed by inserting a cloned PCR product as a \textit{BamHI-SacI} fragment. Constructs contained the entire coding sequences as well as 22 nucleotides 5' to the first ATG. \textit{Agrobacterium tumefaciens} cultures of strain GV3101 pMP90 containing the binary vector expression constructs were prepared at an OD$_{600}$ of 1.0 in LB medium containing 200 $\mu$M acetosyringone and infiltrated into \textit{L. marginale} leaves by syringe. \textit{A. tumefaciens} containing an empty vector were used across all plants as a negative control. Phenotypes were scored after 14 days. Where transient expression of any of the \textit{AvrP123} or \textit{AvrP4} variants induced an HR-type necrotic response in one of the differential host lines, this was interpreted as a specific recognition of the Avr allele by a corresponding R gene in the host. All host/allele combinations inducing a positive response were retested to confirm reproducibility.

3.2.4. Data Analyses

3.2.4.1. Gene diversity and phylogenetic analyses
The DNAsp v4.10.9 software package (Rozas and Rozas 1999) was used to calculate nucleotide diversity statistics and perform McDonald-Kreitman tests for selection (McDonald and Kreitman 1991). The significance of differences between average pairwise nucleotide differences among coding and flanking regions, and \textit{Avr}, ITS and $\beta$-tubulin genes was assessed by $t$ test. Neighbor Joining trees were generated in MEGA ver3 (Kumar et al. 2004) using the Kimura 2-parameter evolutionary model.
and 1000 bootstrap replicates. For phylogenetic analyses, ITS and β-tubulin gene sequences were concatenated.

3.2.4.2. Tests for selection

Several different approaches were used to test for selection. For tests based on comparisons of nonsynonymous substitution \((dN)\) rates with the rate of synonymous substitution \((dS)\) rates, only the amino acid variants were considered. Positive selection was initially tested for by comparing the rate of nonsynonymous substitutions \((dN)\) with the rate of synonymous substitutions \((dS)\). A \(dN/dS\) ratio significantly greater than 1 indicates positive selection. If purifying selection removes non-synonymous substitutions the \(dN/dS\) ratio will be less than one. If non-synonymous mutations are neutral then the \(dN/dS\) ratio will equal one. The program CODEML in the software package PAML (Phylogenetic Analysis by Maximum Likelihood) (Yang 1997) was used to estimate the non-synonymous/synonymous rate ratio parameter \(o\) and so identify codons and lineages under positive selection. Rates of nonsynonymous versus synonymous mutations at Avr loci were calculated via maximum likelihood analyses using the M0 model of the CODEML program. Tests for positive selection were performed by identifying codons that have been repeatedly subjected to natural selection using models (models M1, M7) that do not permit positive selection, compared to models (models M2, M8) that permit sites to evolve under positive selection. The presence of sites under positive selection was tested for significance by comparing twice the log-likelihood difference (M2 vs M1, M8 vs M7) in a chi square test with two degrees of freedom. Codons that were identified as having evolved under positive selection with high posterior probabilities \((P > 0.95)\) were highlighted on an amino acid alignment of the two genes.

The McDonald-Kreitman test (McDonald and Kreitman 1991) was used to contrast patterns of within-species polymorphism and between species divergence at synonymous and non-synonymous sites in the coding region of a gene. This approach assumes that under neutrality, for both synonymous and nonsynonymous mutations, the number of substitutions between two species and the number of polymorphic changes within a species will both be proportional to the mutation rate. If the ratio of synonymous to nonsynonymous changes differs significantly for within versus
between species comparisons, then the assumption of neutrality can be rejected (Otto 2000). Based on the high levels of divergence between the A and B genome, this test was focused on polymorphism within the cluster of ‘A’ alleles for each locus, using the divergent B alleles as a between species comparison. Statistical departure from neutrality was tested with a G-test on a 2x2 contingency table of synonymous and non-synonymous mutations within and between species. The associated neutrality index \((NI)\) (Rand and Kann 1996) which reflects the extent to which the levels of amino acid variation within species depart from the neutral model, is also reported. Under neutrality, an index of 1 is expected. Values greater than 1 indicate an excess of non-synonymous substitutions within species, and values less than 1 indicate an excess of non-synonymous substitutions among species, relative to the number of synonymous substitutions.

3.2.4.3. Recombination

Aligned sequence variants were tested for recombination using the software package RDP (Recombination Detection Program) v2 (Martin et al. 2005). This software package implements a range of statistical algorithms for detecting recombination, and uses a consensus approach to overcome concerns about the limitations of individual methods. The six methods applied here are RDP (Martin and Rybicki 2000), GENECONV (Padidam et al. 1999), Bootscan (Salminen et al. 1995), MaxChi (Smith 1992), Chimaera, and SiScan (Posada and Crandall 2001). The null hypothesis is that sequence variation is generated by nucleotide mutations. Recombination was deemed to occur at a locus if the null hypothesis was rejected by three or more tests at a significance level of 0.01.

3.2.3.4. Population genetic analyses

For both \(AvrP123\) and \(AvrP4\), allele frequencies were calculated, and estimates obtained of the observed heterozygosity \((H_o)\) for each population using the software GENALEX v6.0 (Peakall and Smouse 2005). Haplotype (gene) diversity and its standard deviation (Nei 1987) was calculated for each population using the software DNAsp v 4.10.9 (Rozas and Rozas 1999). To explore any relationships between genetic divergence and geographical distance among populations, matrices of pairwise genetic distance \((\Phi_{PT})\) and geographic distance were subject to a Mantel test.
(Mantel 1967), for both AvrP123 and AvrP4, using GENALEX v6.0 (Peakall and Smouse 2005) with 10000 permutations.

3.3. Results

3.3.1. Phylogenetics and molecular diversity
Sequences from the ITS and β-tubulin loci were recovered for a set of 33 isolates. These were used to provide comparative phylogenies and levels of nucleotide diversity at neutrally evolving loci with Avr loci. Phylogenetic analyses based on concatenated ITS and β-tubulin sequences revealed three distinct clades. Clades 1 and 2 consist of haplotypes recovered from L. marginale (both lineage AA and AB) and L. usitatissimum respectively. Clade 3 consists of haplotypes recovered from lineage AB isolates only (Fig. 3.2). Lineage AB isolates were consistently heterozygous for alleles belonging to clades 1 and 3. Among the isolates collected from L. usitatissimum, five concatenated haplotypes were identified based on the ITS and β-tubulin sequences. The isolates collected from the Americas formed a distinct cluster (bootstrap support of 75%) sister to the single isolate collected from New Zealand (Clade 2). All isolates collected from L. marginale contained a single clade 1 haplotype (Fig. 3.2).

Primers designed to amplify conserved flanking regions of the AvrP4 and AvrP123 genes were next used to amplify these genes from the set of rust isolates. Homologs of these genes were identified in all isolates, but showed considerable sequence variation. Comparative analyses among loci showed that haplotype and gene (π) diversities were significantly higher for AvrP123 and AvrP4 than for either ITS or β-tubulin (Table 3.1). Genotype diversity at Avr loci was similar between lineages, with 11 multilocus genotypes from 15 lineage AA isolates, and 9 multilocus genotypes from 14 lineage AB isolates. All four L. usitatissimum isolates presented unique multilocus avr genotypes. Observed multilocus Avr gene heterozygosity was 0.23 for lineage AA, 0.25 for L. usitatissimum isolates and 1 for lineage AB. Both the AvrP123 and AvrP4 gene regions were tested for evidence of recombination using the six test algorithms included in the software RDP. No evidence of recombination was detected at either locus.

Chapter 3: Evolution of virulence
Neighbor joining trees based on AvrP123 and AvrP4 coding sequences were inferred from the coding portion of the nucleotide sequences of the AvrP123 and AvrP4 genes (Fig. 3.3 A, B). For the samples collected from L. marginale, the 14 individuals belonging to lineage AB were fixed for heterozygosity at both Avr loci, with one allele that clustered with alleles isolated from both lineage AA and flax isolates (hereafter referred to as the ‘A’ clade), and a second highly divergent allele (the ‘B’ clade). The haplotypes recovered from the Australian ‘A’ genome and the cultivated isolates cluster together at both loci into a clearly delineated clade A. There were no clear patterns of divergence between Avr alleles from flax and Australian isolates, although alleles were not shared among Australian and flax isolates (with the exception of the AvrP123 alleles Lm-4 and Lu-5 which were identical at the amino acid level) (Fig. 3.4).

A total of 12 DNA and 10 amino acid alleles were recovered for locus AvrP123. One allele in a rust strain collected from L. usitatissimum was truncated but occurred in a heterozygous state along with a full length variant. The truncated allele (Lu-5) has a large (186 bp) internal frame deletion and was assumed to be non-functional (Fig. 3.3). This allele was therefore excluded from the phylogenetic and statistical analyses. Considering amino acid variants for isolates collected from L. marginale only, 9 ‘A’ type alleles and one highly divergent ‘B’ allele (two DNA variants) were recovered (Figs. 3.3A, 3.4A, Table 3.2). The ‘B’ allele was present in all 14 lineage AB isolates. Five alleles identified in clade A (Figure 3.3A) were recovered from L. marginale isolates. Three of these alleles were shared among lineages (Lm-1, Lm-2, Lm-3), with a single clade A allele found to be specific to both lineage AA (Lm-4) and AB (Lm-5). The L. marginale isolates were dominated by three closely related haplotypes (Lm-1, Lm-2, Lm-3).

Four alleles from clade A were recovered from samples collected from L. usitatissimum. One amino acid variant recovered from a sample collected from L. marginale (Lm-4) was identical to an allele recovered from a sample collected from L. usitatissimum (Lu-5). Most of the polymorphic sites were scattered across the
mature protein, with some fixed polymorphisms between clade A and B haplotypes in the signal peptide (Fig. 3.3A).

Figure 3.2. An unrooted NJ tree showing relationships among concatenated ITS and β-tubulin haplotypes. Major clades discussed in text are indicated by numbers placed above appropriate lineages. Group 1 indicates haplotypes recovered from lineage AA and lineage AB isolates collected from *L. marginale*. Group 2 indicates haplotypes recovered from isolates collected from *L. usitatissimum*. Group 3 indicates haplotypes recovered only from lineage AB isolates collected from *L. marginale*. Bootstrap proportions of 1000 bootstrap replicates >75 are indicated on the branches.
Figure 3.3. Amino acid alignments for A: AvrP123 and B: AvrP4, with sites under significant ($P < 0.05$) positive selection (M2) (marked ! underneath the alignments). Alleles recovered from L. marginale isolates are in black font and prefixed with Lm. Alleles recovered from L. usitatissimum isolates are in blue font and prefixed with Lu. Only those amino acids that differ from the top sequence are shown, with identical residues indicated by dots. The conserved Cysteine amino acid residues are highlighted in grey.
For the *AvrP4* locus a total of 16 DNA variants that translated into 16 amino acid variants were recovered. For isolates collected from *L. marginale* in Australia, 11 clade ‘A’ alleles and two highly divergent clade ‘B’ alleles (Lm-12 and Lm-13) from lineage AB individuals were recovered (Figs. 3.3B, 3.4B, Table 3.3). Three of the clade A alleles (Lm-1, Lm-7, Lm-8), were shared among lineages AA and AB, while the remaining eight were specific to either lineage AA (Lm-2, Lm-3, Lm-4, Lm-5) or AB (Lm-6, Lm-9, Lm-10, Lm-11). Of the ‘B’ alleles, one (Lm-13), was recovered from only one isolate which was collected from Tasmania. The second B allele (Lm-12) was present in all the remaining 13 hybrid isolates. Three clade A alleles were recovered from samples collected from *L. usitatissimum*. No common alleles were found among isolates collected from the two different hosts. Polymorphic sites were scattered across the entire protein, with high levels of variation in the 3’ terminal end of the protein (Fig. 3.3B).

**Table 3.1.** Gene diversity estimates for nucleotide sequence data collected from avirulence genes and neutrally evolving loci in 33 isolates of the flax rust fungus *Melampsora lini.*

<table>
<thead>
<tr>
<th>Locus</th>
<th>AvrP123*</th>
<th>AvrP4*</th>
<th>ITS</th>
<th>β-tubulin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length (bp)</td>
<td>351</td>
<td>289</td>
<td>790</td>
<td>810</td>
</tr>
<tr>
<td>No. of <em>L.m.</em> alleles</td>
<td>7</td>
<td>13</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>No. of <em>L.u.</em> alleles</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Segregating sites</td>
<td>118</td>
<td>80</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>Polymorphism (π)</td>
<td>0.124</td>
<td>0.078</td>
<td>0.005</td>
<td>0.004</td>
</tr>
</tbody>
</table>

*Avr statistics based on coding sequence only

1 Isolates collected from *Linum marginale* (n = 33)

2 Isolates collected from *Linum usitatissimum* (n = 4)

3.3.2. Selection

Examinations of patterns of polymorphism between loci are consistent with the hypothesis that diversifying selection is generating variation at *AvrP123* and *AvrP4*. Nucleotide diversity at both of the *Avr* loci is more than an order of magnitude higher than in either ITS or β-tubulin (Table 3.1) (χ² test for significance: all comparisons, *P* < 0.001), suggesting accelerated rates of evolution at both *Avr* loci, and at *AvrP123* in
particular. Also, a significant excess of nucleotide substitutions was found in the \textit{Avr}
coding regions (\textit{AvrP123}: 140 in 353 bp; \textit{AvrP4}: 96 in 289 bp) relative to the flanking
DNA (\textit{AvrP123}: 37 in 301 bp; \textit{AvrP4}: 31 in 292 bp) when all different haplotypes
were considered ($\chi^2$ test: both loci, $P < 0.001$). Similarly, when considering variation
within loci, nucleotide diversity within the coding sequence for both \textit{Avr} genes is
significantly higher than in the associated flanking sequence (Tables 3.2, 3.3).

Of note are the contrasting patterns of polymorphism between the different lineages,
and between haplotypes in clades A and B (Tables 3.2, 3.3). All of the variation
within lineage AB is found within alleles sourced from the A genome, with the
exception of one highly divergent clade B allele (Lm-13) at \textit{AvrP4}. This lack of
diversity is in strong contrast to the pattern seen among alleles from clade A, where
there are 8 and 14 variant haplotypes at \textit{AvrP123} and \textit{AvrP4} respectively, and
suggests that there is little or no selection for variation acting on clade B alleles.
However, the high levels of divergence between the different \textit{Avr} gene allelic classes
in lineage AB, particularly when compared to ITS and $\beta$-tubulin, suggests instead that
strong directional selection may be acting to promote divergence between clades A
and B. The presence of a single, divergent third allele at \textit{AvrP4} (Lm-13) further
suggests that strong directional selection may be acting on this gene.

Further evidence that clade B alleles are not under strong diversifying selection can be
seen when examining patterns of nucleotide polymorphism among the different
lineages and clades (Tables 3.2, 3.3). Overall, the ratio of nucleotide polymorphism at
synonymous and non-synonymous sites is close to 1 in lineage AB. In contrast, the
ratio of non-synonymous to synonymous polymorphisms in lineage AA and cultivated
isolates suggests that a non-synonymous mutation is between two and five times more
likely to be fixed in these isolates.
Figure 3.4. Genealogical relationships of *Melampsora lini* isolates as determined by the coding sequences of *Avr* loci using a Neighbor Joining Analysis. A: *AvrP123* and B: *AvrP4*. Bootstrap proportions of 1000 bootstrap replicates >75 are indicated on the branches. Alleles from isolates collected from *L. marginale* are prefixed with Lm. Alleles from isolates collected from *isolated collected from L. usitatissimum* are prefixed with Lu. The number of times each haplotype was recovered is indicated in brackets after each allele code (with a double haplotype inferred for each homozygote).
Table 3.2. Polymorphism and gene diversity estimates for AvrPI23 nucleotide sequences across 33 isolates of Melampsora lini.

<table>
<thead>
<tr>
<th>Lineage</th>
<th>No. of isolates</th>
<th>No. of haplotypes</th>
<th>No. of segregating sites</th>
<th>( \pi )</th>
<th>( \pi^\text{coding} )</th>
<th>( \pi^\text{flanking} )</th>
<th>( \pi^\text{syn} )</th>
<th>( \pi^\text{non} )</th>
<th>( \pi^\text{non}/\pi^\text{syn} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>33</td>
<td>9</td>
<td>118</td>
<td>0.124</td>
<td>0.048</td>
<td>0.105</td>
<td>0.131</td>
<td>1.29</td>
<td></td>
</tr>
<tr>
<td>AA(^1)</td>
<td>15</td>
<td>4</td>
<td>48</td>
<td>0.010</td>
<td>0.003</td>
<td>0.003</td>
<td>0.013</td>
<td>4.74</td>
<td></td>
</tr>
<tr>
<td>AB(^2)</td>
<td>14</td>
<td>5</td>
<td>118</td>
<td>0.159</td>
<td>0.065</td>
<td>0.148</td>
<td>0.165</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>L. u.(^3)</td>
<td>4</td>
<td>4</td>
<td>56</td>
<td>0.070</td>
<td>0.015</td>
<td>0.023</td>
<td>0.084</td>
<td>3.62</td>
<td></td>
</tr>
<tr>
<td>Clade A</td>
<td>8</td>
<td>62</td>
<td>0.013</td>
<td>0.002</td>
<td>0.003</td>
<td>0.017</td>
<td>5.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clade B</td>
<td>1</td>
<td>1</td>
<td>0.0004</td>
<td>0.002</td>
<td>0.002</td>
<td>0</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)Lineage AA isolates collected from Linum marginale  
\(^2\)Lineage AB isolates collected from Linum marginale  
\(^3\)Isolates collected from Linum usitatissimum

Table 3.3. Polymorphism and gene diversity estimates for AvrP4 nucleotide sequences obtained from 33 isolates of Melampsora lini.

<table>
<thead>
<tr>
<th>Lineage</th>
<th>No. of isolates</th>
<th>No. of haplotypes</th>
<th>No. of segregating sites</th>
<th>( \pi )</th>
<th>( \pi^\text{coding} )</th>
<th>( \pi^\text{flanking} )</th>
<th>( \pi^\text{syn} )</th>
<th>( \pi^\text{non} )</th>
<th>( \pi^\text{non}/\pi^\text{syn} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>33</td>
<td>16</td>
<td>80</td>
<td>0.078</td>
<td>0.014</td>
<td>0.086</td>
<td>0.075</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>AA(^1)</td>
<td>15</td>
<td>7</td>
<td>11</td>
<td>0.013</td>
<td>0.003</td>
<td>0.007</td>
<td>0.014</td>
<td>2.02</td>
<td></td>
</tr>
<tr>
<td>AB(^2)</td>
<td>14</td>
<td>7</td>
<td>77</td>
<td>0.112</td>
<td>0.019</td>
<td>0.134</td>
<td>0.106</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>L. u.(^3)</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td>0.012</td>
<td>0.006</td>
<td>0</td>
<td>0.016</td>
<td>( \infty )</td>
<td></td>
</tr>
<tr>
<td>Clade A</td>
<td>14</td>
<td>27</td>
<td>0.020</td>
<td>0.006</td>
<td>0.010</td>
<td>0.040</td>
<td>4.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clade B</td>
<td>2</td>
<td>52</td>
<td>0.028</td>
<td>0.010</td>
<td>0.027</td>
<td>0.028</td>
<td>1.04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)Lineage AA isolates collected from Linum marginale  
\(^2\)Lineage AB isolates collected from Linum marginale  
\(^3\)Isolates collected from Linum usitatissimum

The AvrP123 locus showed the highest level of nucleotide polymorphism (\( \pi = 0.124 \)), with an overall ratio of \( \pi^\text{non} \) to \( \pi^\text{syn} \) of 1.29. All lineages displayed \( \pi^\text{non}/\pi^\text{syn} \) ratios in a range between 1.12 and 4.77, with lineage AB being the lowest, and lineage AA the highest. When only clade A alleles were considered for analysis, the \( \pi^\text{non}/\pi^\text{syn} \) ratio increased to 5.67. Variation within clade B was limited to a single silent change (Table 3.2). Overall levels of polymorphism were comparatively lower at AvrP4 (\( \pi = \))
0.078), with an overall $\pi_{\text{non}}/\pi_{\text{syn}}$ ratio of 0.87. The $\pi_{\text{non}}/\pi_{\text{syn}}$ ratio was 0.78 for lineage AB and 2.02 for lineage AA. Only nonsynonymous changes were recorded at the *AvrP4* locus for the cultivated flax isolates. When only clade A alleles were considered for analysis, the $\pi_{\text{non}}/\pi_{\text{syn}}$ ratio increased to 4.22. Examination of these statistics among the different clades reveals that synonymous divergence between clades A and B are largely responsible for the decreased ratios of synonymous and non-synonymous mutations seen in lineage AB.

