ECOLOGICAL PERSPECTIVES ON HOST-PARASITE COEVOLUTION

Samuel Thormond Dobbie

Submitted for the degree of PhD, November 2013

University of East Anglia

John Innes Centre/School of Biological Sciences

ACKNOWLEDGEMENTS

Contrary to the official line, no-one produces a PhD on their own. It takes academic supervision, technical help and a large contingent of friends and family willing to put up with your whining, panicking and occasional mental break-downs. Actually the last one might just be me. With that in mind, there are a few people I should thank.

Firstly, obviously, my supervisor James Brown, who not only took me on, directed my early research and reacted with excitement and encouragement when I had interesting initial results but was also very supportive and encouraging when I needed it. Secondly the rest of my advisory committee, Richard Morris and Luzie Wingen, and the student advisory committee who oversaw my intercalation, re-instatement, deadline extension, second deadline extension and a whole bunch of related admin stuff I wouldn't want to touch with a barge-pole.

My parents, without whose encouragement I'd never have started or resumed this project and without whose more practical support I'd long since have starved to death.

The Brown research group, for an enormously supportive, fun and tea-drinking working environment. Also for listening to my highly mathematical talks and nodding politely.

My friends, without whom I'd have gone crazy a lot earlier.

Thanks, guys.

ABSTRACT

It is a truth universally acknowledged that polymorphism at host immunity loci and corresponding parasite antigenicity loci, maintained by coevolution in pathosystems, is common and can persist for millions of years. Such polymorphisms and how they persist or break down are both fundamentally interesting and important for human health and agriculture. Examples include the major histocompatibility complex in vertebrates and the gene-for-gene (GFG) relationships in plants and their parasites.

GFG systems are well-understood genetically and an important source of disease resistance for plant breeders. Therefore considerable effort has gone into studying their evolutionary dynamics in natural pathosystems and modelling the conditions under which long-term polymorphism persists or breaks down. Polymorphism in GFG systems is common and in many cases ancient in wild pathosystems. Conversely, in agriculture the introduction of a resistance gene normally results in the matching parasite avirulence gene rapidly becoming locally extinct.

Simple genetic models of GFG coevolution do not produce stable polymorphism. Various more complex models do but are difficult to analyse. Recent work has shown a factor common to stable models is negative direct frequency-dependent selection, so at least one genotype becomes less fit as it grows more common regardless of genotype frequencies in the other species. This selection is not present in simplified models, but is generated in real pathosystems by various ecological and epidemiological factors.

Here I expand on previous work by demonstrating that realistic demography, specifically density-dependent regulation of parasite incidence, can generate negatively self-regulating stabilising pressure. This is loosely analogous to negative frequency-dependent selection and, similarly, can stabilise polymorphism in GFG pathosystems. I show this density-dependent regulation can stabilise both non-spatial deterministic and spatial stochastic systems. I also study how this stabilising factor interacts with the complicating biological factors of limited dispersal and resultant spatial structure in populations, variable host density and the presence of a second parasite.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1 - INTRODUCTON

OVERVIEW

Polymorphism at host immunity loci and corresponding parasite antigenicity loci is common and can persist for millions of years. Examples include the major histocompatibility complex in vertebrates and the gene-for-gene (GFG) relationships in plants and their parasites. Such polymorphisms and the conditions under which they persist or break down are important from both pure and applied scientific perspectives. From a pure science perspective, the dynamics explain the frequencies and diversities of immunity genotypes observed