Formal selection tests were performed to clarify if these patterns of polymorphism are a result of selection. Tests comparing rates of synonymous ($dS$) and nonsynonymous ($dN$) substitution within the coding regions of each of the haplotypic variants were performed using a maximum likelihood approach (Yang et al. 2000). For *AvrP123* $dN$ exceeded $dS$ in 61 of 66 possible comparisons among allelic variants. For *AvrP4* $dN$ exceeded $dS$ in 80 of 105 possible comparisons. The mean $dN/dS$ ratios were also calculated across the entire coding sequence. Average $dN/dS$ ratios calculated for the entire *AvrP123* and *AvrP4* genes were 4.95 and 1.68 respectively. Table 3.5 lists parameter estimates under models of variable $o_\sigma$ values among sites. Model M2 fitted the data for both avirulence genes significantly better than the M1 model that (Table 3.4), indicating significant levels of positive selection occurring at these loci. Comparison between models M7 and M8 produced analogous results (data not presented). Individual amino acid sites predicted to be subject to diversifying selection across all haplotypes are indicated in Figure 3.2. For *AvrP123*, 23 sites were under significant ($P < 0.05$) levels of positive selection, distributed throughout the length of the protein following the signal peptide region. For *AvrP4*, 6 sites were under significant levels of positive selection ($P < 0.05$), clustered at the 3’ end of the secreted protein.

McDonald-Kreitman tests were also used to test for positive selection. Based on the high level of divergence between the A and B clades, and the clear hybrid status of lineage AB, MK tests were conducted using sequences in clade A for the intraspecific comparison, and both clade B alleles from *AvrP123*. Because of the strong divergence and uncertain origin of the second divergent *AvrP4* B allele, only the common variant from *AvrP4* was used in among-species comparisons. Table 3.5 shows patterns of
synonymous and nonsynonymous variation and the neutrality index (NI) (Yang 1997; Yang et al. 2000) in both Avr genes for polymorphisms within clade A and fixed changes between clades A and B. Together, these data suggest a significant departure from neutrality for both loci, and show a large excess of amino acid polymorphism within species (NI = 10.9 and 4.5 for AvrP123 and AvrP4 respectively). This indicates that positive selection is driving diversification at these loci within clade A.

Table 3.4. dN/dS ratios and parameter estimates generated by maximum likelihood model comparisons\(^1\) for the AvrP123 and AvrP4 genes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Haplotypes</th>
<th>dN/dS *</th>
<th>(\omega)</th>
<th>2Δl</th>
<th>df</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AvrP123</td>
<td>All haplotypes</td>
<td>2.33</td>
<td>0.36</td>
<td>12.52</td>
<td>64.8</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>A haplotypes</td>
<td>7.0</td>
<td>0.45</td>
<td>20.41</td>
<td>46.5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Australian A haplotypes</td>
<td>7.95</td>
<td>0.47</td>
<td>29.96</td>
<td>36.0</td>
<td>2</td>
</tr>
<tr>
<td>AvrP4</td>
<td>All haplotypes</td>
<td>0.99</td>
<td>0.16</td>
<td>6.94</td>
<td>16.2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>A haplotypes</td>
<td>3.23</td>
<td>0.42</td>
<td>7.88</td>
<td>7.0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Australian A haplotypes</td>
<td>2.58</td>
<td>0.62</td>
<td>4.16</td>
<td>2.4</td>
<td>2</td>
</tr>
</tbody>
</table>

* dN/dS ratio calculated using the model M0 (assuming a single evolutionary rate among codons)

\(^1\) For the model comparison, \(\omega\) represents parameter estimates of dN/dS in the highest class and \(p_s\), the proportion of sites estimated to belong to that class. A comparison between M1 (neutral) vs M2 (selection) is presented here. A comparison of models M7 (beta) vs M8 (beta and \(\omega\)) produces highly similar results.

3.3.3. Avr diversity and structure in local populations

Within ten Australian populations, 'A' type AvrP123 and AvrP4 amplicons were recovered from a total of 257 individual isolates. The haplotypes recovered represent a subset of the alleles described at the continental scale (Figs. 3.2, 3.3). In total, four AvrP123 and five AvrP4 haplotypes were recovered from Australian populations in the mountains and plains regions. Levels of nucleotide diversity (\(\pi\)) within local populations (population mean: AvrP123: 0.0029; AvrP4: 0.0041) was much reduced compared to levels of diversity within the wider continental dataset. The frequency and distribution of Avr haplotypes varied among both populations and lineages (Tables 3.6, 3.7). For AvrP123, the number of haplotypes per population ranged between one and four. The lowest diversity was found in the Kiandra and Larras Lee populations; in both cases only a single haplotype was found. The highest diversity...
was in Marrar, where four haplotypes were recovered. For \textit{AvrP4}, the number of haplotypes per population also ranged between one and four. The lowest diversity was found in the Kiandra and SH2 populations, where only single haplotypes were present. The highest diversity was in the Molong population (four haplotypes).

Among lineage AA populations (plains region), high levels of heterozygosity were revealed for Marrar (0.48) and Molong (0.90), at the \textit{AvrPl23} and \textit{AvrP4} loci respectively (Table 3.6). For the remaining lineage AA populations, levels of observed heterozygosity were generally low or nonexistent, for both \textit{AvrPl23} and \textit{AvrP4} (Table 3.6).

\textbf{Table 3.5. Summary of McDonald-Kreitman tests}

<table>
<thead>
<tr>
<th></th>
<th>Fixed differences between clades A &amp; B</th>
<th>Polymorphic sites within clade A</th>
<th>G-test results</th>
<th>NI \textsuperscript{1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{AvrPl23}</td>
<td>Synonymous</td>
<td>19</td>
<td>2</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>Nonsynonymous</td>
<td>47</td>
<td>54</td>
<td>\textless 0.001</td>
</tr>
<tr>
<td>\textit{AvrP4} \textsuperscript{1}</td>
<td>Synonymous</td>
<td>17</td>
<td>2</td>
<td>4.55</td>
</tr>
<tr>
<td></td>
<td>Nonsynonymous</td>
<td>34</td>
<td>18</td>
<td>0.033</td>
</tr>
</tbody>
</table>

\textsuperscript{1}The calculation of fixed differences between clades A and B for \textit{AvrP4} were performed using only the single clade B haplotype Lm-12 (see materials and methods). \textsuperscript{2} Probability determined by a G-test. \textsuperscript{3} \textit{NI} refers to the neutrality index (Rand and Kann 1996).

Population genetic analyses showed strong geographic structure in the data. For \textit{AvrPl23}, AMOVA (\(\Phi_{PT} = 0.29\), \(P = 0.001\)) assigned 20% of the diversity to differences among lineages, 9% between populations within a lineage, and the largest proportion of the variation among isolates within populations (71%). PCoA analysis of pairwise estimates of \(\Phi_{PT}\) among all populations suggests that this structure is largely due to the divergence of three mountains populations, B2, Kiandra and SH2 (Fig. 3.5). For \textit{AvrP4}, AMOVA also revealed significant structure in the data (\(\Phi_{PT} = 0.49\), \(P = 0.001\)). However, none of this variation could be attributed to differences among lineages. Instead, 49% of the diversity was assigned to differences among populations, and 51% of the variation among isolates within populations.

Visualisation of the pairwise estimates of \(\Phi_{PT}\) among populations based on haplotypic data shows no clear pattern of clustering among either lineages or populations (Fig. 3.5).
Table 3.6. Gene diversity statistics for *AvrP123* and *AvrP4* in local populations. 

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Population</th>
<th>No. isolates</th>
<th>( N_A )</th>
<th>( H_0 ) ( (\text{S.D.}) )</th>
<th>Hap. Div. ( \pi ) ( \text{coding} )</th>
<th>( N_A )</th>
<th>( H_0 ) ( (\text{S.D.}) )</th>
<th>Hap. Div. ( \pi ) ( \text{coding} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>Garra</td>
<td>30</td>
<td>2</td>
<td>0.21 (0.09)</td>
<td>0.00064</td>
<td>2</td>
<td>0.13 (0.08)</td>
<td>0.0030</td>
</tr>
<tr>
<td></td>
<td>Gundagai</td>
<td>24</td>
<td>2</td>
<td>0.09 (0.09)</td>
<td>0.00086</td>
<td>2</td>
<td>0.22 (0.06)</td>
<td>0.0032</td>
</tr>
<tr>
<td></td>
<td>Larras Lee</td>
<td>27</td>
<td>1</td>
<td>0.00</td>
<td>0.00000</td>
<td>3</td>
<td>0.32 (0.09)</td>
<td>0.0071</td>
</tr>
<tr>
<td></td>
<td>Marrar</td>
<td>21</td>
<td>4</td>
<td>0.48 (0.06)</td>
<td>0.02355</td>
<td>3</td>
<td>0.16 (0.09)</td>
<td>0.0053</td>
</tr>
<tr>
<td></td>
<td>Molong</td>
<td>21</td>
<td>2</td>
<td>0.26 (0.10)</td>
<td>0.00074</td>
<td>4</td>
<td>0.61 (0.04)</td>
<td>0.0052</td>
</tr>
<tr>
<td>AB</td>
<td>B2</td>
<td>25</td>
<td>3</td>
<td>*</td>
<td>0.38 (0.09)</td>
<td>4</td>
<td>*</td>
<td>0.37 (0.10)</td>
</tr>
<tr>
<td></td>
<td>B3</td>
<td>26</td>
<td>2</td>
<td>*</td>
<td>0.19 (0.09)</td>
<td>2</td>
<td>*</td>
<td>0.19 (0.09)</td>
</tr>
<tr>
<td></td>
<td>Kiandra</td>
<td>33</td>
<td>1</td>
<td>0.00</td>
<td>0.00000</td>
<td>1</td>
<td>0.00</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>N2</td>
<td>28</td>
<td>2</td>
<td>*</td>
<td>0.33 (0.09)</td>
<td>2</td>
<td>*</td>
<td>0.19 (0.09)</td>
</tr>
<tr>
<td></td>
<td>SH2</td>
<td>22</td>
<td>2</td>
<td>*</td>
<td>0.35 (0.09)</td>
<td>1</td>
<td>*</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*For lineage AB isolates, only ‘A’ type alleles were sequenced. If the fixed presence of the ‘B’ type allele in these isolates is assumed, then all populations have an observed heterozygosity of 1.*

Table 3.7. Distribution of *AvrP123* and *AvrP4* haplotypes within and among 10 *Melampsora lini* populations originating from two geographic regions.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele</th>
<th>Plains (Lineage AA)(^1)</th>
<th>Mountains (Lineage AB)(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>AvrP123</em></td>
<td></td>
<td>Garra</td>
<td>Gundagai</td>
</tr>
<tr>
<td>Lm-1</td>
<td>0.93</td>
<td>0.88</td>
<td>1.00</td>
</tr>
<tr>
<td>Lm-2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lm-3</td>
<td>0.07</td>
<td>0.13</td>
<td>0</td>
</tr>
<tr>
<td>Lm-4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>AvrP4</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lm-1</td>
<td>0.94</td>
<td>0.70</td>
<td>0.13</td>
</tr>
<tr>
<td>Lm-3</td>
<td>0</td>
<td>0.30</td>
<td>0.03</td>
</tr>
<tr>
<td>Lm-4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lm-7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lm-8</td>
<td>0.06</td>
<td>0</td>
<td>0.84</td>
</tr>
</tbody>
</table>

\(^1\)For the plains populations, data is presented for lineage AA isolates only

\(^2\)For the mountains populations, data is presented for lineage AB isolates only

Considering all ten populations, for both *AvrP123* and *AvrP4*, Mantel tests revealed no significant association between the degree of genetic population differentiation (\(\Phi_{PT}\)) and geographic distance (km) (*AvrP123*, \( P = 0.12; \) *AvrP4*, \( P = 0.18\)). In addition, there were no significant relationships between genetic differentiation among populations (\(\Phi_{PT}\)) and geographic distance for either lineage AA (*AvrP123*, \( P \))
= 0.35; \textit{AvrP4}, P = 0.57) or lineage AB (\textit{AvrP123}, P = 0.39; \textit{AvrP4}, P = 0.47) populations.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure3_5.png}
\caption{Principal coordinates plots based on genetic distances among populations of \textit{Melampsora lini} (as determined by pairwise $\Phi_{PT}$ estimates) for the A type alleles for (A) \textit{AvrP123} and (B) \textit{AvrP4}. The \% variation explained by each PCoA axis is shown on the figure.}
\end{figure}

**3.3.4. Transient expression of \textit{AvrP123} and \textit{AvrP4} in \textit{L. marginale}**

In order to test whether any of the variation between the \textit{AvrP4} and \textit{AvrP123} genes is associated with differences in recognition by \textit{L. marginale} lines, 8 and 6 \textit{Avr} variants respectively were transiently expressed in a set of \textit{L. marginale} differential lines using \textit{Agrobacterium}-mediated transient expression. In this assay, resistance-gene mediated
recognition of the expressed Avr protein is expected to lead to induction of a HR-like necrotic response. Two AvrP123 variants (AvrP123-Lm-1; AvrP123-Lm-3) recovered from *L. marginale*, and a single AvrP123 variant (AvrP123-Lu-4) recovered from *L. usitatissimum*, induced such a response in one *L. marginale* differential line (line L), indicating that the differential line L has a resistance gene or allele capable of detecting these allelic variants (Fig. 3.6).

![Figure 3.6](image)

**Figure 3.6.** Transient expression of the *M. lini* AvrP123 and AvrP4 genes in *L. marginale* causes Avr gene dependent necrosis. (A). Leaves of a single *L. marginale* genotype (differential line L) displaying a necrotic response following infiltration with *A. tumefaciens* cultures containing the *M. lini* AvrP123 alleles Lm-1 and Lm-3. (B). Leaves of a single *L. marginale* genotype (differential line UU) displaying a necrotic response following infiltration with *A. tumefaciens* cultures containing the AvrP4 alleles Lm-1, Lm-5, Lm-7, Lm-8 and Lu-2.

Chapter 3: Evolution of virulence
AvrP123-Lm_1 and AvrP123 Lm_3 differ by only a single amino acid, suggesting that they may be functionally identical with regards to recognition by the L differential. The remaining AvrP123-differential combinations failed to elicit a necrotic response. Transient expression of 4 AvrP4 variants (AvrP4-Lm-1; AvrP4-Lm-5; AvrP4-Lm-7; AvrP4-Lm-8) recovered from isolates infecting *L. marginale*, induced an HR-like necrotic response in one *L. marginale* differential line (line UU), indicating that this differential has a broad resistance to multiple AvrP4 allelic variants (Fig. 3.6). The remaining Avr-differential combinations failed to elicit a necrotic response. For both AvrP123 and AvrP4, the necrotic response seen in the infiltration experiments was not uniform in all leaves, so that some experimental leaves appeared unaffected. This may be due to insufficient or failed infiltration/transformation in some leaves. However the positive responses were reproducibly observed in multiple infiltration experiments.

3.4. Discussion

3.4.1. Adaptive evolution of AvrP123 and AvrP4

The *AvrP123* and *AvrP4* genes, which function as avirulence elicitors in the interaction with the host plant *Linum usitatissimum* (Catanzariti *et al.* 2006), show high levels of polymorphism in *M. lini* isolates from the Australian native host species, *L. marginale*. Several lines of evidence suggest that selection driven by host-pathogen coevolution is generating and maintaining this diversity. Patterns of nucleotide variation within the coding regions of the *Avr* genes demonstrate a striking and significant excess of non-synonymous compared to synonymous polymorphisms relative to expectations based on neutrality. In addition, patterns of sequence variation at the *Avr* loci when compared to sequence data from non-coding, neutrally evolving regions revealed an order of magnitude more nucleotide polymorphism at *AvrP123* and *AvrP4*. Patterns of sequence variation among local pathogen populations are strongly geographically structured, suggesting that spatially heterogeneous patterns of directional selection (pathogen local adaptation) plays an important role in maintaining genetic variation across the entire species. Furthermore, variants of both *AvrP123* and *AvrP4* can induce HR responses in some *L. marginale* differential lines,
suggesting that resistance genes in this host species confer specific recognition of alleles of these \textit{Avr} loci, and may be the driving force behind the observed \textit{Avr} gene diversification.

The inactivation of avirulence genes can occur via a variety of mechanisms, including alteration of the avirulence protein through sequence mutation, deletion of the entire gene, or mutations resulting in the loss of function (Schurch \textit{et al.} 2004). Here, it has been demonstrated across a genetically and geographically broad sample that functional changes in the \textit{AvrP4} and \textit{AvrP123} genes occur almost exclusively via non-synonymous mutations. No evidence was found for genetic recombination driving sequence variation, nor deletion of the entire gene. None of the observed mutations render the genes non-functional, with the exception of a single, large internal deletion of coding sequence in one \textit{AvrP123} allele. This allele exists in a heterozygous state, meaning that the effects of the mutation on the fitness of the particular isolate in which it is found is difficult to ascertain. However, it is likely that virulence has not been gained for this isolate since it still has a functional copy of \textit{AvrP123}. These results support the hypothesis of positive diversifying selection acting to modify \textit{AvrP123} and \textit{AvrP4} to avoid recognition by plant resistance proteins while simultaneously retaining their effector functions. Similar genetic signatures have been detected at other \textit{R} and \textit{Avr} loci that interact via direct protein-protein interactions. For example, at the flax L resistance locus, 13 alleles describing different resistance specificities to \textit{M. lini} have been demonstrated to be evolving under strong positive selection (Ellis \textit{et al.} 1999). Studies at the interacting \textit{AvrL567} locus of \textit{M. lini} also demonstrate high levels of allelic diversity and provide evidence that diversifying selection is driving this amino acid diversification (Dodds \textit{et al.} 2006).

While these results provide evidence for the selective maintenance of sequence variation at \textit{AvrP123} and \textit{AvrP4}, the high levels of divergence between clades A and B, and the contrasting patterns of polymorphism between them, complicate interpretation of the data. All sequences recovered that were assigned to clade B in lineage AB isolates were highly divergent from clade A alleles, and of uncertain taxonomic and pathogenic origin. To date, it has not proven possible to identify the putative BB parental lineage in Australia, despite sampling at wide geographic scales.
as well as more intensively in local populations. However, the extensive genetic differences between the A and B genomes suggest that the two parental lineages evolved in isolation for a long period of time before the formation of hybrid lineage AB. The formation of this lineage is therefore likely to have involved the migration of one of the parents beyond its normal geographic range, and suggests that the potential for repeated hybridisation events may be geographically constrained. Given that lineage AA isolates have been sampled within the study area extensively, and that there are no other native hosts, it is suggested that the parental BB lineage typically occurs outside our sampling area, where it infects a different host species.

Other than the single, highly divergent \textit{AvrP4} allele in a single lineage AB isolate noted above, clade B \textit{Avr} alleles lack amino acid variation. Such a pattern may be expected where there is little selective advantage to maintaining variation in the clade B alleles. Instead, given the large differences in the levels of nucleotide diversity at \textit{Avr} compared to the neutral loci, it seems likely that strong directional selection has promoted divergence between clades A and B. The lack of synonymous and flanking sequence polymorphism in the majority of clade B \textit{Avr} alleles (with the exception of the variants found in the isolate Tas1) further suggest a relatively recent origin for these alleles. In addition, maximum likelihood analyses for \textit{AvrP123} and \textit{AvrP4} alleles, revealed markedly higher \(dN/dS\) ratios for both genes when clade B alleles were excluded from the analyses. McDonald-Kreitman tests examining variation within clade A alleles, using clade B alleles as a between species comparison, revealed a significant excess of nonsynonymous polymorphisms within clade A, compared to among clades A and B. These patterns of sequence variation and nucleotide polymorphism are consistent with host-pathogen coevolutionary interactions occurring exclusively via clade A alleles at both the \textit{AvrP123} and \textit{AvrP4} loci, with diversifying selection promoting divergence of the A alleles while the B alleles have been fixed.

The role of the B genome alleles in determining pathogenicity in the hybrid lineage AB is therefore uncertain. In \textit{M. lini} (where virulence is a recessive trait), classic studies on the genetics of pathogenicity (Flor 1955), and recent studies exploring the molecular basis of avirulence (Dodds \textit{et al}. 2004; Dodds \textit{et al}. 2006) indicate that all

Chapter 3: Evolution of virulence
allelic copies of genes determining virulence are expressed. Thus, given that lineage AB isolates are fixed for heterozygosity at both *AvrP123* and *AvrP4*, lineage AB isolates should carry more potential targets for host R genes than lineage AA isolates, which have a mean observed heterozygosity of around 25%. However, lineage AB isolates are instead nearly 20% more virulent than lineage AA isolates (chapter 2). Furthermore, the invariance of clade B *Avr* haplotypes suggests that these would seemingly provide an easy target for host R genes. One hypothesis is that lineage AB may be functionally haploid with regard to interactions with *L. marginale* (Nuismer and Otto 2004). Given the high levels of divergence between clade A and B alleles, and the uncertain origin of the B genome, it is possible that *L. marginale* has not evolved the capacity to recognise clade B alleles. This would act to decrease the mean likelihood of detection of lineage AB genotypes by *L. marginale*, as potential avirulence elicitors would only be present in the ‘A’ nucleus.

### 3.4.2. *Avr* diversity and structure in local populations

Patterns of sequence variation at the level of local populations can provide further insight into the various evolutionary forces generating and maintaining *avr* polymorphisms in *M. lini*. Previous studies of the *L. marginale-M. lini* interaction suggest that selection for virulence has the potential to strongly influence local patterns of genetic variation in *M. lini* (Thrall and Burdon 2003). Large amounts of variation for host resistance exist within and among populations, and pathogen populations have never been observed to reach fixation for particular virulence phenotypes (Burdon and Jarosz 1991; Jarosz and Burdon 1991; Burdon et al. 1999; JJ Burdon unpublished data). Comprehensive cross-inoculation trials have also demonstrated strong local adaptation of the pathogen to host populations (Thrall et al. 2002). Furthermore, pathogen virulence and host resistance have been shown to be strongly correlated, so that broadly virulent pathogens occur more frequently in highly resistant host populations, whereas avirulent pathogens dominate susceptible host populations (Thrall and Burdon 2003). Consistent with these earlier results, frequent polymorphisms were found within populations, and strong geographic structure at both *AvrP123* and *AvrP4*. 
Geographically variable selection on \textit{Avr} genes may lead to the maintenance of different functional alleles in partially isolated populations, and therefore to high levels of amino acid polymorphism in the species generally. Strong genetic structure at \textit{AvrP4} ($\Phi_{PT} = 0.49$), and to a lesser extent \textit{AvrP123} ($\Phi_{PT} = 0.29$), among local populations, suggests that local patterns of selection have the potential to contribute to regional maintenance of diversity at Avr loci in \textit{M. lini}. However, highly stochastic demographic forces in populations of \textit{M. lini}, associated with boom-and-bust epidemic dynamics and significant rates of local extinction and re-colonisation, suggest that any genetic polymorphism in \textit{M. lini} may be also strongly influenced by neutral genetic drift. In \textbf{chapter 4} the potential for neutral drift to influence patterns of virulence within lineages of \textit{M. lini} is examined, through comparison of population structure at \textit{AvrP123} and \textit{AvrP4} with population structure determined using phenotypic virulence markers and AFLP markers. Strong structure at neutral (AFLP) loci demonstrates the potential for drift to contribute to the generation of genetic divergence among populations at Avr loci. However, examination of patterns of divergence among populations shows no significant relationships among different marker types (in sexually reproducing populations), suggesting that local selection is more important than drift in driving phenotypic divergence among populations, as well as population divergence in the frequencies of \textit{AvrP123} and \textit{AvrP4} alleles.

Within populations, the maintenance of allelic polymorphisms may be promoted by a range of variables, not all of them necessarily adaptive. Where populations are strongly locally adapted, or frequency dependent dynamics are spatially and temporally asynchronous, dispersal and gene flow among populations may mean that alleles from neighbouring populations might contribute to the maintenance of genetic variation within populations (e.g. Yeaman and Jarvis, 2006). The potential for these kinds of dynamics to generate polymorphisms in local populations of \textit{M. lini} is illustrated by longitudinal data showing the frequency of different pathotypes in two closely proximate populations (Thrall and Burdon 2003). Within each population, locally common pathotypes are maintained at high frequency over multiple years, suggesting that they are locally adapted. These locally common sympatric pathotypes, while highly virulent and able to invade nearby susceptible host populations, are only intermittently present in allopatric populations, and rarely reach high frequencies,
suggesting that isolates dispersing into populations may have lower fitness than locally adapted genotypes.

Selection for polymorphism within populations may also lead to the local maintenance of diversity at \textit{AvrP123} and \textit{AvrP4}. Allelic polymorphisms at \textit{Avr} loci may potentially be maintained by selection within local \textit{M. lini} populations in two major ways: heterozygote advantage and frequency dependent selection. Polymorphisms can be maintained in populations when heterozygotes (at any given locus) exhibit higher fitness than homozygotes [i.e. overdominant selection; (Maruyama and Nei 1981)]. However, as discussed previously, theoretical studies suggest that selection should favour pathogen individuals that express the fewest potential elicitors of resistance in the host. Considering sexually reproducing populations, where individual alleles are not under strong selection to overcome host resistance, it is possible that the pathogenicity effector function of \textit{Avr} genes could be enhanced in the heterozygous state, providing a fitness benefit to the individual where genes for resistance are not common in the host population. Indeed, for one population (Molong; Table 3.6), the frequency of heterozygotes was greater than expected, with 90\% of isolates showing heterozygosity for two alleles at \textit{AvrP4}.

Single population theoretical models of gene-for-gene interactions predict that frequency dependent selection can maintain polymorphisms within pathogen populations (and in corresponding host populations), assuming that particular virulence alleles may be either costly or beneficial depending on the level of resistance in corresponding host populations (Jayakar 1970; Antonovics and Thrall 1994; Roy and Kirchner 2000; Thrall and Burdon 2002). For the interaction between \textit{M. lini} and \textit{L. marginale}, there is no direct empirical evidence for these kinds of fluctuating genetic dynamics taking place within local populations. However, the potential for frequency dependent dynamics to maintain genetic polymorphisms within populations of \textit{M. lini} is evidenced by the existence of considerable diversity for virulence within populations (at both phenotypic and molecular levels). In addition, for \textit{M. lini}, the potential for costs to maintain variation in pathogen virulence has been confirmed via experimental inoculations, which demonstrated trade-offs between spore production and virulence in isolates infecting the Australian host \textit{L.}

Chapter 3: Evolution of virulence

75
marginale (Thrall and Burdon 2003). From the host side, there is also evidence for changes in resistance structure within populations following the impact of a substantial epidemic. Large declines in overall population size were associated with a marked shift in host resistance structure such that the resistance phenotypes at high frequency in the pre-epidemic population declined significantly in the post-epidemic population (Burdon and Thompson 1995).

Clearly not all allelic variants identified are necessarily potential targets for selection imposed by host resistance. For example, many allelic variants common in local populations of M. lini differ by only single or a few amino acids. Although single amino acid mutations can confer functional changes with regard to host recognition (P. Dodds, unpublished manuscript), not all of the variation detected within populations necessarily has coevolutionary implications. Further experimentation examining the performance of different allelic variants on different local hosts (i.e. local adaptation at the molecular level) will help to further resolve the functional significance of these allelic differences, and thus to what extent geographic structure at Avr loci drives local adaptation.

3.4.3. Host lineages
Avr gene homologs in one host-pathogen interaction need not necessarily function as Avr elicitors in related hosts. Thus, the fact that the AvrP123 and AvrP4 genes identified from rust isolates of cultivated flax, L. usitatissimum, can elicit resistance responses in the native host, L. marginale, is of considerable interest. The congeneric status of L. usitatissimum and L. marginale suggests that the ability to recognise AvrP123 and AvrP4 proteins may be the result of shared, ancestral R gene homologs among the two hosts. If this is indeed the case, then the association between M. lini and Linum spp. is likely to be of long evolutionary standing, predating speciation in at least some species. Unfortunately, there is little information available regarding evolutionary relationships among plant species in the genus Linum. However, strong geographic separation (L. usitatissimum has its origin in Central Asia and the Mediterranean: Zeven and Zhokovsky, 1975), and large differences in chromosomal number between L. marginale (2n = 84) and L. usitatissimum (2n = 30) suggest a sustained period of genetic isolation between the two species. Such conservation of
interactions between specific R and Avr gene homologs through speciation events has been demonstrated by Kruijt et al. (2005), who showed that the Cf-4 and Cf-9 R genes mediate recognition of the Avr4 and Avr9 genes of *Cladosporium fulvum* across multiple species in the genus *Lycopersicon*.

However, any protein secreted into the plant cell during early stages of the infection process has the potential, over evolutionary time, to trigger an arms race between the pathogen (to avoid recognition) and the host (to evolve resistance). Thus, an alternative explanation for the shared specificity of *L. marginale* and *L. usitatissimum* is that particular specificities may have evolved independently in the two species as a result of convergent evolution. Such convergent evolution underlies shared R gene specificities in *A. thaliana* (RPM1) and *G. max* (Rpg1-b) to the type III effector protein from *P. syringae* (Ashfield et al. 2004). Clearly, analogous investigation of R gene structure and diversity in populations of *L. marginale* is needed to better understand the evolution of plant disease resistance mechanisms.

The small number of isolates examined from *L. usitatissimum* means that it is difficult to draw firm conclusions regarding the possible existence of genetically discrete lineages of pathogens on different hosts in *M. lini*. However, patterns of nucleotide polymorphism found in the clade A alleles at AvrP123, AvrP4, ITS and β-tubulin are consistent with discrete but very closely related *L. marginale* and *L. usitatissimum* host lineages. Individual haplotypes recovered from isolates collected from the two different host species are largely specific to those hosts (with the exception of a single AvrPJ23 haplotype, which was found in isolates infecting both host species). Previous ecological and evolutionary studies suggest that in Australia *M. lini* is well adapted to *L. marginale*, and that the coevolutionary interaction between *M. lini* and *L. marginale* is long-term. *M. lini* occurs throughout the range of its host and frequently occurs in small, highly isolated *L. marginale* populations. Pathogenicity tests clearly show that Australian isolates have high variability on *L. marginale* differential lines, but very limited variability and lower average virulence on the *L. usitatissimum* set of differential lines (Lawrence 1989). Thus, diversification at the AvrP123 and AvrP4 loci may have began prior to host speciation, followed by subsequent evolution and specialisation on the different hosts.
In contrast, phylogenetic analysis of *AvrP123* and *AvrP4* demonstrates that clade A haplotypes collected from the different host species do not form discrete evolutionary lineages. Furthermore, genetic divergence between Australian clade A haplotypes, and *L. usitatissimum* haplotypes, at the ITS and β-tubulin loci is very small. This suggests the possibility of historical gene flow between the two host lineages. Given the aerially dispersed nature of *M. lini*, long-distance dispersal events and subsequent gene-flow are not unfeasible. For example, long distance dispersal (via intercontinental migration), followed by inter-lineage somatic hybridisation has been previously demonstrated to be responsible for changes in pathogenicity of wheat stem rust (*Puccinia graminis* f.sp *tritici*) populations in Australia (Burdon and Silk 1997). Furthermore, isolates of *L. usitatissimum* have been demonstrated to be virulent against some *L. marginale* host lines (Lawrence and Burdon 1989) (some *L. marginale* hosts are susceptible to all isolates they have been tested against), meaning that should long distance spore dispersal take place, susceptible host substrate is potentially available. However, a more comprehensive examination of diversity in isolates of *M. lini* infecting *L. usitatissimum* would be required before any strong conclusions can be made regarding the existence of genetically discrete host lineages in *M. lini*.

### 3.4.4. Concluding remarks

This study was undertaken to investigate the evolutionary processes that maintain pathogenic variability using the model *Linum-Melampsora* interaction. Overall, the results indicate the importance of undertaking evolutionary studies across multiple populations and a hierarchical range of spatial scales. *AvrP123* and *AvrP4* homologs were conserved across all isolates, with functional changes at these loci occurring almost exclusively via non-synonymous changes. Comparisons between *AvrP123* and *AvrP4* and two neutrally evolving loci, together with selection analyses of the ratio of nonsynonymous to synonymous polymorphism, revealed strong diversifying selection acting on the two *Avr* loci. Importantly, it has been shown that locally divergent patterns of selection are likely to be key drivers of the generation and maintenance of diversity at *AvrP123* and *AvrP4*. Furthermore, the maintenance of allelic diversity within local pathogen populations suggest the potential for negative frequency
dependent selection, overdominant selection, or other non-equilibrium processes to generate and maintain diversity within local populations. A corresponding survey of $R$ gene diversity and structure in populations of $L.\ marginale$ would provide deeper insight into reciprocal patterns of selection, coevolution, and the maintenance of diversity in resistance and virulence in natural populations.
3.5 Literature Cited


Chapter 3: Evolution of virulence


CHAPTER 4

Population structure and diversity across sexual and asexual populations of the pathogenic fungus *Melampsora lini*
4.1 Introduction

Host-parasite interactions are thought to be a major force driving coevolutionary change and generating biological diversity (Thompson 2005). In host-pathogen systems, coevolutionary dynamics are largely contingent upon interactions between host resistance and pathogen infection strategies. Genetic variation in these characters in both host and pathogen populations is common, and thought to be a crucial factor influencing disease dynamics in human (Cooke and Hill 2001), plant (Alexander et al. 1993) and animal (Bishop and Morris 2007) systems. However, the underlying microevolutionary processes that generate and maintain such diversity in natural host-pathogen interactions, and the epidemiological consequences of such variation, remain poorly understood (Thrall and Burdon 2003).

Pathogenic species exhibit extensive diversity in their modes of reproduction (i.e. relative level of sexual vs. asexual reproduction) (Kohn 1995; Milgroom 1996), with potentially profound consequences for levels of genetic variation within populations (Milgroom 1996), population growth rate (Heitman 2006), persistence within populations (chapter 2), and evolutionary change (McDonald and Linde 2002). The role and significance of sexual reproduction in interactions between pathogens and their hosts in particular has been heavily debated (e.g. Keeling and Rand, 1995; Fox et al. 1996; Lively et al. 2004; Salathe et al. 2006). For example, the Red Queen Hypothesis (RQH) predicts that sexual reproduction in host species is an adaptation against rapidly evolving parasites because it enables hosts to respond quickly to the appearance of novel pathogen genotypes through recombination of resistance genes (Hamilton 1980; Clay and Kover 1996). By extension, the RQH also predicts that parasites infecting sexual hosts should be genetically more variable than those infecting asexual hosts because recombination of resistance genes will drive negative frequency-dependent selection on the corresponding virulence genes of parasites (Clay and Kover 1996; Ooi and Yahara 1999).

Spatial and environmental heterogeneity in the interactions between parasites and their hosts can have a major influence on coevolutionary dynamics across landscapes (Thompson 2005). Many host species, particularly those in natural ecosystems, exist
as interacting groups of small, geographically and genetically differentiated populations (Burdon 1992), subdividing pathogen populations into relatively small, discrete units. Such spatial and environmental heterogeneity among pathogen populations is likely to strongly influence spatial patterns of disease incidence and persistence (Thrall and Burdon 1997), with populations of many species shown to undergo frequent local extinctions and recolonisations (Antonovics et al. 1994; Ericson et al. 1999; Thrall et al. 2001a; Smith et al. 2003; Laine and Hanski 2006). Thus, advances in the understanding of broader evolutionary processes in pathogenic species are most likely to come from a metapopulation approach that accounts for populations structured into interconnected demes with ongoing local processes of extinction and recolonisation.

Within such a metapopulation framework, variation in different micro-evolutionary forces among demes may act to generate and maintain resistance and virulence polymorphisms. Local host populations, particularly in natural ecosystems, often vary spatially in the identity and diversity of resistance genotypes present (Jarosz and Burdon 1991; Bevan et al. 1993; Laine and Hanski 2006), and local adaptation of pathogens to their hosts has been demonstrated as a strong driver of pathogen population genetic structure in a number of host-pathogen interactions (Greischar and Koskella 2007). Within populations, negative frequency dependent selection, where pathogen (or host) genotypes have higher fitness when rare, is a common assumption in many models of host-parasite interactions, and may contribute to the maintenance of resistance and virulence polymorphisms within populations (Gillespie 1975; Clarke 1976; Anderson and May 1982). In addition, it is also likely that there is strong potential for non-selective factors, such as random genetic drift, founder events and selection on linked traits, to influence host and pathogen evolution (Parker 1991; Burdon and Thompson 1995; Salathé et al. 2005).

Comparative analyses of geographic variation in genes controlling pathogenic traits and neutral molecular markers can provide insight into the relative roles of natural selection and genetic drift in driving divergence among pathogen populations (Zhan et al. 2005). For example, when selection is important and favours different pathotypes or particular virulence genes in different populations, it would be expected that
among-population variation in pathogenicity traits to exceed that of neutral markers. Conversely, when selection favours similar pathogenic traits or functional genes among populations, estimates of population differentiation calculated for the pathogenic traits are expected to be lower than estimates derived from neutral markers (Lewontin and Krakauer 1973; Latta 2004).

In this study, the focus is on the fungal plant pathogen *Melampsora lini*. In Australia, *M. lini* infects *Linum marginale*, an endemic wild, herbaceous plant species. The interaction between *M. lini* and *L. marginale* follows a gene-for-gene model (Burdon 1994), and spatially structured variation in pathogen virulence has been demonstrated at scales ranging from individual populations to the entire range of the interaction in Australia (Burdon and Jarosz 1992; Thrall et al. 2002; Burdon et al. 2002). In Australia, isolates of *M. lini* have been shown to fall into two distinct lineages (termed AA and AB; (chapter 2) that differ in key genetic, reproductive and life-history traits. Continental scale sampling and microsatellite genotyping shows that while lineage AA isolates have low overall genetic diversity and heterozygosity, lineage AB isolates maintain fixed heterozygosity at corresponding loci. Furthermore, while lineage AB isolates consistently have one allele in common with lineage AA isolates, the second allele is consistently singular to lineage AB. Similarly, sequence data shows fixed heterozygosity in lineage AB isolates for ITS and β-tubulin loci, with one allele identical to the allele recovered from lineage AA isolates, and low nucleotide divergence between the A and B clades (<2%; Chapter 3). No intermediate genotypes have been detected to date. These results (Barrett et al. 2007) suggest that lineage AB is a fixed F1 cross between two genetically divergent lineages of *M. lini* (i.e. a hybrid between lineage AA and a yet to be identified lineage BB), and that sexual recombination (i.e. meiosis) is unlikely to occur in lineage AB. In addition, the apparent inability of lineage AB isolates to effectively endure as resting spores (due to the lack of a sexual cycle) appears to drive disjunct geographic distributions among the lineages (Chapter 2), although the strength of this separation at local scales remains unclear.

A mixture of molecular and phenotypic approaches is used to explore the processes structuring genetic variation in *M. lini* across two distinct biogeographic regions.
(mountains and plains) in the state of New South Wales. These regions differ markedly in a range of factors likely to influence the geographic distributions of lineage AA and AB. The plains populations are comparatively hot and dry compared to the subalpine mountains populations; and the main growing seasons of the host are disjunct between regions (Burdon et al. 1999; chapter 2). Furthermore, in the plains populations, where host plants survive over summer largely as underground rootstock, pathogens survive between epidemics exclusively in the dormant teliospore stage (during which meiosis takes place) on senescent host stems. In contrast, in the mountains populations, pathogens rarely form telia, and instead overwinter largely in the clonal urediospore stage on dormant green host shoots (chapter 2). Given that lineage AB isolates seem unable to effectively form resting teliospores (Barrett et al. 2007), these observations lead to the expectation that lineage AA and AB will be largely geographically disjunct among the mountains and plains. This scenario thus offers an excellent opportunity to explore how genetic diversity and structure in a pathogenic species might be influenced by mode of reproduction. AFLP markers were used to assess the geographic distribution of lineages AA and AB, and combined AFLP markers with phenotypic virulence data and allelic data from two avirulence (Avr) loci to examine patterns of genetic and genotypic diversity within and between lineages AA and AB. These markers were further used to evaluate the relative importance of local selection, neutral drift and gene flow in generating divergence within lineages and among local pathogen populations through comparison of estimates of population genetic structure.

4.2 Materials and Methods

4.2.1 Study sites

The pathogen populations used in this study occur in two geographically and environmentally distinct areas (Fig. 4.1). One of these areas lies within the western plains region of NSW. Of the five plains populations examined here, three occur in a cluster to the north and east (Garra, Molong, Larras Lee), with the other two occurring further to the west and south (Gundagai, Marrar) (Fig. 4.1). Together these populations are part of a formerly patchy, but continuous, distribution of L. marginale across the western plains region of NSW (an area that has been extensively cleared for
agriculture). The second region encompasses the Kiandra and Wild Horse Plains, in the northern part of the Kosciuzko National Park (the mountains populations). While the five mountains populations are all within 20 km of each other, the populations are spatially discrete, and earlier studies show populations in this region are quite distinct with regard to pathogen virulence and host resistance structure (e.g. Thrall et al 2001b) (Fig. 4.1, Table 4.1).

Figure 4.1. Map with insert region showing locations of study regions and populations in New South Wales, Australia.

Phenological and epidemiological patterns in the *Linum-Melampsora* interaction differ markedly between mountains and plains populations. In the mountains region, plants over-winter as underground rootstocks with a few short shoots protected from frost and snow by the surrounding vegetation. With the coming of spring, fresh shoots develop and plants flower in mid- to late-summer before dying back with the onset of autumn frosts. Disease epidemics reach their peak in late summer, and the pathogen overwinters as dormant (clonally reproducing) uredial infections on occasional green shoots (Jarosz & Burdon 1991; chapter 2). In the plains populations, very hot dry summers and mild winters result in a virtual reversal of the patterns observed in the
mountains. Epidemics start in the autumn and reach their peak in early spring. In the plains populations, the pathogen survives the summer drought as dormant teliospores (specialized resting spores which are a necessary precursor to the sexual cycle) which are produced in large quantities as shoots senesce (chapter 2).

4.2.2 Sample collection

Rust samples were collected from mountains populations in the summer of 2004/2005, and from plains populations in the spring of 2005 (mid-epidemic in both regions). At each site, 21-33 single pustule-derived individuals were recovered (Table 4.1) from 40 samples haphazardly collected from different infected plants by rubbing cotton buds across sporulating uredia. In the laboratory, urediospores were inoculated onto the universally susceptible *Linum usitatissimum* cultivar Hoshangabad. Inoculated plants were left in a humid atmosphere overnight before being transferred to a glasshouse. Approximately one week later, single pustules were isolated and put through up to three cycles of increase on Hoshangabad, to ensure that each isolate consisted of a single genotype, and that sufficient urediospores were available for DNA extraction and pathotype analysis.

Table 4.1. Site locations, approximate host population sizes, sample sizes and the number of lineage AA and AB isolates recovered from each population of *Melampsora lini*.

<table>
<thead>
<tr>
<th>Region</th>
<th>Population</th>
<th>Host Population size (approx)</th>
<th>Sample size</th>
<th>n AA</th>
<th>n AB</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mountains</td>
<td>Kiandra</td>
<td>5000</td>
<td>33</td>
<td>0</td>
<td>33</td>
<td>S 35°48', E 148°30'</td>
</tr>
<tr>
<td></td>
<td>SH2</td>
<td>2000</td>
<td>22</td>
<td>0</td>
<td>22</td>
<td>S 35°53', E 148°31'</td>
</tr>
<tr>
<td></td>
<td>B2</td>
<td>500</td>
<td>25</td>
<td>0</td>
<td>25</td>
<td>S 35°48', E 148°34'</td>
</tr>
<tr>
<td></td>
<td>B3</td>
<td>1000</td>
<td>26</td>
<td>0</td>
<td>26</td>
<td>S 35°48', E 148°35'</td>
</tr>
<tr>
<td></td>
<td>N2</td>
<td>1000</td>
<td>28</td>
<td>1</td>
<td>27</td>
<td>S 35°49', E 148°34'</td>
</tr>
<tr>
<td>Plains</td>
<td>Garra</td>
<td>1000</td>
<td>31</td>
<td>30</td>
<td>1</td>
<td>S 33°07', E 148°45'</td>
</tr>
<tr>
<td></td>
<td>Gundagai</td>
<td>1000</td>
<td>26</td>
<td>24</td>
<td>2</td>
<td>S 35°03', E 148°06'</td>
</tr>
<tr>
<td></td>
<td>Larras Lee</td>
<td>500</td>
<td>31</td>
<td>27</td>
<td>4</td>
<td>S 32°59', E 148°51'</td>
</tr>
<tr>
<td></td>
<td>Marrar</td>
<td>200</td>
<td>28</td>
<td>21</td>
<td>7</td>
<td>S 34°49', E 147°22'</td>
</tr>
<tr>
<td></td>
<td>Molong</td>
<td>200</td>
<td>25</td>
<td>21</td>
<td>4</td>
<td>S 33°05', E 148°51'</td>
</tr>
</tbody>
</table>

4.2.3 Pathotype identification

Chapter 4: Population structure and diversity
The virulence of each of 275 isolates was assessed using the ‘standard differential set’ of 11 lines of *L. marginale* that has been used extensively in previous work assessing variation in virulence within *M. lini* populations (Burdon & Jarosz 1991; Jarosz & Burdon 1991; Burdon & Roberts 1995; Thrall *et al.* 2001b). Pathogenicity reactions were scored after 14 days, with the interaction between each pathogen isolate and individual differential host lines being classified as either virulent (1) or avirulent (0). Individual pathotypes were thus defined by their unique combination of infectivity responses across the *L. marginale* differentials. The 10 pathogen populations were then analysed both in terms of the frequency and distribution of individual pathotypes.

4.2.4 DNA Extraction

Prior to DNA extraction, approximately 100 mg of urediospores contained within an eppendorf tube were allowed to germinate overnight in 100 uL of sterile water. Total genomic DNA was extracted from these urediniospores using a DNeasy plant mini kit (Qiagen), following standard protocols.

4.2.5 AFLP analyses

Although microsatellite markers have been developed for *M. lini*, they proved too invariant for population genetic studies (Barrett & Brubaker 2006; *chapter 2*). Here, an AFLP method adopted from Becerra Lopez-Lavalle & Brubaker (2007) was used, with modifications as follows. Eight *EcoK1/ Msel* primer combinations in the selective amplification (GC/GT; ACC/GT; ACT/GG; AGA/GG; GC/GA; ACC/GC; ACT/GC; AGA/GT) were used to generate markers ranging from 50 to 500 bp in size. In each case, the *EcoK* selective primer was labelled with one of the fluorescent dyes FAM, VIC, PET or NED. Selective PCR products were separated on an ABI 3130 automated sequencer (PE Applied Biosystems) with a genescan LIZ 500 internal size standard. Electropherograms were subsequently analysed using GENEMAPPER version 4.0 (PE Applied Biosystems). The intensity of each individual peak was normalised on the basis of the total signal intensity and the peak was considered only if its intensity exceeded a fixed threshold. To test the repeatability of AFLP results, 10 individuals from each sample were completely replicated; no differences between the two sets of samples were observed. Only samples that could be scored unambiguously by eye were retained for further analysis. The presence and absence of AFLP
fragments in each sample was analysed using the software GENEMAPPER v4.0 (PE Applied Biosystems). Fragments obtained using each primer combination were scored as either present or absent for each size-specific locus. AFLP fragments were treated as dominant marker loci with two states, presence (1) or absence (0).

4.2.6 *AvrP123* and *AvrP4* amplification and sequencing

The avirulence genes *AvrP123* and *AvrP4* were initially cloned and characterised in the interaction between *M. lini* and the cultivated flax species *L. usitatissimum* (Catanzariti et al. 2006), where multiple virulence polymorphisms are maintained via amino acid differences in the expressed *Avr* proteins (Dodds et al. 2006). *AvrP4* encodes a 95 amino acid protein with a predicted 28 amino acid cleavable secretion signal peptide and *AvrP123* encodes a 117 amino acid protein with a predicted 23 amino acid cleavable secretion signal peptide. For the interaction between *M. lini* and *L. marginale*, results presented in chapter 2 demonstrated that both *AvrP123* and *AvrP4* are also recognised by *L. marginale* host plants, and that strong diversifying selection is acting at both these loci in Australian populations. This parallel study further demonstrated that lineage AB isolates are consistently heterozygous at these loci, and that haplotype diversity in lineage AB isolates at these loci is confined to alleles from the ‘A’ genome. One of these alleles (the ‘A’ allele), is very closely related to alleles recovered from lineage AA isolates. The second allele (the ‘B’ allele), which is highly divergent, displayed no variation.

Due to the lack of variation in ‘B type’ alleles, primers and PCR protocols were designed to amplify only the ‘A type’ alleles in lineage AB isolates (chapter 3). For *AvrP4*, a region including 180 bp of 5’ flanking sequence, 285 bp of open reading frame, and 103 bp of 3’ flanking sequence (F-CATCAAAATCTAACCNGTAC and R-TTGTTTCAGGATAGATAGTGC) was amplified. PCR amplifications were performed on a Hybaid Express thermocycler under the following conditions: 95 °C for 3 min, 34 cycles at 94 °C for 30 s, 56 °C for 45 s, 72 °C for 90 s, followed by a 4 °C holding step. For *AvrP123* a region including 119 bp of 5’ flanking sequence, 351 bp of open reading frame, and 128 bp of 3’ flanking sequence (F-ATTGTGAACCTTTTGAAGGAC and R-CGCCATGGTATTGTTCAGAC) was amplified. PCR amplifications were performed as for *AvrP4* except for a 58 °C
annealing temperature. PCR products were cleaned using Multiscreen PCR96 filter plates (Millipore) and sequenced directly using the forward primers in conjunction with BigDye™ Terminator Cycle Sequencing Reaction Kit (PE Applied Biosystems). Sequencing products were resolved on an ABI 3130 automated sequencer. To assist with alignment of different sequences, alignments were initially performed on the amino acid sequence, using conserved cystine motifs as reference points and then back-translated to the original nucleotides. Avr sequence variation was analysed by classifying amino acid variants as allelic variants. As noted above, all analyses based on data collected for lineage AB isolates are based on the ‘A’ haplotypes only.

4.2.7 Data analyses

Analyses were conducted at both the between-and within-lineage levels, using different approaches. Analyses of genotypic diversity across different levels, and overall genetic differentiation (i.e. AMOVA and Nei’s genetic distance) were performed using all isolates. For comparative analyses of spatial genetic structure (i.e. θ and ANOSIM comparisons) the two lineages were treated separately. Due to low sample size, for these analyses lineage AB isolates collected from the plains populations and the single lineage AA isolate collected from the mountains populations were excluded.

Genotypic diversity among lineages and populations was calculated based on the number of shared multilocus AFLP or Avr genotypes. Phenotypic diversity was calculated based on the shared number of virulence phenotypes. The genotypic or phenotypic identity of the different isolates was evaluated using the software GENOTYPE (Meirmans & Van Tienderen 2004). Genotypic diversity for lineages and populations was estimated via the number of unique multilocus genotypes, Nei’s diversity index (Nei 1987) (calculated using GENODIVE; Meirmans & van Tienderen 2004). A bootstrap re-sampling method as implemented in GENODIVE was used to test whether these indices were significantly different among lineages.

For all other lineage and population level genetic analyses, a clone-corrected dataset was used. This approach was taken so as to minimise the effects of clonal reproduction during epidemics on estimates of genetic diversity and population
differentiation. Clones were identified based on variation between isolates in AFLPs, virulence phenotypes, and avirulence gene genotypes. AFLP genetic diversity was evaluated as the percentage of polymorphic loci.

To test for random associations among AFLP markers, multilocus linkage disequilibrium was measured by calculating the Index of Association, $I_A$ (Maynard Smith et al. 1993), using the software MULTILOCUS v2.2 (Agapow & Burt 2001). The observed $I_A$ for each population and each lineage was compared with an expected $I_A$ under random mating simulated through the reshuffling of data over 1,000 permutations. $I_A$ has an expected value of zero if there is no association of alleles at unlinked loci as expected in a randomly mating population.

The degree of genetic relatedness (AFLP data) among all pairs of populations was evaluated using unweighted pair-group mean analysis (UPGMA) based on estimates of Nei's genetic distance (Nei 1972) calculated using GENALEX v6.0 (Peakall & Smouse 2005). The UPGMA tree was constructed using NT-SYSpc v 2.11 (Rohlf 1993). The genetic similarity among isolates (AFLP) within lineages was visualised using principal coordinate analysis (PCoA) based on Euclidean distances between AFLP multilocus genotypes using GENALEX v6.0.

To examine hierarchical partitioning of molecular variation among lineages and populations, AFLP data were analysed using analyses of molecular variance (AMOVA), using GENALEX v6.0. To test for genetic structure among populations within lineages, a Bayesian approach was used to estimate an $F_{ST}$ analogue among populations within each lineage (denoted $\theta^B$) for the AFLP data, and the two Avr loci, as implemented in the software Hickory (Holsinger et al. 2002). The data were run with the default parameters (burn in = 50000, number of samples = 250000, thinning factor = 50). The data were analysed using the $f$-free model, which does not assume any prior knowledge regarding the degree of inbreeding within populations. Comparisons of $\theta^B$ between datasets were performed by calculating the difference between paired random samples of the posterior distribution of $\theta^B$ for each data set, and declaring the difference in $\theta^B$ significant if the 95% credible interval (CRI) of this difference excluded zero. This approach allowed us to estimate $\theta^B$ for both AFLP and
Avr gene loci, and to formally test hypotheses regarding variation in genetic structure among neutral and selected markers. For comparative purposes genetic structure was also estimated within lineages with AMOVA, using the software GENALEX v6.0.

To estimate levels of geographic structure in the virulence phenotypic data, analysis of similarity (ANOSIM) (Clarke 1993) was used. This approach was adopted because the nature of the phenotypic virulence data precluded more conventional approaches to measuring population divergence (e.g. Fst or Qst, or their analogues). Each virulence phenotype is under direct genetic control, and is controlled by a limited but undetermined number of genes, each with major effects. Each individual response on each differential line is thus informative, but can not be considered as an independent marker, because the phenotype is determined by multiple genes. At the same time, the data cannot be considered as a continuously distributed quantitative trait, because each phenotype is controlled by genes with major effects, and is quantified in a binary manner. Similar to AMOVA, ANOSIM can be used to assess similarity among individuals when grouped at different hierarchical levels. The method does not make assumptions about the nature of the data, and uses rank order of dissimilarity values among individual samples as its base. Under the null hypothesis of no differentiation, the test statistic, R, changes little when population labels are rearranged randomly. R values range between 1 (maximum separation among populations) and 0 (completely random grouping) (Clarke 1993). The associated test for significance was calculated over 1000 permutations. For comparative purposes, these analyses were further performed using the AFLP dataset. ANOSIM was performed using the software PRIMER v6 (PRIMER-E Ltd, Plymouth Marine Laboratory, Plymouth).

To explore how much AFLP genetic variation between populations (within a lineage) was explained by geographical distance, matrices of pairwise genetic distance (Nei 1972) and geographic distance were subject to a Mantel test (Mantel 1967) using GENALEX v6.0 (Peakall & Smouse 2005) with 10000 permutations.

4.3 Results

4.3.1 Patterns of diversity among isolates
AFLP multilocus genotypes, virulence phenotypes and genotypes at the two *Avr* loci were recovered for 275 isolates of *Melampsora lini*, sampled from 10 natural populations across two geographic regions. Among the mountains populations, lineage AB isolates were sampled exclusively, with the exception of a single lineage AA isolate in population N2. Among the plains populations, isolates from both lineages were sampled in all populations, although isolates belonging to lineage AA were predominant (population average: 87%; Table 4.2).

For the AFLP dataset, a total of 208 AFLP markers were scored and retained for further analysis. Of these, 88 markers were polymorphic across the data set. Of the 208 fragments, 60 were unique to lineage AB, and four were unique to lineage AA. In total, 124 isolates were assigned to lineage AA, and 151 isolates to lineage AB.

Genetic diversity as determined by AFLP was generally low, with only 12% and 14% of all scored AFLP markers being polymorphic within lineages AA and AB respectively.

**Table 4.2.** Lineage composition, AFLP and *Avr* multilocus genotypic diversity and virulence phenotypic diversity for plains and mountains populations of *Melampsora lini*.

<table>
<thead>
<tr>
<th>Region</th>
<th>Population</th>
<th>n</th>
<th>nAA</th>
<th>nAB</th>
<th>AFLP MLG</th>
<th>Vir Phen</th>
<th>Avr MLG</th>
<th>Avr div</th>
<th>AFLP div</th>
<th>Vir div</th>
<th>Avr div</th>
<th>ncc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plains</td>
<td>Garra</td>
<td>31</td>
<td>30</td>
<td>1</td>
<td>16</td>
<td>7</td>
<td>5</td>
<td>0.89</td>
<td>0.69</td>
<td>0.38</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gundagai</td>
<td>26</td>
<td>24</td>
<td>2</td>
<td>19</td>
<td>16</td>
<td>7</td>
<td>0.96</td>
<td>0.94</td>
<td>0.80</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Larras Lee</td>
<td>31</td>
<td>27</td>
<td>4</td>
<td>15</td>
<td>10</td>
<td>6</td>
<td>0.88</td>
<td>0.58</td>
<td>0.49</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marrar</td>
<td>28</td>
<td>21</td>
<td>7</td>
<td>25</td>
<td>20</td>
<td>7</td>
<td>0.99</td>
<td>0.97</td>
<td>0.79</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molong</td>
<td>25</td>
<td>21</td>
<td>4</td>
<td>23</td>
<td>10</td>
<td>7</td>
<td>0.99</td>
<td>0.86</td>
<td>0.56</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>141</td>
<td>123</td>
<td>18</td>
<td>92</td>
<td>46</td>
<td>19</td>
<td>0.98</td>
<td>0.92</td>
<td>0.83</td>
<td>119</td>
<td>84</td>
</tr>
<tr>
<td>Mountains</td>
<td>B2</td>
<td>25</td>
<td>0</td>
<td>25</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>0.48</td>
<td>0.52</td>
<td>0.49</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B3</td>
<td>26</td>
<td>0</td>
<td>26</td>
<td>7</td>
<td>9</td>
<td>3</td>
<td>0.47</td>
<td>0.76</td>
<td>0.22</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kiandra</td>
<td>33</td>
<td>0</td>
<td>33</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0.12</td>
<td>0.19</td>
<td>0</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N2</td>
<td>28</td>
<td>1</td>
<td>27</td>
<td>9</td>
<td>13</td>
<td>5</td>
<td>0.76</td>
<td>0.87</td>
<td>0.43</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SH2</td>
<td>22</td>
<td>0</td>
<td>22</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>0.48</td>
<td>0.45</td>
<td>0.31</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>134</td>
<td>133</td>
<td>25</td>
<td>19</td>
<td>6</td>
<td>0</td>
<td>0.65</td>
<td>0.74</td>
<td>0.59</td>
<td>84</td>
<td></td>
</tr>
</tbody>
</table>

`n` = sample size in each population. AFLP MLG = number of AFLP multilocus genotypes. Vir Phen = number of virulence phenotypes. Avr MLG = number of genotypes considering both *AvrP123* and *AvrP4*. AFLP div = Simpson’s diversity index for AFLP markers. Vir div = Simpson’s diversity index.
for virulence phenotypes. \( Avr \text{ div} = \) Simpson’s diversity index for \( Avr \) genotypes. \( n_{\text{nc}} = \) number of clone corrected genotypes using all genotype and phenotype information. Mountains and plains regions differ significantly at \( P \leq 0.01 \).

Pathogenicity analyses revealed extensive variation across all isolates, with 62 different virulence patterns detected across the 11 host differential lines. The number of host lines overcome by individual isolates ranged between 3 and 10. Across all isolates, 4 and 5 clade ‘A’ amino acid variants were recovered from \( AvrPI23 \) and \( AvrP4 \) respectively. The sequence data for each allele is deposited in Genbank (accession numbers for \( AvrPI23 \): EU642492, EU642493, EU642494, EU642495; for \( AvrP4 \): EU642476, EU642478, EU642479, EU642482, EU642483). Isolates with identical AFLP multilocus genotypes frequently represented multiple virulence phenotypes, and isolates with the same virulence phenotype had different AFLP genotypes. Similarly, identical \( Avr \) genotypes harboured multiple virulence phenotypes, and vice versa.

![UPGMA phylogram based on pairwise estimates of Nei’s genetic distance (AFLP) among populations of Melampsora lini.](image)

**Figure 4.2.** UPGMA phylogram based on pairwise estimates of Nei’s genetic distance (AFLP) among populations of *Melampsora lini*.

### 4.3.2 Variation between lineages

The two lineages differed significantly in levels of genotypic diversity as determined by AFLP markers, \( Avr \) genes, and virulence phenotypes. In all three cases, lineage AA
harboured significantly \((P < 0.01)\) more genotypic variation than lineage AB (Table 4.3). Combining all of these marker systems, the total number of genotypes in lineage AA was 100 (of 124 isolates), compared to 61 (of 151) in lineage AB. For lineage AB, 18 of these isolates were recovered from plains populations, 16 of which were genotypically unique.

Overall, AMOVA revealed significant and high levels of structure in the dataset \((\Phi_{PT} = 0.94, P = 0.001)\), with 90% of the AFLP diversity attributed to differences among lineages, with the remainder among isolates within populations (6%), or between populations of the same lineage (4%). Consistent with these observations, the UPGMA phylogram based on the genetic distance among populations showed very strong separation between plains and mountains populations (Fig. 4.2).

**Table 4.3.** Summary statistics describing genotypic and phenotypic variation in 124 and 151 isolates of *Melampsora lini* belonging to lineage AA and lineage AB respectively. PLP = percent loci polymorphic. \(I_A\) = Index of association

<table>
<thead>
<tr>
<th>Marker</th>
<th>Measure of diversity</th>
<th>Lineage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AA</td>
</tr>
<tr>
<td>No. of isolates</td>
<td>124</td>
<td>151</td>
</tr>
<tr>
<td>No. of CC* isolates</td>
<td>100</td>
<td>61</td>
</tr>
<tr>
<td>AFLP</td>
<td>83</td>
<td>25</td>
</tr>
<tr>
<td>Simpson’s Diversity*</td>
<td>0.98</td>
<td>0.68</td>
</tr>
<tr>
<td>Shannon’s Index*</td>
<td>1.77</td>
<td>0.77</td>
</tr>
<tr>
<td>(I_A)</td>
<td>0.71(^*)</td>
<td>1.47(^*)</td>
</tr>
<tr>
<td>PLP</td>
<td>12</td>
<td>14.4</td>
</tr>
<tr>
<td>Avirulence loci</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>Simpson’s Diversity*</td>
<td>0.78</td>
<td>0.62</td>
</tr>
<tr>
<td>Shannon’s Index*</td>
<td>0.8</td>
<td>0.53</td>
</tr>
<tr>
<td>Virulence phenotypes</td>
<td>34</td>
<td>31</td>
</tr>
<tr>
<td>Simpson’s Diversity*</td>
<td>0.90</td>
<td>0.28</td>
</tr>
<tr>
<td>Shannon’s Index*</td>
<td>1.22</td>
<td>0.95</td>
</tr>
<tr>
<td>Mean Virulence*</td>
<td>0.51</td>
<td>0.58</td>
</tr>
</tbody>
</table>

*CC = clone corrected using all genetic and phenotypic data.
*Lineage AA and AB differ significantly at \(P < 0.01\).
*Observed \(I_A\) values differ significantly from that expected under random mating at \(P < 0.01\)
PCoA analysis of the Euclidean distance among lineage AA and AB isolates highlights clearly contrasting AFLP genetic structure within the two lineages (Fig. 4.3). Lineage AA isolates, while showing structure at the population level, form a relatively continuous cloud of points across axes 1 and 2 (explaining 36.5% and 20.1% of the variation respectively) (Fig. 4.3A). In contrast, lineage AB isolates form a series of four relatively discrete clusters along PCoA axis 1 (axes 1 and 2 explain 56.1% and 17.1% of all variation respectively) (Fig. 4.3B).

The Index of Association ($I_A$) statistic revealed significant levels of non-random association among AFLP markers for both lineage AA and AB. However, while significant for both lineages, $I_A$ was more than twice as high in lineage AB as in lineage AA. In conjunction with previous results showing fixed heterozygosity at microsatellite and *Avr* loci (chapter 2), these elevated levels of linkage disequilibrium in lineage AB further support the suggestion that this lineage only reproduces clonally. Significant levels of $I_A$ and the Bayesian estimate of the fixation index ($f = 0.7$) for lineage AA suggest frequent inbreeding within this lineage. The frequency distribution of the different AFLP multilocus genotypes within lineage AB further suggest that clonal reproduction is dominant in this lineage, with two distinct AFLP genotypes representing more than 70% of all sampled isolates (Fig. 4.4B). In contrast, while identical AFLP genotypes were present within the lineage AA samples, no single genotype reached frequencies of higher than 10% across populations (Fig. 4.4A).

Similarly, virulence phenotypes show a distinctly uneven frequency distribution in lineage AB compared to lineage AA, and the two lineages also displayed distinctly different virulence structures in terms of the identity and frequency of individual pathotypes (Fig. 4.5A). The specific differences among the two lineages in their ability to overcome different resistance genes are indicated by striking differences in the frequency with which pathogen isolates from the two lineages were able to overcome individual host differential lines (Fig. 4.5B).
4.3.3 Variation among populations

Within each lineage, among-population pathogen diversity as determined by all three marker systems varied widely (Table 4.2). Of the mountain populations (lineage AB), Kiandra was notable for having especially low levels of AFLP and virulence diversity, being dominated by a single clone corrected genotype. In contrast, diversities as indicated by virulence phenotypic diversity and Avr markers within population N2 were equal, or higher, than in several plains (lineage AA) populations. Within lineage AA, low levels of diversity were observed in the Larras Lee and Garra populations compared to other plains populations (Table 4.2).
Figure 4.4. Histograms of frequencies of different AFLP multilocus genotypes including more than one individual out of a total of 275 isolates of Melampsora lini collected from plains and mountains populations. Lineage AA and AB isolates had no multilocus genotypes in common. (A) Lineage AA isolates. A total of 83 multilocus genotypes were recovered from 124 isolates, with 11 occurring more than once. (B) Lineage AB isolates. A total of 25 multilocus genotypes were recovered from 151 isolates, with 10 occurring more than once.
Figure 4.5. Virulence profile of lineage AA and AB isolates. (A) Histograms of frequencies of different virulence phenotypes (including more than one individual) among 124 lineage AA and 151 lineage AB isolates of *Melampsora lini* as determined on 11 differential host lines collected from plains and mountains populations. Individual virulence phenotypes represent a unique combination of virulence specificities on the differential set, and have been assigned an arbitrary number as shown on the x-axis. (B) Mean virulence responses of isolates of *Melampsora lini* belonging to lineage AA and AB on differential lines of the host plant *Linum marginale*.

Within lineages, estimates of divergence among populations revealed consistent patterns of genetic and phenotypic differentiation. Bayesian estimates of $\theta^B$ based on AFLP, and avirulence gene data resulted in much lower DIC values for models allowing population structure than for models allowing no population structure (Table 4.4), providing evidence for population differentiation across all markers. For the AFLP data, estimates of population structure were high for both lineages ($\theta^B = 0.336$ for AA, $\theta^B = 0.367$ for AB). Strong population structure within lineages was also revealed for *AvrP4*, although the genetic structure among lineage AB populations was weaker ($\theta^B = 0.325$ for AA, $\theta^B = 0.186$ for AB). In contrast, estimates of population
structure for *AvrP123* were much lower, although still significant, for both lineages ($\theta^B = 0.112$ for AA, $\theta^B = 0.046$ for AB). Posterior comparisons of $\theta^B$ revealed significantly lower levels of population differentiation for *Avr* loci compared to AFLP data, with the exception of the estimate of $\theta^B$ for *AvrP4* among lineage AA populations, which was not significantly different from estimates of population differentiation for AFLP markers ($\theta^B = 0.325$ and 0.336 respectively) (Table 4.4). Analysis of structure in the phenotypic virulence data using analysis of similarity also revealed significant spatial structure within both lineages (lineage AA: $R = 0.27$; lineage AB: $R = 0.12$), indicating moderate and low levels of separation among populations respectively. Within lineages, Mantel tests revealed no significant association between the degree of population differentiation as estimated by AFLP markers ($\Phi_{PT}$) and geographic distance (km) (lineage AA, $P = 0.08$; lineage AB, $P = 0.08$).

**Table 4.4.** Estimates of population structure ($\theta^B$), standard deviations (s.d.) and credible intervals (CRI) from Bayesian modelling for AFLP fingerprints and *Avr* loci for two genetic lineages of *Melampsora lini*. Deviance information criteria (DIC) are provided for alternative models ($\theta^B = \text{variable}$ and $\theta^B = 0$).

<table>
<thead>
<tr>
<th>Marker</th>
<th>Lineage</th>
<th>$\theta^B$</th>
<th>s.d.</th>
<th>95% CRI</th>
<th>DIC $\theta^B = \text{variable}$</th>
<th>DIC $\theta^B = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFLP</td>
<td>AA</td>
<td>0.336</td>
<td>0.042</td>
<td>(0.220, 0.382)</td>
<td>362.13</td>
<td>785.27</td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>0.367</td>
<td>0.068</td>
<td>(0.243, 0.502)</td>
<td>205.88</td>
<td>372.14</td>
</tr>
<tr>
<td><em>AvrP123</em></td>
<td>AA</td>
<td>0.112</td>
<td>0.041</td>
<td>(0.003, 0.154)</td>
<td>70.1</td>
<td>79.2</td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>0.046</td>
<td>0.071</td>
<td>(0.016, 0.287)</td>
<td>114.41</td>
<td>122.66</td>
</tr>
<tr>
<td><em>AvrP4</em></td>
<td>AA</td>
<td>0.325</td>
<td>0.076</td>
<td>(0.198, 0.491)</td>
<td>104.73</td>
<td>264.73</td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>0.186</td>
<td>0.068</td>
<td>(0.077, 0.339)</td>
<td>119.84</td>
<td>163.34</td>
</tr>
</tbody>
</table>

Relationships between patterns of phenotypic and genetic divergence among populations were not concordant among lineages. For lineage AA, pairwise comparisons of among-population differentiation as determined by AFLP multilocus genotypes ($\Phi_{PT}$), *Avr* gene genotypes ($\Phi_{PT}$), and virulence phenotypes ($R$) were not significantly correlated (Mantel tests of matrix comparisons, $P > 0.05$). In contrast, for clonal lineage AB, comparisons of pairwise differentiation among the different
markers using Mantel tests revealed a number of statistically significant relationships (for AFLP vs virulence phenotypes, \( P = 0.01, R = 0.79 \); for AFLP vs \( Avr \) multilocus genotypes, \( P = 0.01, R = 0.93 \); for virulence phenotypes vs \( Avr \) multilocus genotypes \( P = 0.02, R = 0.8 \)). \( AvrP123 \) was significantly correlated to both AFLP genotypes (\( P = 0.04, R = 0.88 \)) and virulence phenotypes (\( P = 0.03, R = 0.75 \)). There were no significant relationships between \( AvrP4 \) and either AFLP genotypes or virulence phenotypes (\( P > 0.3 \)).

4.4 Discussion

The geographic mosaic theory of coevolution predicts that traits influencing the outcome of coevolutionary interactions will differ among populations, thereby generating a selection mosaic that varies across landscapes (Thompson 1994, 2005). In this context, our results have important implications for understanding how spatial, environmental and life-history variation influence the population and regional genetic structure of \( M. lini \). Through the integration of AFLP, \( Avr \) gene and phenotypic virulence data, strong genetic and pathogenic divisions among populations of \( M. lini \) within two biogeographic regions are revealed, driven by key differences between pathogen lineages in genetic composition and mode of reproduction. Within this broad genetic and geographic framework, selection and genetic drift interact to generate genetic and pathogenic divergence among local populations.

4.4.1 The geographic distribution of lineage AA and AB

Analysis of molecular and phenotypic virulence data at the regional scale confirm the strong genetic separation between the \( M. lini \) lineages AA and AB previously suggested by geographically broad sampling (chapter 2). This pattern and the dominance of lineage AA isolates in the plains environment is likely a consequence of life-history differences associated with the hybrid origin of lineage AB. The loss or lack of the sexual mode of reproduction effectively restricts lineage AB isolates to the clonal (urediospore) stage of the lifecycle (chapter 2). In the hotter, drier, plains populations, the regular occurrence of summer drought results in the widespread death or senescence of the host, \( L. marginale \). Without green host tissue, \( M. lini \) is unable to survive in the uredial stage, meaning that telial formation is critical for long-term
persistence. A severe summer drought in the year prior to sampling during which isolates in three monitored plains populations survived exclusively as telia, suggests that lineage AB isolates recovered in plains populations are likely to have dispersed from more mesic sites during that growing season. In some years, summer rains may result in year-round persistence of green host tissue, reducing the likelihood of local extinction of lineage AB isolates in plains populations.

In contrast to the plains environment, host phenology is essentially reversed in the mountains region, and the striking dominance of lineage AB isolates in the mountains populations is more difficult to explain. The presence of only a single lineage AA isolate in the five mountains populations implies strong barriers to the establishment of this lineage in this environment. Given that mountains hosts are not universally or even unusually resistant to lineage AA isolates (L.G. Barrett, unpublished ms), this distribution is unlikely to be driven by among-region differences in host resistance. Instead, it is possible that an increased tendency to form telia during warm weather, while crucial to survival in the plains, may affect other fitness parameters relating to disease spread such as spore production and transmission. Consistent with this idea, glasshouse experiments demonstrate that lineage AB isolates are significantly less likely to form telia under warm conditions than lineage AA isolates (chapter 2). Lineage AB isolates may therefore be better at exploiting host plants during the warm summer months, which coincides with the main growing season for *L. marginale* in the mountains.

4.4.2 Genetic and pathogenic structure within lineages

It is generally expected that populations able to reproduce sexually will generate and maintain higher genotypic diversity than asexual populations. Indeed, a number of studies show greater phenotypic and genetic diversity in sexual compared to asexual populations of plant pathogens (Milgroom 1996). Here, it is demonstrated that levels of genotypic diversity within lineages of *M. linii* are consistent with such expectations, with lineage AA harbouring significantly higher AFLP, Avr, and virulence diversity than lineage AB despite census populations being generally smaller and more isolated in the plains region where AA predominates (Burdon *et al.* 1999). The two lineages also differ strikingly in how variation is distributed among isolates. Under strict
asexual reproduction, founding genotypes should diversify over time into discrete clonal lineages characterised by a complex of closely related genotypes. In contrast, the exchange of genetic information among individuals in sexually reproducing populations should result in a more homogenous population structure. Consistent with this expectation, PCoA analysis of the AFLP data showed that genomic variation is distributed relatively evenly across isolates for lineage AA, while lineage AB isolates cluster into a few discrete genotypic groups.

For lineage AA isolates in the plains populations, the high diversity of genotypes in each population suggests that sexual reproduction is common, despite the strong potential for clonal multiplication during epidemics. Only at Larras Lee, and to a lesser extent, Garra, was evidence found for clonal dynamics strongly influencing population structure. Generally, high genotypic diversity was maintained in plains populations despite evidence for significant levels of inbreeding. The low number of identical clones in these populations may reflect a high number of sexual recombinants as primary inoculum, or selection for different virulence phenotypes by individual host plants. While all populations were sampled during established phases of infection, these results represent only a snapshot of pathogen population structure during a single epidemic. Clearly, the clonal structure of populations may change throughout the growing season, with individual pathotypes having the potential to rapidly increase in frequency as epidemics progress (Jarosz & Burdon 1991; Burdon & Jarosz 1992).

Although the population structure detected for the mountains (lineage AB) populations was consistent with that expected under clonal reproduction, within-population diversity varied markedly, with some populations maintaining unexpectedly high levels of variation. For example, in population N2, 21 of the 28 isolates sampled were genetically unique, with AFLP and virulence diversity being higher than some of the sexually reproducing lineage AA populations. In contrast, only six genetically unique individuals from 33 sampled isolates in Kiandra were sampled, reflecting dominance by a single pathotype. Interestingly, this pathotype is also the most common pathotype across all mountains populations and has dominated Kiandra for many years (Burdon & Jarosz 1991; Thrall et al. 2001b; JJ Burdon...
unpublished data). Thus, while patterns of diversity in some populations suggest strong selection for variation in virulence, low levels of diversity in others agree with the general-purpose-genotype hypothesis (Lynch 1984) that predicts selection of a few asexual genotypes characterised by broad tolerances for fluctuating environmental or biotic conditions.

Although difficult to evaluate, given that the putative BB parent has not been detected, the presence of multiple shared \textit{Avr} gene alleles between AA and AB isolates suggests that either multiple hybridisation events, multiple origin of the same alleles, or somatic exchange of genetic information, must have occurred at some stage. Genetically discrete isolates within clusters of closely related AFLP genotypes, and isolates with identical AFLP multilocus genotypes, but different virulence phenotypes, are most likely to have been generated via stepwise patterns of mutation, following establishment of a founding genotype. Stepwise mutation has been demonstrated as an important mechanism for generating diversity in populations of asexual pathogen species (Jimenez-Gasco et al. 2004; Fisher et al. 2005; Hovmøller & Justetson 2007), however horizontal gene exchange may also be important in driving divergence between otherwise very closely related isolates. Thus, although lineage AB \textit{Avr} genotypes and AFLP multilocus genotypes on average are strongly linked, there were several examples where lineage AB isolates with identical AFLP genotypes carried different \textit{Avr} gene alleles. While the generation of these genotypes via multiple hybridisation events between identical AFLP genotypes cannot be discounted, horizontal gene exchange via hyphal fusion among genetically different individuals seems equally parsimonious. The contribution of such non-sexual genetic exchange to generating pathogen diversity has been clearly demonstrated in several species (Spiers & Hopcroft 1994; Burdon & Silk 1997; Friesen et al. 2006).

Higher overall diversity in plains compared to mountains populations could significantly impact on host-pathogen coevolutionary interactions at the regional scale. As suggested by the Red Queen Hypothesis, where hosts and pathogens are naturally coevolving, high levels of diversity in pathogen populations may promote higher diversity through sexual reproduction in corresponding host populations (Haldane 1949). Thus, the results presented here are particularly intriguing when...
interpreted in light of host resistance variation and mating system in plains and mountains environments (Burdon et al. 1999; Thrall et al. 2001b). In the mountains populations (lineage AB), hosts are almost completely selfing. In contrast, in the plains environment, where the majority of isolates reproduce sexually, hosts show significant levels of outcrossing. These differences in host mating system are reflected in significant variation among regions in the diversity of resistance phenotypes, with host populations in the plains region showing consistently higher values. Furthermore, strong differentiation between pathogen lineages in virulence responses to individual host differential lines, and in the identity of associated virulence phenotypes between lineages, corresponds to the observed divergence between mountains and plains host populations in the identity of resistance phenotypes (Burdon et al. 1999).

4.4.3 Genetic structuring among populations

Analysis of the AFLP data demonstrates relatively strong genetic divergence between populations of both lineages. This structure is maintained despite the fact that *M. lini* produces large numbers of urediospores that have the potential to disperse aerially over long distances. Indeed, the presence of lineage AB isolates in all plains populations at low frequencies, and the lack of evidence for any isolation by distance effects, suggests that long-distance spore dispersal is both stochastic and relatively frequent. Despite this, levels of among-population genetic differentiation are considerably higher than those reported for many wind-dispersed, foliar plant pathogens of cultivated crops, such as *Mycosphaerella graminicola* (*Gst* = 0.05; Linde et al. 2002), *Phaeosphaeria nodorum* (*θ* = 0.05; Pimental et al. 2000), and *Tapesia yallundae* (*θ* = −0.008, Douhan et al. 2002). It is important to note that these agricultural examples largely represent pathogens of high density crops, which can survive saprophytically to varying degrees.

Strong bottlenecks during epidemic troughs and the stochastic nature of re-colonisation processes are likely to be key factors driving divergence among populations (Brown and Hovmøller, 2002). Given that *M. lini* is a biotrophic foliar pathogen, and occurs in small, fragmented and genetically diverse host populations, these kinds of stochastic demographic dynamics are common, with annual boom-and-bust epidemic dynamics likely promoting frequent local extinction, stochastic
recolonisation and subsequent genetic drift (Wade & McCauley 1988; Burdon & Jarosz 1992; Ingvarson 1997; Pannell & Charlesworth 1999; Haag et al. 2006). Heterogeneity in terms of host availability determined by resistance structure and environmental conditions further limit opportunities for the establishment and persistence of pathogen migrants which are likely to be accentuated by clonal reproduction (lineage AB), and inbreeding during the sexual phase of the life-cycle (lineage AA).

Previous studies (largely within the mountains) suggest that selection for virulence in the \( L. \ marginale-M. \ lini \) interaction can strongly influence local patterns of genetic variation in \( M. \ lini \). Considerable variation for host resistance exists within and among host populations (Burdon & Jarosz 1991; Jarosz & Burdon 1991) and comprehensive cross inoculation trials have demonstrated strong local adaptation of the pathogen to host populations (Thrall et al. 2002). Furthermore, pathogen virulence and host resistance are highly correlated, such that broadly virulent pathogens occurred more frequently in highly resistant host populations, whereas avirulent pathogens dominate susceptible host populations (Thrall & Burdon 2003). Consistent with expectations arising from these observations, significant structuring was found between populations, in both lineage AA and AB, for virulence phenotypes, and both \textit{Avr} gene loci.

Direct comparisons of population structure revealed through phenotypic characters or genes under selection with that shown by neutral genetic markers can provide insight into the relative roles of drift and selection in driving genetic divergence among populations (Merilä & Crnokrak 2001). However, comparisons between different marker systems used to make inferences about local adaptation require that markers are comparable and unbiased. One concern is that the virulence phenotypes, although under direct genetic control may underestimate the true level of divergence among populations relative to AFLP markers, particularly if either homoplasy or epistasis are common. Furthermore, the use of a standard set of host differential lines may also not provide a complete estimate of the diversity of pathogen populations when infecting their local hosts. Despite these caveats, our AFLP estimates of genetic structure...
provide a useful baseline against which to compare the structure of variation in phenotypic virulence characters, and Avr gene allele frequencies.

Most studies comparing phenotypic traits (Merila & Crnokrak 2001; McKay & Latta 2002) or genes under selection (Conway et al. 2001; Abdel-Muhsin et al. 2003; Anderson et al. 2005) with neutral markers demonstrate higher levels of differentiation among populations for phenotypic traits than for neutral marker genes. Such findings are generally interpreted as evidence for local selection driving the divergence of specific traits or genes. In contrast, in the present study, data from two Avr loci and for virulence phenotypes showed levels of spatial genetic structure either significantly less, or similar to, structure determined by the AFLP markers. Both AvrP123, (\(\theta^B_{AA} = 0.05, \theta^B_{AB} = 0.11\)) and AvrP4 (\(\theta^B_{AA} = 0.32, \theta^B_{AB} = 0.19\)) are significantly differentiated among populations, however, with the exception of estimates of \(\theta^B\) for AvrP4 for lineage AA, these values are significantly lower than estimates for the AFLP markers. Similarly, global R statistics describing divergence among populations for AFLP markers and virulence phenotypes indicate that separation among populations was more than twice as strong for AFLP than for virulence markers.

The strong population genetic structure in the AFLP data compared to the Avr genes and virulence phenotypes could be a consequence of selection for particular virulence genes or phenotypes acting to constrain among-population divergence in virulence structure (i.e. selection for the same virulence phenotypes in different populations). Thus, it is possible that selection imposed by host resistance structure is preventing population divergence beyond that seen for AFLP markers. Given the potential for very large annual fluctuations in demographic population size in M. lini, such a scenario is not implausible. However, even where selection is acting to limit divergence among populations, drift may also have the potential to influence the pathogenic identity of local populations, particularly where individual virulence phenotypes of Avr gene alleles might be under only weak selection, or selectively neutral in local host populations (Salathé et al. 2005).
To evaluate the relative influence of drift versus local selection in generating divergence among populations, pairwise patterns of divergence among populations were examined for the different genetic and phenotypic characters. In the sexual lineage AA, no significant relationships among populations were found for the degree of divergence as determined by any of the various genetic and phenotypic characters. These results demonstrate the potential for different evolutionary forces to act on these markers independently, and suggest that local patterns of selection contribute to phenotypic and Avr gene divergence among lineage AA populations. In contrast, for lineage AB, highly significant correlations were found between all phenotypes and genotypes at the population level. However, comparisons among populations for lineage AB are almost certainly complicated by a lack of sexual reproduction. Given that very strong linkages among all genetic elements in lineage AB are likely, unravelling the relative roles of drift and selection is highly problematic.

4.4.4 Conclusions

The pattern of variation in genetic and pathogenic structure of *M. lini* provides strong evidence that geographic structure across both regional and local scales is an important component in the generation and maintenance of virulence diversity in this system. The spatial distribution of the variation is consistent with evolutionary dynamics acting to structure diversity at two distinct spatial scales. At the regional scale, patterns of environmental heterogeneity interact with host and pathogen life-history traits to create two broad areas (mts vs plains) with distinct genetic and demographic dynamics. Importantly, correlated patterns of pathogen virulence diversity and host resistance diversity (Burdon *et al.* 1999) indicate contrasting rates of coevolutionary response among the two regions, suggesting the possibility of a scenario where distinct coevolutionary hotspots and coldspots (Thompson 2005) are generated through corresponding modes of host and pathogen reproduction. Within these regions local metapopulation processes act to structure the distribution of variation, driven by local patterns of selection and neutral genetic drift. A major goal for future studies will be to quantify the influence of these patterns on disease epidemiology, host mating system and host resistance.
4.5 Literature Cited


Chapter 4: Population structure and diversity


Chapter 4: Population structure and diversity
CHAPTER 5

Life-history determines genetic structure and evolutionary potential of host-parasite interactions
5.1 Introduction

Population genetic studies can provide insight into parasite evolutionary histories (Wilson et al. 2005, Miura et al. 2006, Munkacsi et al. 2007), as well as identifying causal factors contributing to disease dynamics and distribution (Thrall and Burdon 1999, Grenfell et al. 2004). Hence, accurately interpreting measures of genetic variation and its distribution within host-parasite systems is central to many applied and basic issues relating to human, plant and animal populations. These include the emergence and spread of new diseases (Wilson et al. 2005, Fargette et al. 2006, Friesen et al. 2006, Miura et al. 2006), effects of infection on host mortality and reproduction (Baumler et al. 2000), assessing risks posed by invasive parasites (Altizer et al. 2003, Desprez-Loustau et al. 2007), and predicting the evolutionary response of parasite populations to new host resistance genes or vaccines (Anderson 1998, McDonald and Linde 2002).

Parasites are a heterogeneous group of organisms that show a remarkable diversity of transmission modes, life-history strategies and spatial structures. Across the spectrum of plant and animal parasites, examples include sexually transmitted species where infection causes host sterility (e.g. anther smuts, gonorrhoea), parasites with complex life-cycles requiring multiple hosts (e.g. rust fungi, digenean trematodes), soil parasites that quickly kill their hosts (damping-off diseases, anthrax), and aerially-dispersed species that individually have only limited effects on their hosts (e.g. foliar plant pathogens, common cold). Many species infect hosts opportunistically or can attack multiple hosts, while others are more specialised, relying on living tissue for survival. Host species are similarly heterogeneous and differ in key traits such as spatial structure (population size and distribution), longevity and resistance diversity. However, despite the importance of such traits for determining disease incidence, prevalence and severity (Burdon 1992, Lockhart et al. 1996, Thrall and Burdon 2004, Carlsson-Graner and Thrall 2006), the causal relationships between spatial structure, life-history and the evolutionary dynamics of parasite populations have received little attention.

Here, we highlight the idea that variability in key host and parasite life-history traits distinctively influences disease epidemiology, genetic variation and underlying
evolutionary dynamics within populations of parasitic organisms (Box 1).
Importantly, many of the underlying assumptions of classical population genetics
theory (e.g. infinite population size, marker neutrality, random mixing, and free
genetic recombination) on which analyses and interpretations of genetic variation in
parasite populations are based, are frequently violated as a result of strong
demographic fluctuations and variable selection patterns in both hosts and parasites
(Thrall and Burdon 2003, Huyse et al. 2005). Drawing on these considerations we
derive a series of inferences regarding the likely impact of interacting host-parasite
life-history traits on parasite population dynamics and genetics.

5.2 Impacts of parasite life-history on population demography and genetic
structure

Parasite species exhibit a range of life-history strategies that affect disease dynamics
and epidemiology and, through this, the genetic composition and spatial structure of
their populations. The examples given below, and summarised in Table 1, illustrate
the potential for particular traits to strongly influence the genetic structure and
evolutionary trajectory of disease-causing organisms.

5.2.1 Host specificity

Parasites vary in the extent of host specialisation, as well as overall dependence on a
parasitic mode of life (Woolhouse et al. 2001). One expectation is that the
demographic and evolutionary dynamics of parasites with broad host ranges will
generally differ from those that can only infect single host species. For example,
parasites with narrow host ranges are more likely to be locally adapted to their hosts
than generalist species (Lajeunesse and Forbes 2002). In addition, obligate parasites
specialised to single host species are more likely to experience frequent local
extinction and recolonisation events than generalists, particularly in small and
fragmented wild host populations (Ericson et al. 1999, Thrall et al. 2001). Such
among-population processes might promote loss of genetic diversity within parasite
populations and generate among-population genetic differences (e.g. through genetic
drift) as has been observed for the biotrophic rust pathogen Melampsora lini infecting
wild Australian flax Linum marginale, which is typified by ‘boom-and-bust’ epidemic
dynamics (Burdon et al. 1999, chapter 4). In contrast, pathogens able to infect multiple hosts or survive in the absence of a host are unlikely to regularly experience such extreme levels of population stochasticity, and thus should maintain comparatively higher levels of within-population genetic variation. For example, the aerially dispersed wheat pathogen *Mycosphaerella graminicola* (wheat leaf blotch) survives saprophytically between epidemics on plant debris. Populations are genetically stable over time (Zhan et al. 2001), and worldwide studies indicate remarkably low genetic differentiation among populations with more than 90% of variation distributed within populations (Zhan et al. 2003).

**Inference:** *The degree to which pathogen species maintain stable population sizes via free-living stages, or the ability to infect multiple host species, strongly influences the effective size, variability and genetic structure of their populations.*

5.2.2 Mode of reproduction

Parasite reproductive mode is a complex trait that reflects the interplay between environmental, genetic, and demographic factors (Heitman 2006). While sexual reproduction is common, many parasite species are either clonal or primarily inbreeding; others exhibit both clonal and sexual reproductive modes that vary in time and space. In some bacteria and fungi, horizontal exchange of genetic information (both within and among species) also occurs (Friesen et al. 2006). Thus, patterns of reproduction in many species greatly depart from expectations under classical population genetic models (Prugnolle et al. 2005). Of perhaps most obvious effect, variation in levels of sexual recombination should strongly influence patterns of genetic variation, with populations that undergo sexual recombination generally exhibiting higher levels of genotypic diversity than populations that are exclusively asexual (Fox et al. 1996, Liu et al. 1996, Huyse et al. 2005). Studies investigating the genetic structure of sexual and asexual populations of the wheat stem rust *Puccinia graminis* f.sp. *tritici* illustrate some of these points. In areas where the alternate host common barberry *Berberis vulgaris* is present, sexual reproduction is common and pathogen populations harbour comparatively high levels of genotypic and phenotypic diversity for infectivity. In contrast, where barberry is absent, the rust reproduces only
asexually, and populations contain comparatively low pathogenic diversity (Burdon and Roelfs 1985).

Patterns of disease incidence and prevalence can also strongly influence mating patterns and thus the genetic structure of pathogen populations. For example, the malarial parasite *Plasmodium falciparum* has an obligate sexual phase that occurs in the mosquito vector prior to transmission. In regions where disease incidence is high, populations maintain high levels of genotypic diversity and low genetic differentiation among locations. In these areas, high disease incidence means the vector is frequently infected by multiple *Plasmodium* genotypes, which then mate and recombine before transmission to a new host. In contrast, parasites in regions with comparatively low disease incidence have a lower probability of encountering novel genotypes in the vectors, leading to frequent self-fertilization, low genetic diversity, and high variation between geographic locations (Anderson *et al.* 2000).

**Inference:** *Parasite reproductive mode is a variable and complex trait that strongly influences the genetic structure and evolutionary potential of populations. Clonal reproduction and inbreeding are generally expected to decrease genotypic variation within populations.*

### 5.2.3 Transmission and dispersal

Understanding how parasites move within and between host populations is important for correctly interpreting patterns of genetic diversity. Disease dispersal gradients vary with mode of dispersal, environmental conditions and vector behaviour. For example, soil-borne and splash dispersed species generally show steep declines in dispersal with distance, while wind-borne diseases show much flatter curves (Gregory 1968). Host-dispersed and vector-transmitted species might be expected to lie somewhere between these extremes, although in these cases, host and vector behaviour (e.g. in relation to resource distribution) will largely determine the precise shape of parasite dispersal gradients.

Between-host transmission is primarily horizontal for the majority of parasites, with spread occurring between related and unrelated host individuals via a variety of
mechanisms (e.g. direct contact, vector transmission or aerial dispersal). In some instances, transmission can be vertical, with the pathogen passing directly from parent to offspring without an intervening free dispersal stage. Differences in transmission can significantly affect the persistence of disease in populations, as well as parasite genetic diversity and population structure. Thus, vertical transmission of grass endophytes typically results in populations that maintain high and consistent levels of infection through time (Clay 1996). Stable population sizes associated with vertically transmitted species can therefore be expected to promote the retention of variation within populations.

For vector dispersed species, the extent of gene flow is primarily determined by vector behaviour. Thus, strong local spatial genetic structure in the anther smut *Microbotryum violaceum* is associated with locally restricted insect dispersal (Giraud 2004). Host preferences shown by vectors might also play a role in sympatric pathogen race formation by promoting reproductive isolation among host-specific strains (Van Putten *et al.* 2005). Similarly, for parasites of animal species, patterns of host dispersal have the potential to strongly influence parasite gene-flow (Blouin *et al.* 1995). For example, the dispersal ability of intermediate hosts has been shown to have a strong effect on the genetic structure of different species of trematode parasites of salmonid fishes (Criscione and Blouin 2004).

In contrast, patterns of dispersal in wind-dispersed species are likely to be highly stochastic and dependent on prevailing environmental conditions. Long distance aerial dispersal has been shown to result in significant gene flow among populations at both continental and global scales (Brown and Hovmoller 2002). However, for specialist pathogens with a high degree of host specificity, effective dispersal at any scale can only occur if there is a susceptible host in the new location. For that reason, long-distance dispersal events are likely to be more evident for pathogens that infect widely dispersed species (e.g. agricultural crops, livestock). Occasionally, such events can be detected or reliably inferred. For example, in 1969, two previously undetected genotypes of wheat stem rust (*Puccinia graminis* f.sp. *tritici*) dispersed to Australia from southern Africa in the jet-steam (Watson 1980). While such events might be
relatively rare, the result will be founder populations that have a lower genetic diversity than the source population because of genetic drift.

The spatial scale of parasite dispersal is an important determinant of extinction-recolonisation dynamics, particularly in wild host metapopulations, and thus influences regional patterns of disease occurrence and persistence. In turn, spatial variation in disease incidence and severity affects the intensity of selection on parasite infectivity and aggressiveness, as well as host resistance (Thrall and Burdon 1999, 2002). Furthermore, parasites that experience high levels of dispersal (and thus high gene flow) should have higher within-population genetic diversity than those with more limited dispersal. In addition, high gene flow tends to counteract the effects of genetic drift and homogenise adjacent populations, thereby increasing the spatial area encompassed by a deme. However, the outcome of host-parasite coevolution means that frequent dispersal among populations might not always result in a homogenous population structure. Theoretical studies indicate that parasites are more likely to be locally adapted to their hosts when their among-population migration rates exceed those of their hosts. The reverse is expected when parasites migrate less than their hosts (Gandon et al. 1996).

**Inference:** Pathogen dispersal is a critical factor determining disease dynamics and persistence. Pathogens infecting sessile hosts, or lacking specialisations promoting long distance dispersal, should show stronger patterns of spatial genetic structure and isolation by distance.

### 5.2.4 Life-cycle complexity and epidemiology

Parasitic organisms display an array of life-histories and infection strategies (Box 3), ranging from species that complete their life-cycle on a single host, through species with alternate parasitic and free-living phases, to species that have multiple stages, each on a different host species. For parasites with obligate multi-host life-cycles, the selective forces experienced by different genotypes are likely to vary for different stages, given the requirement for establishing on potentially unrelated host species. Such trade-offs between fitness components might be expected to promote the maintenance of genetic polymorphisms due to disruptive or fluctuating selection.

Chapter 5: Life history
favouring different alleles in different obligatory hosts (Gower and Webster 2004). The availability of alternate hosts has further been proposed to promote outcrossing and transmission (Rauch et al. 2005) and might also increase population stability, thus reducing the severity of bottlenecks and decreasing genetic drift.

Similarly, for host-pathogen interactions typified by epidemic dynamics, pathogen traits under selection will vary at different stages in the epidemiological cycle. During the transmission and establishment phase, selection for the ability to encounter and infect a broad range of hosts will be high; later as population growth enters an exponential phase, fecundity can become increasingly important; while later still, as populations collapse there is likely to be a strong shift towards traits favouring off-season survival (Burdon 1992). For example, end-of-season population reductions can be massive for specialist pathogens of plants that have no reliable resting spore stages or alternate hosts (e.g. some biotrophic rusts (Roelfs 1982). In contrast, survival might be less problematic for generalist species, although evolutionary constraints imposed by differential patterns of selection between parasitic and free-living phases of the disease cycle might act to reduce diversity (Abang et al. 2006). Particularly for systemic parasites, trade-offs can occur such that greater dispersal ability might allow genotypes to persist even when they are inferior within-host competitors (Amarasekare and Nisbet 2001). A key point is that inferences regarding pathogen evolutionary history must account for within and among-population variation in within-season disease dynamics, severity and incidence.

**Inference:** Increasing life-cycle complexity is likely to generate higher levels of genetic variation in response to increased demographic stability and the potential disruptive selection pressures being experienced by different infective stages.

**Inference:** Variation in selection intensity during the epidemiological cycle provides opportunities for diversifying selection, competition and genetic drift to influence the diversity and structure of pathogen populations.

5.2.5 Parasite infection strategies

Parasite species vary in infection strategies that, in combination with different suites of host life-histories, broadly determine opportunities for disease invasion, spread and
persistence, and thus the rate and direction of coevolution. The ways in which parasites impact host fitness can be broadly classed into those that cause rapid host death (‘killers’), those that directly attack host reproductive organs (‘castrators’), and those that reduce overall host fitness (‘debilitators’) (Lafferty and Kuris 2002).

Results from a recent metapopulation model exploring the dynamics of pathogens that decrease host fertility versus ones that increase mortality support the contention that these groups might differ broadly in persistence, the degree to which they reduce host population size, and the strength of selection on host resistance (Thrall and Burdon 2004). In particular, the evolution of resistance was more rapid for pathogens affecting host mortality than those reducing fecundity. Interestingly, while increasing pathogen effects on fecundity resulted in a monotonic decrease in the number of patches occupied by the host, the impact of increasing mortality was nonlinear, with the greatest reductions in host population size at intermediate virulence levels; when virulence was high, the parasite temporarily reduced host occupancy of patches to the point that it caused its own extinction. The magnitude of these effects is at least partly determined by host life-history (e.g. longevity) (Miller et al. 2007). Parasites also differ with regard to which parts of their hosts they colonise. Resultant variability in the durability of infection is likely to strongly influence the genetic diversity and demographic dynamics of parasite populations. For example, foliar plant pathogens frequently undergo annual epidemic cycles with population sizes fluctuating greatly, while systemic pathogens generate long-term infections of perennial hosts and maintain relatively stable population sizes (Burdon 1992).

**Inference:** Different infection strategies provide opportunities for population size, persistence, diversifying selection, competition and genetic drift to influence the genetic diversity and structure of pathogen populations.

### 5.3 Impact of host life-history and spatial structure on parasite dynamics and evolution

The obligate dependence of many parasitic organisms on their hosts for long-term survival makes the size, structure and distribution of host populations an important
determinant of the genetic structure of parasite populations. Hosts represent an inherently patchy and dynamic resource that varies spatially and temporally in the time available for infection, levels and types of resistance (e.g. quantitative versus genes of major effect), and population size and persistence. Furthermore, host species differ in life-history traits such as longevity, mating system and phenology which also influence parasite population dynamics. Despite this, there has been little consideration of how such heterogeneities affect the genetic structure and evolutionary trajectories of parasite populations.

5.3.1 Spatial structure
The extent to which the spatial structure of parasite populations mirrors patterns seen in host populations depends on the degree to which pathogens are obligately dependent on that host. For animal parasites, estimates of spatial structure must also take into account host dispersal (McCoy et al. 2003) (Box 2) as well as heterogeneities in host social behaviour and group structures. Generally, increasing among-population isolation should reduce genetic variability within parasite populations and promote higher levels of genetic differentiation among demes (Ellstrand and Elam 1993). Population subdivision also increases parasite population vulnerability to environmental stochasticity and the possibility of local extinction of particular genetic variants, or even whole demes.

Population size and the degree of population subdivision are both likely to differ significantly between parasites infecting human or domesticated species, and those found on wild hosts. For example, agricultural crops are typically planted in large genetically uniform stands, which massively amplify available host resources for suitable pathogen races, thereby increasing effective pathogen population size. In contrast, wild plant populations are often demographically small and at least partially isolated, thus decreasing the chance of migration between pathogen populations, and increasing the propensity for drift, divergent selection and genetic differentiation among populations. The possibility of near, or total, extinction of pathogen populations also depends on host numbers. For example, an 11-year survey of >130 populations of the rust pathogen Triphragmium ulmariae infecting its wild host meadowsweet Filipendula ulmaria found that extinction and recolonisation events
were common, and that disease incidence was positively correlated with host population size (Smith et al. 2003).

Ecological differences between the plant host species red campion *Silene dioica* and white campion *S. latifolia* provide an interesting comparison of the effects of host life-history on populations of the anther smut *Microbotryum violaceum* in Europe (Bucheli et al. 2001). *S. latifolia* is found in ruderal habitats associated with frequent population extinction and recolonisation. In contrast, *S. dioica* is found in more undisturbed natural habitats, and has more stable population dynamics. As might be predicted based on these differences, *M. violaceum* collected from *S. latifolia* showed less microsatellite variation and higher differentiation among populations than that from *S. dioica*. Another key factor is host population connectivity; more continuous host populations favour the evolution of resistance as disease spreads globally (Carlsson-Graner and Thrall 2002) which could in turn influence the evolution of pathogen virulence.

**Inference:** Host population distribution affects parasite population dynamics and among-population movement. The metapopulation structure typical of many wild hosts results in complex heterogeneous mosaics of local populations, dictating that pathogen population sizes are limited, near-extinction events are common, and drift and gene flow are influential.

5.3.2 Host longevity and phenology

Theoretical models indicate that long-term persistence of horizontally transmitted parasites exploiting single host species within populations increases with lower intrinsic host mortality (Anderson and May 1982). Thus, if hosts are short-lived, pathogen populations are more likely to experience regular extinction and colonization dynamics. In contrast, if hosts are long-lived and provide a perennial resource, then pathogen populations are likely to be more stable. Empirical support for this prediction is found in the interaction between *M. violaceum* and plant host species in the Caryophyllaceae which vary in longevity, mating system and vector identity (Thrall et al. 1993). Variation in the timing of different host life-history stages (or ephemerality of the tissues being attacked) can also alter host availability.

Chapter 5: Life history
and hence disease epidemiology within populations. For example, because of seasonal
differences in leaf availability, foliar pathogens of perennial deciduous trees
experience demographic dynamics more akin to pathogens of short-lived host species
than systemic pathogens of long-lived species (Roslin et al. 2007).

The timing of host availability is also a strong selective force at multiple levels of
spatial organisation. Thus, temporal variation in the timing of leaf production and
abscission among individual trees promotes reproductive isolation and adaptation of
specialist insect herbivores (Mopper et al. 2000). Experimental studies of
entomopathogenic nematodes indicate that variation in host availability and
transmission opportunities among host populations might play a role in the evolution
of infection strategies, as well as maintaining overall diversity (Crossan et al. 2007),
while differences in flowering phenology of the congeneric hosts S. latifolia and S.
dioica promote strong host-related genetic differentiation in associated populations of
the anther smut M. violaceum (van Putten et al. 2007).

**Inference:** Host longevity and availability affects the predictability and timing of
substrate availability for pathogens and, through this, the size of populations,
adaptation to local conditions, and the potential for gene flow among populations.

5.3.3 Host resistance

Hosts resist attack by parasites and pathogen via a diverse array of mechanisms that
have the potential to influence pathogen population genetic structure. However, broad
relationships between host resistance mechanisms and the genetic structure of
pathogen populations remain largely unconsidered. For example, in vertebrate hosts,
differential exposure to either acquired or innate immune mechanisms should have the
potential to strongly influence the genetic structure and evolutionary dynamics of
pathogen populations (e.g. Futse et al. 2008). Similarly, pathogen evolutionary
dynamics can also be strongly influenced by qualitative vs. quantitative genetic
control over innate resistance in plant and invertebrate hosts (Gandon and Michalakis
2000). However, to date, the general influence of such variation on the strength and
direction of selection on pathogen genes and the genetic consequences for populations
is largely unknown.
Currently, the clearest example of how differences in mechanisms of host resistance can influence the genetic structure of pathogen and parasite systems is in gene for gene interactions between plant hosts and their microbial pathogens. Under the current paradigm, there are essentially two models to explain how hosts and pathogens interact (via gene-for-gene dynamics) at the molecular level. Firstly the products of host resistance genes can interact directly with the pathogen molecules they recognize (known as effector or avirulence proteins) to trigger resistance (e.g. flax and flax rust; Dodds et al. 2006). Secondly hosts can detect certain effector proteins indirectly by responding to changes induced in host target proteins. For example, the \textit{RPS2} and \textit{RPM1} resistance proteins in \textit{Arabidopsis} apparently recognize corresponding \textit{Pseudomonas syringae} effector gene products by detecting changes induced in the intermediary host protein \textit{RIN4} by the pathogen products (Mackey et al. 2002, Axtell and Staskawicz 2003). It has been suggested that these differences are likely to lead to qualitatively different outcomes in terms of the diversity of host resistance and pathogen infectivity (Van der Hoorn et al. 2002). Thus, indirect recognition will lead to simple, binary, balanced polymorphisms for host resistance and pathogen infectivity, such as observed for the \textit{RPM1} and \textit{RPS2} loci in \textit{Arabidopsis} (Stahl et al. 1999, Mauricio et al. 2003). In contrast, direct interactions might favor the classical ‘arms race’ model of host-pathogen coevolution, whereby continual changes in pathogen \textit{Avr} genes are matched by changes in host \textit{R} genes. This could lead to high levels of phenotypic, allelic and nucleotide diversity, such as are observed at interacting host and pathogen loci in the flax/rust system (Dodds et al. 2006).

Hosts can also vary widely in the diversity of resistance genotypes within and among populations. The potential for variation in host resistance to influence the generation and maintenance of genetic polymorphisms for pathogen virulence is well established both theoretically (e.g. Hamilton 1980, Anderson and May 1982) and empirically. For example, high levels of virulence diversity within and between populations (Jarosz and Burdon 1991), local adaptation of pathogen and host populations (Thrall et al. 2002), correlations between levels of host resistance and pathogen virulence in local populations, and trade-offs between spore production and virulence (Thrall and
Burdon 2003) all implicate selection by *L. marginale* as an important source of variation in populations of *M. lini*. In contrast, genetically homogeneous crops grown over large areas impose strong directional selection on local pathogen populations, resulting in selective sweeps favouring particular genotypes and lower overall virulence diversity in pathogen populations (Wichmann et al. 2005). For pathogens that either infect multiple hosts or are vector-transmitted, population structure can reflect selection pressure from different host defences. For example, allele and genotype frequencies in populations of the cattle intracellular parasite *Theileria parva* are markedly changed after passage through both the tick vector *Rhipicephalus appendiculatus* and the bovine host (Katzer et al. 2007). Similarly, within species, male and female hosts can differ in infection patterns, and thus harbour parasite populations with different levels of genetic diversity (Caillaud et al. 2006).

**Inference:** Heterogeneity for host resistance can result in complex patterns of disruptive selection being imposed on pathogen populations, and through this, dramatically influence levels of diversity in host-pathogen systems.

### 5.4 Genetic expectations for host-pathogen interactions

Conclusions regarding parasite evolutionary history and potential that are based on univariate approaches (e.g. characterisation of patterns at neutral marker loci) are likely to be incomplete. Empirical examples presented in this essay support the argument that host and pathogen life-history features partly determine key variables like connectivity and stochasticity (particularly as it relates to the predictability of encounter between individual host and parasite genotypes, and the overall degree to which local dynamics are endemic versus epidemic) that affect disease patterns, genetic variation and genetic structure (Table 5.1). For example, demographic stability is predicted to decline with reductions in host and pathogen population sizes, decreasing host range, shorter host lifespans, reduced ability of pathogens to survive in the absence of living hosts, and smaller spatial scales of dispersal. Among-population connectivity is a function of both host and pathogen dispersal scale. Based on the assumption that neutral loci are generally more sensitive to changes in either stochasticity or connectivity than loci under selection, it is possible to make general
predictions about how changes in connectivity and stochasticity might influence genetic patterns for neutral versus selected loci (Fig. 5.1). As implied by Figure 5.1, absolute levels of variation, as well as relative variability in neutral and selected loci are subject to the influence of suites of host and parasite traits which also determine patterns of disease persistence and among-population movement. An important point is that low genetic diversity per se does not necessarily imply that a parasite is of recent origin or newly introduced.

**Figure 5.1.** General genetic expectations for neutral versus selected loci in pathogen populations. a) Overall genetic diversity in relation to host and pathogen population stability. b) Population genetic structure (e.g. as measured by standard statistics such as $F_{st}$) in relation to among-population connectivity. Expectations for changes in neutral genetic variation are based on metapopulation models exploring the impact of variation in rates of population turnover (Pannell and Charlesworth 1999).
These predictions assume that genes under selection for variation (e.g. avirulence loci) will be buffered to some extent, particularly in natural coevolutionary interactions where pathogens are responding to spatial heterogeneity in host resistance structure. Thus, levels of diversity in host resistance structure will partly determine the degree to which genes under selection respond to changes in either stochasticity or spatial structure. Clearly the precise shapes of the relationships shown in Figure 5.1 could take many forms, particularly given that stochasticity and connectivity are not completely independent of each other. Moreover, one could also imagine situations where these relationships would not hold. Thus, if local selection were particularly strong, genetic structure at selected loci could potentially be stronger than at neutral loci. Characterising host resistance structure is therefore crucial to interpreting patterns of variation in parasite populations. Despite these caveats, this analysis provides a starting point for exploring inter-relationships between life history, spatial structure and the epidemiological and genetic dynamics of host-parasite systems. Furthermore, it strongly highlights the value of comparative studies across multiple populations or species as well as modelling studies that integrate demographic and genetic dynamics, particularly where key life-history features can be varied. Host-parasite systems where genes relevant to the interaction can be characterised in addition to neutral molecular markers are likely to be especially informative.

5.5 Inferring demographic and evolutionary dynamics

Host-parasite systems represent suites of interacting life-histories that collectively determine disease epidemiology, and therefore patterns of genetic variation and evolutionary trajectories of parasite populations. In particular, parasite genetic diversity might be affected in an extreme fashion when life-history parameters interact additively (Table 5.1). For example, high levels of demographic and genetic stochasticity resulting from a combination of traits including host specialisation, metapopulation dynamics (with associated extinction-recolonisation dynamics) and asexual reproduction should significantly increase the fixation rate of neutral and nearly neutral mutations and reduce the effective size of parasite populations (Huyse et al. 2005) relative to more demographically stable situations. Moreover, selective sweeps leading to the fixation of favourable alleles might further lead to reductions in
neutral genetic diversity across parasite genomes (Kaplan et al. 1989, Wichmann et al. 2005).

Thus, if relevant biological parameters are not considered, expectations for population genetic patterns that are based on particular demographic or evolutionary episodes (e.g. a recent host shift) are likely to be inaccurate, and might lead to a failure to detect the predicted pattern or to false detection of a non-existent pattern (Morrison and Hoglund 2005). For example, some wild species, even ones with nearly cosmopolitan distributions, possess little variation at neutral marker loci (Sivasundar and Hey 2003, Rydholm et al. 2006). Thus, low genetic variation per se does not necessarily indicate recent bottlenecks caused by invasion or population declines, and instead might reflect various aspects of the underlying biology of the species concerned. Conclusions regarding pathogen evolutionary potential and history must therefore explicitly consider the influence of both host and parasite life-history on patterns of genetic variation within contemporary parasite populations.

Table 5.1. Some factors associated with host and pathogen life-history and population structure that are likely to influence the genetic structure and effective size of populations.

<table>
<thead>
<tr>
<th>Factors that generally increase effective population size</th>
<th>Factors that generally decrease effective population size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Host exploitation and specificity</strong></td>
<td></td>
</tr>
<tr>
<td>Opportunistic and/or generalist species, multiple host species</td>
<td>Specialised parasite, single host species</td>
</tr>
<tr>
<td><strong>Mode of pathogen reproduction</strong></td>
<td></td>
</tr>
<tr>
<td>Sexual</td>
<td>Clonal or inbreeding</td>
</tr>
<tr>
<td><strong>Pathogen dispersal</strong></td>
<td></td>
</tr>
<tr>
<td>Long distance dispersal</td>
<td>Restricted, local dispersal</td>
</tr>
<tr>
<td><strong>Environmental stochasticity</strong></td>
<td></td>
</tr>
<tr>
<td>Stable environment and host population dynamics</td>
<td>Frequent population extinction and recolonisation, short lived hosts</td>
</tr>
<tr>
<td><strong>Host longevity; ephemerality of tissues attacked</strong></td>
<td></td>
</tr>
<tr>
<td>Perennial or long lived host</td>
<td>Annual or ephemeral hosts</td>
</tr>
<tr>
<td><strong>Host population size and structure</strong></td>
<td></td>
</tr>
<tr>
<td>Large, inter-connected host populations</td>
<td>Small, fragmented host populations</td>
</tr>
<tr>
<td><strong>Epidemiological dynamics</strong></td>
<td></td>
</tr>
<tr>
<td>Endemic, systemic</td>
<td>Epidemic, boom &amp; bust</td>
</tr>
</tbody>
</table>

Chapter 5: Life history
5.6 Conclusions

Understanding the evolutionary drivers of disease outbreak and emergence is of clear importance, especially given the health, economic and ecological costs associated with infectious diseases. Here, we demonstrate that variation in key pathogen life-history features drives epidemiological and genetic dynamics within and between pathogen populations. Similarly, we propose that pathogen populations are likely to respond to a range of variable host life-history traits and population structures (Table 1). This argues for better integration of information on host and pathogen life-history into theoretical and empirical studies of disease. Of particular value will be focused comparative studies of the genetic structure of pathogen and parasite species that differ in key features of their life-history, life-cycle and associated traits of the host, using similar sampling designs and genetic markers. In addition, we highlight the importance of combining demographic and genetic studies at the population level. As examples from both natural and anthropogenically modified systems accumulate, and better genomic tools become available, it should be possible to gain a clearer, more mechanistic understanding of the variability and adaptability of pathogen populations in relation to life history and population structure. Quantifying the consequences of these interactions should lead to predictions that can advance our understanding of host-parasite interactions as well as contribute to the development of a broad conceptual framework for understanding the role of life-history in parasite evolution.
5.7 Literature Cited


van Putten, W. F., J. A. Elzinga, and A. Biere. 2007. Host fidelity of the pollinator guilds of *Silene dioica* and *Silene latifolia*: Possible consequences for


Zhan, J., C. C. Mundt, and B. A. McDonald. 2001. Using restriction fragment length polymorphisms to assess temporal variation and estimate the number of ascospores that initiate epidemics in field populations of Mycosphaerella graminicola. Phytopathology 91:1011-1017.

CHAPTER 6

Synthesis and Conclusions
6.1. Overview: *M. lini* in the geographic mosaic

A major recent advance for understanding the coevolution of parasites and their hosts has arisen from observations that the dynamics and evolutionary outcomes of species interactions can vary across landscapes. This geographic heterogeneity forms the basis of the geographic mosaic theory of coevolution (Thompson 1994; Thompson 2005). There are three central components to the geographic mosaic theory of coevolution. First, the intensity and direction of selection between interacting species varies across landscapes (i.e., geographic selection mosaics). Second, the rate and intensity of coevolution will vary across locations (i.e., coevolutionary coldspots and hotspots). Third, gene flow, genetic drift and other stochastic and non-equilibrium factors continually alter the spatial distribution of genetic variation for potentially coevolving alleles and traits.

This work provides one of the clearest empirical examples to date of the potential for spatial and environmental heterogeneity to drive coevolutionary diversification in host-pathogen interactions. Across broad geographic scales, significant genetic, pathogenic and geographic differences between two lineages (AA and AB) of *M. lini* in Australia are revealed. In particular, it is shown that lineage AA and AB differ in a range of key reproductive, life-history, and epidemiological traits, that when combined with distinct patterns of environmental heterogeneity and host phenology, generate striking regional divergence in the distribution, diversity and identity of pathogenic variation. Importantly, the key processes involved in the generation and maintenance of these among-population and regional differences include gene flow and subsequent hybridisation in one lineage (AB). Within the resultant geographic and genetic framework, critical reproductive and life-history differences between the two lineages, in conjunction with local selective and non-equilibrium dynamics, strongly influence evolutionary and genetic dynamics at local and regional scales. In addition, strong diversifying selection, resulting from interactions between *M. lini* and *L. marginale*, is generating high levels of genetic variation at genes underlying pathogen virulence, and it is proposed that geographically divergent patterns of natural selection across different host and pathogen populations are responsible for maintaining this variation.
Overall, these patterns of differentiation are consistent with a mosaic of different selective scenarios operating across the range of the interaction between *M. lini* and its host, *L. marginale*. At the species level, frequent polymorphisms and strong statistical evidence for positive diversifying selection at gene loci underlying pathogen virulence confirm the importance of coevolutionary interactions in generating pathogenic diversity. At regional scales, strong genetic and pathogenic structure are uncovered, and it is proposed that correlated patterns of pathogen virulence diversity and host resistance diversity are indicative of contrasting rates of coevolutionary response. This suggests a scenario where distinct coevolutionary hotspots and coldspots are generated at regional scales through corresponding differences in the mode of host and pathogen reproduction. Within the resultant genetic and geographic framework, local patterns of directional and balancing selection, combined with gene flow, neutral genetic drift and other non-equilibrium dynamics interact to create a shifting, inter-connected mosaic of coevolutionary dynamics.

6.2. The ecological and evolutionary determinants of virulence polymorphisms in *M. lini*

In this thesis, molecular, experimental and field approaches have been combined to gain important and novel insight into the ecological and evolutionary drivers of genetic variation in pathogen virulence. Below, the main results with respect to the general focus of this thesis are discussed: *How are genetic polymorphisms for pathogen virulence generated and maintained across geographic landscapes?*

Frequent phenotypic polymorphisms for virulence exist within and among natural populations of *M. lini* (Burdon and Jarosz, 1991; Jarosz and Burdon, 1991; Burdon, Thrall and Brown, 1999; Thrall and Burdon, 2003; chapter 4). In chapter 3, it is demonstrated that frequent allelic polymorphisms also exist at *Avr* gene loci that (among other genes) determine pathogen virulence phenotypes in the interaction with *L. marginale*. Statistical analysis of patterns of nucleotide polymorphism shows strong evidence for positive selection driving this allelic diversification, indicating

Chapter 6: Synthesis and conclusions
that ongoing coevolutionary dynamics between *M. lini* and *L. marginale* are important for generating virulence polymorphisms. Allelic polymorphisms at *Avr* loci may potentially be maintained by selection within the broader *M. lini* population in three major ways: heterozygote advantage, frequency dependent selection, and geographically divergent selection.

Polymorphisms can be maintained in populations when heterozygotes (at any given locus) exhibit higher fitness in a population than homozygotes (i.e. overdominant selection; (Maruyama and Nei 1981). However, for pathogen species, theoretical studies suggest that selection should favour pathogen individuals that express the fewest potential elicitors of resistance in the host. Hence, for example, haploid parasite species or genotypes should hold an advantage over diploids (Nuismer and Otto 2004). By extension, unless mechanisms exist to limit gene expression, increasing heterozygosity in the dikaryotic *M. lini* should equate to an increasing probability of being recognised by any given host. However, as shown in chapter 2, highly heterozygous lineage AB isolates are significantly more virulent than the typically homozygous lineage AA isolates.

As shown subsequently in chapter 3, this finding does not in fact contradict theoretical expectations for haploid vs. diploid parasites. Instead the high levels of divergence between clade A and B alleles, a lack of variation among clade B alleles, and a potential alien origin for the B genome, suggests that *L. marginale* has not evolved the capacity to recognise clade B alleles. Such a scenario would therefore act to decrease the mean likelihood of detection of lineage AB genotypes by *L. marginale*, as potential avirulence elicitors would only be present in the ‘A’ nucleus (i.e. lineage AB isolates are functionally haploid). A second possibility may be that the pathogenicity effector function of *Avr* genes could be enhanced in the heterozygous state, providing a fitness benefit to the individual where genes for resistance are not common in the host population (chapter 3).

Many theoretical models predict that frequency dependent selection can maintain polymorphisms within pathogen populations (and in corresponding host populations), assuming that particular virulence alleles may be either costly or beneficial depending
on the level of resistance in corresponding host populations (Jayakar 1970; Antonovics and Thrall 1994; Roy and Kirchner 2000; Thrall and Burdon 2002). For the interaction between \textit{M. lini} and \textit{L. marginale}, there is no direct empirical evidence for these kinds of fluctuating genetic dynamics taking place within local populations. However, the potential for frequency dependent dynamics to maintain genetic polymorphisms within populations of \textit{M. lini} is evidenced by the existence of considerable diversity for virulence within populations (both at the phenotypic and molecular levels) \textbf{(chapter 3, 4)}. In addition, for \textit{M. lini}, the potential for costs to maintain variation in pathogen virulence has been confirmed via experimental inoculations, which demonstrated trade-offs between spore production and virulence in isolates infecting the Australian host \textit{L. marginale} (Thrall and Burdon 2003). From the host side, there is also evidence for changes in resistance structure within populations following the impact of a substantial epidemic. Large declines in overall population size were associated with a marked shift in host resistance structure such that the resistance phenotypes at high frequency in the pre-epidemic population declined significantly in the post-epidemic population (Burdon and Thompson 1995).

Where such frequency dependent dynamics exist, sexual reproduction provides a powerful means to generate locally rare combinations of genotypes. Under the Red Queen Hypothesis (RQH), sexual reproduction in hosts is seen as an adaptation against rapidly evolving parasites because it enables host organisms to evolve faster by recombination of resistance genes (Hamilton 1980; Hamilton \textit{et al.} 1990). In addition, the RQH predicts that pathogens infecting sexual hosts should be genetically more variable than those infecting asexual hosts because recombination of resistance genes will drive frequency-dependent selection on the corresponding virulence genes of pathogens (Ooi and Yahara 1999). Consistent with these predictions, results from \textbf{chapter 2 and chapter 4} demonstrate that sexual (lineage AA) and asexual (lineage AB) populations of \textit{M. lini} are strongly geographically separated. Furthermore, genotypic and pathogenic diversity were significantly higher in sexual than in asexual pathogen populations \textbf{(chapter 4)}. Interestingly, these differences appear to be associated with corresponding differences in host mating system and the resistance structure and diversity of host populations (Burdon \textit{et al.} 1999; LG Barrett; unpublished manuscript). This situation, in which clonal pathogens interact with ...
Inbreeding hosts and sexual pathogens interact with outcrossing hosts, is consistent with a coevolutionary scenario where the comparative rates of frequency dependent change in the host and pathogen are moderated through corresponding changes in the modes of host and pathogen reproduction.

In this thesis, perhaps the clearest evidence for how virulence diversity is maintained in *M. lini* has been found at broader population and regional scales. Geographically variable selection for different virulence phenotypes and genotypes almost certainly contributes to the maintenance of different phenotypes and *Avr* gene alleles in different partially isolated populations in *M. lini*. Strong patterns of genetic (*AvrP123 and AvrP4*) and phenotypic divergence among local populations (chapters 3 and 4), suggests that local patterns of selection have the potential to contribute to the maintenance of virulence diversity among local populations. These results are consistent with previous analyses of the *L. marginale-M. lini* interaction that have demonstrated strong local adaptation by *M. lini* to its host populations within a single metapopulation (Thrall et al. 2002). Previous studies have also demonstrated that at broader spatial scales, there are significant differences in the partitioning of host resistance within and among the plains and mountains regions (Burdon et al. 1999; LG Barrett unpublished manuscript). Through demonstrating corresponding regional structure in the distribution of virulence genotypes and phenotypes (chapters 2, 3 and 4), it has been confirmed that these patterns likely reflect divergent coevolutionary dynamics across broad regional and continental scales in the *L. marginale-M. lini* interaction.

Local adaptation, frequency dependence and heterozygote advantages are only likely to be part of the raw material for the overall processes by which pathogens evolve differentially across geographic landscapes. In addition, various other deterministic, non-selective and stochastic factors are likely to influence how diversity is generated and maintained within groups of interconnecting populations (Thompson 2005). For example, patterns of environmental heterogeneity, spatial population structure, gene flow, extinction/recolonisation dynamics and various non-selective factors are all likely to influence how variation is generated and maintained across different spatial scales.
The spatial structure of natural populations has important consequences for both population dynamics and evolutionary change (chapters 4 and 5). Demographically, the scale and pattern of pathogen dispersal and gene flow between local populations can impact on disease dynamics and persistence and the genetic structure of local populations (Antonovics et al. 1994; Nuismer et al. 1999; Thrall and Burdon 2002). Genetically, the level of gene flow will influence to what extent, and at what scale, local populations diverge (Latta 2004) For example, where strong patterns of local selection promote genetic divergence between populations (chapter 3 and 4), dispersal and gene flow could potentially maintain polymorphisms for virulence within populations, even in the absence of local overdominant or fluctuating selection (Levene 1953; Gillespie 1974; Hedrick 1974). Furthermore, highly stochastic demographic forces in populations of M. lini, associated with boom and bust epidemic dynamics (reviewed in chapter 5), suggest that gene flow is likely important for maintaining diversity and preventing extreme divergence in the face of frequent bottlenecks and strong potential for genetic drift (chapter 4). Importantly, long distance dispersal, phylogenetic divergence, and subsequent hybridisation are the likely processes behind the generation of lineage AB within Australia (chapter 2). The hybrid nature of lineage AB appears to directly and critically influence broad-scale coevolutionary dynamics through the modification of a range of life-history traits influencing pathogen distribution, virulence identity and diversity, and disease dynamics (chapters 2, 3 and 4).

Parasites (and the hosts that they exploit) exhibit a broad range of life-history strategies that may influence disease dynamics, persistence within populations, and population genetic structure (chapter 5). For M. lini, elements of both host and pathogen life-history are likely to interact with local environmental conditions to influence demographic and genetic dynamics across both local and broad regional scales (chapters 2, 4 and 5). For example, at local scales, host populations are generally very small (often on the order of 100’s of individuals), and go through an annual phase of senescence. This means that local pathogen populations (which are adapted to local conditions: chapters 2 and 4) experience strong annual fluctuations in populations size, resulting in extreme bottlenecks (if not complete local extinction)
followed by phases of rapid demographic increase. The genetic outcome of these kind of dynamics is that genetic variability may be strongly reduced within populations by random genetic drift (Ingvarsson 1997; Pannell and Charlesworth 1999). Thus, ongoing patterns of gene flow over long distances among \emph{M. lini} populations across Australia are likely to be important for the maintenance of local variation (as discussed previously). Within this dynamic, reproductive patterns such as the degree of inbreeding and clonal reproduction may further influence effective population sizes (\textit{chapter 4}). As discussed above, across broad spatial scales, the generation and maintenance of pathogenic diversity can be strongly influenced by interactions between host and pathogen life-history traits (which have in turn been influenced by patterns of gene flow) and environmental heterogeneity (\textit{chapter 2, 4 and 5}).

6.2.1 Conclusions

One of the central challenges for increasing our general knowledge of how pathogen virulence evolves is to understand how genetic polymorphisms are maintained across space and time. This thesis work demonstrates that the ecological and evolutionary factors involved in the generation and maintenance of virulence and aggressiveness in \emph{M. lini} are complex, and encompass multiple spatial scales. Clearly, heterozygote advantage and frequency dependent selection may contribute to the maintenance of polymorphisms within populations. However, the patterns of variation seen in \emph{M. lini} provide dramatic evidence that the maintenance of variability within \emph{M. lini} probably relies upon evolutionary divergence among geographically structured populations, at both local and broader regional scales. Within this framework, gene flow, stochastic population dynamics, host and pathogen life history, environmental variation, and environmental variation all influence the processes by which genetic variation is generated and maintained. In turn, such variation underlies the distribution and abundance of diseases across landscapes - the visible outcomes of interactions between hosts and their pathogens.


Levene, H. 1953. Genetic equilibrium when more than one ecological niche is available. Am. Nat. 87:331-333.


**APPENDIX**


**Abstract**

We developed and characterized primers for 11 variable microsatellite loci present in the genome of the flax rust, *Melampsora lini*. The microsatellite loci were identified by sequencing clones from a library of EcoRI DNA fragments enriched for four simple sequence repeat motifs (AAG, AAT, TC, and TG). All 11 primer pairs successfully amplified DNA fragments from a sample of 102 *M. lini* isolates (98 isolated from *Linum marginale* and 4 from *L. usitatissimum*), revealing a total of 32 alleles. Allelic diversity at the 11 loci ranged from 0.030 to 0.449.
Melampsora lini is an autoecious fungal pathogen and the causal agent of rust disease on a number of different plant species in the family Linaceae, including flax and linseed (Linum usitatissimum) (Lawrence 1989). Melampsora lini is a macrocyclic species capable of repeated cycles of asexual reproduction or of initiating a process of sexual recombination under appropriate environmental conditions. During the epidemic phase and in the glasshouse, M. lini exists as clonally reproducing dikaryotic urediniospores. While there is a good understanding of the ecological and coevolutionary interactions between M. lini and its Australian host, L. marginale (Thrall et al. 2002; Thrall and Burdon 2003), there is little information regarding the neutral population genetic structure of either the host or the pathogen. Ultimately to further an understanding of the underlying population genetic structure and how processes such as sexual reproduction and gene flow are influencing the co-evolutionary dynamics between M. lini and Linum spp., genetic analysis with neutral markers is essential.

Although Australian M. lini populations exhibit high levels of genetic variation for pathogenicity characters (Jarosz and Burdon 1991), a previous study using isozyme and RFLP markers revealed only limited genetic variation (Burdon and Roberts 1995). Investigating the relationships between neutral genetic variation and selected traits (i.e., virulence, aggressiveness, and resistance) in this system clearly requires genetic studies using more informative molecular markers. To progress these investigations, 11 polymorphic microsatellite loci were identified and characterized from M. lini to provide a suite of new, and potentially more informative, codominant markers.

Prior to DNA extraction, 100 mg of urediniospores were allowed to germinate in 100 uL of sterile water in a sterile 1.5 ml eppendorf tube. Total genomic DNA was extracted using a DNeasy plant mini kit (Qiagen) following standard protocols. A library enriched for microsatellite containing fragments was constructed following the FIASCO procedure of Zane et al. (2002) with some modifications. Briefly, 250 ng of genomic DNA was digested with EcoRI (New England Biolabs) and simultaneously ligated to EcoRI-adapters. An initial round of PCR amplification of adapter-ligated fragments was performed using primers specific to the EcoRI adapters. Fragments containing putative microsatellite repeats were magnetically captured using four
biotinylated SSR probes [(AAG)\textsubscript{10}, (AAT)\textsubscript{10}, (TC)\textsubscript{10}, (TG)\textsubscript{10}] and streptavidin coated magnetic particles (Promega). Enrichments for each repeat motif were performed separately. The resulting products were amplified in a second PCR amplification using the EcoRI adapter-specific primers and ligated into pGEM T-Easy (Promega) vectors following the manufacturer’s instructions. The plasmids were transformed into competent Escherichia coli JM109 cells (Promega). Plasmid DNA from putative positive clones was isolated by touching a sterile pipette tip to the colony which was then washed in 10 \(\mu\)L of 10 mM Tris by repeated pipetting to create a dilute suspension of cells. The cell suspensions were boiled for 10 minutes to lyse the cells. The cellular contents were pelleted by centrifugation, and 2 \(\mu\)L of the DNA-Tris supernatant was used as a template for PCR. The multi-cloning site was amplified using the flanking M13 priming sites in a standard PCR reaction. Genomic inserts larger than 500 bp were sequenced on an ABI 310 Genetic Analyser using the BigDye Terminator Cycle Sequencing kit (Applied Biosystems).

A total of 192 clones (48 from each enrichment motif) were sequenced. Of these, 30 (16\%) contained novel microsatellite loci with 5 or more repeats. The AAG probe was the most efficient, returning 16 of the 30 recovered loci. Based on these results, a further 96 AAG clones (excluding inserts of a size that had already been sequenced to prevent sequencing of duplicated loci) were sequenced. Of these, 30 (31\%) contained novel microsatellite loci with 5 or more repeats.

The microsatellite repeat was located at the one end of the fragment in 19 of the 60 fragments, preventing primer development. Primer pairs specific to the remaining 41 SSR-containing loci were designed using the program Oligo (Molecular Biology Insights). A M13 tag (5'–CACGACGTTGAAAACGAC) was added to the 5’ tail of the forward primer following Schuelke (2000). These primers were initially tested on an arbitrary set of eight isolates. Seven loci either failed to amplify or produced uninterpretable banding patterns when electrophoresed in 2.0% agarose gels. The remaining 34 microsatellite loci were tested on a larger sample of 102 \textit{M. lini} isolates. Within this sample, 98 isolates were collected from \textit{L. marginale} and were selected on the basis of their unique virulence profiles and ability to represent the geographical range of the \textit{M. lini} in Australia. The remaining four isolates of \textit{M. lini} were collected.
from a second host species (*Linum usitatissimum*) and were used to test for cross-
amplification and variability of microsatellite loci.

Each 20 uL PCR reaction contained 1 x PCR Buffer (10 mM Tris-HCl, 50 mM KCl),
200 uM of dCTP, dGTP, dTTP and dATP, 3.0 mM MgCl₂, 5 pmol labelled M13
primer (6-Fam, Hex and Ned—CACGACGTGTAAAAACGAC) 1 pmol 5’ tagged
forward primer, 5 pmol reverse primer, 1 U Taq polymerase (Applied Biosystems)
and 30 ng of rust genomic DNA. PCR amplifications were performed on a Hybaid
Express thermocycler under the following conditions: 95 °C 4 min, 15 cycles of 94°C
30 s, 65-50°C 30 s decreasing by 1°C per cycle, 72°C 80 s, 30 cycles of 94°C 15 s,
50°C 15 s, 72°C 45 s, followed by a 4°C holding step.

For Genescan electrophoresis, fluorescent samples were diluted by adding 2 uL of 6-
Fam samples, 6 uL of Hex samples and 2 uL of Ned samples plus sterile water to a
total volume of 100 uL. 2.5 uL of this solution was then added to a 64-well plate and
dried in a PCR block at 65°C. Prior to gel electrophoresis, samples were resuspended
in 1.5 uL of loading mix (1.2 uL Blue Dextran loading solution, 0.3 uL Internal Lane
Standard 600, Promega). Denaturing acrylamide gels (5.0%) were prepared from
Long Ranger XL Singel Packs (Cambrex), 1.5 uL of sample was loaded onto a 64
lane membrane comb (The Gel Company), and the gels were run on an ABI 377
sequencer using Genescan software (Applied Biosystems). Fragment sizes were
estimated using Genescan and Genotyper software (Applied Biosystems).

Of the 34 microsatellites tested, 11 produced repeatable banding patterns and were
polymorphic (Table 1). Based on the presence of unique alleles, the generated data
clearly discriminated between two lineages of *M. lini* collected from *L. marginale* and
a third lineage consisting of isolates collected from *L. usitatissimum*. Overall, a
surprisingly high level of clonality was observed with microsatellites given the
geographic and pathotypic variation represented in the sample. For example, the 98
samples (94 unique pathotypes) collected Australia wide from *L. marginale* are
represented by only 8 unique microsatellite genotypes. The presence of linkage
disequilibrium for individual loci was tested using the program FSTAT (Goudet,
1995). A high level of linkage disequilibrium (40 of 55 possible comparisons, P<
0.05) was observed, reflecting the low levels of genetic variation within Australian

Appendix
lineages and the genetic differentiation between the two Australian lineages collected from *L. marginale* and isolates collected from *L. usitatissimum*. It does however suggest that these markers will be useful for studying the evolution of host specialized lineages of *M. lini*. The low genetic diversity observed in the Australian wide collection of *M. lini* isolates suggests that these markers have a limited use in population studies, but will be most useful to track migration of unique genotypes.

**Table 1.** Polymorphic microsatellite loci from *Melampsora lini*. Shown are the locus name, repeat motif, primer sequences, number of alleles, allele sizes and gene diversity obtained for 102 isolates.

<table>
<thead>
<tr>
<th>Locus</th>
<th>repeat motif</th>
<th>Primer sequences (5'-3')</th>
<th>Label</th>
<th>no. alleles (size)</th>
<th>Genbank accession no.</th>
<th>Gene diversity*</th>
</tr>
</thead>
<tbody>
<tr>
<td>M10</td>
<td>(GAA)n</td>
<td>CATTCCGTAACCATC</td>
<td>HEX</td>
<td>3</td>
<td>DQ393817</td>
<td>0.403</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TCAATGAGAAATGAAATCT</td>
<td></td>
<td></td>
<td>DQ393818</td>
<td>0.449</td>
</tr>
<tr>
<td>M11</td>
<td>(AG)n</td>
<td>AGACGCTGATCTATTTCAAG</td>
<td>FAM</td>
<td>5</td>
<td>DQ393819</td>
<td>0.449</td>
</tr>
<tr>
<td>M12</td>
<td>(GTA)n</td>
<td>ACCTCATTATGCTGATG</td>
<td>FAM</td>
<td>5</td>
<td>DQ393820</td>
<td>0.413</td>
</tr>
<tr>
<td>M13</td>
<td>(AAG)n</td>
<td>AACACACACACACACCACTC</td>
<td>HEX</td>
<td>3</td>
<td>DQ393821</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TCCACCCACACACACACAC</td>
<td></td>
<td></td>
<td>DQ393822</td>
<td>0.407</td>
</tr>
<tr>
<td>M14</td>
<td>(AC)n</td>
<td>CTGGGGCTGCTAAGTTT</td>
<td>FAM</td>
<td>3</td>
<td>DQ393823</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCTGCTGCTGCTGCA</td>
<td></td>
<td></td>
<td>DQ393824</td>
<td>0.403</td>
</tr>
<tr>
<td>M15</td>
<td>(TTTGTATT)n</td>
<td>TTGAGAGCATGATCTTCATC</td>
<td>FAM</td>
<td>3</td>
<td>DQ393825</td>
<td>0.396</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAAAATCTGCTACAC</td>
<td></td>
<td></td>
<td>DQ393826</td>
<td>0.030</td>
</tr>
<tr>
<td>M16</td>
<td>(AAG)n.d(TTGATT)n</td>
<td>CCGGTTAGTTAAGCTCTGCAACG</td>
<td>FAM</td>
<td>2</td>
<td>DQ393827</td>
<td>0.407</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AATCGGTTAGTTAAGCTCTGCA</td>
<td></td>
<td></td>
<td>DQ393828</td>
<td>0.030</td>
</tr>
<tr>
<td>M17</td>
<td>(TCT)n</td>
<td>CCGTCTAAGATGTGCTGCTG</td>
<td>FAM</td>
<td>2</td>
<td>DQ393829</td>
<td>0.401</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CTTGCTAAGATGTGCTGCTG</td>
<td></td>
<td></td>
<td>DQ393830</td>
<td>0.404</td>
</tr>
<tr>
<td>M18</td>
<td>(AGA)n</td>
<td>TCAAGAAGAAGGTTAAGCTC</td>
<td>NED</td>
<td>2</td>
<td>DQ393831</td>
<td>0.396</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CTGCTAAGAAGGTTAAGCTC</td>
<td></td>
<td></td>
<td>DQ393832</td>
<td>0.030</td>
</tr>
<tr>
<td>M19</td>
<td>(TTA)n</td>
<td>TTAATCTCCTTGCTCAACCC</td>
<td>NED</td>
<td>2</td>
<td>DQ393833</td>
<td>0.396</td>
</tr>
<tr>
<td></td>
<td>(AGA)n</td>
<td>GATTACCTACTCCTCAACC</td>
<td></td>
<td></td>
<td>DQ393834</td>
<td>0.030</td>
</tr>
</tbody>
</table>

*Gene diversity \((1 - \sum p_i^2)\), where \(p_i\) is the frequency of the \(i\)th allele (Nei 1973)
Literature cited


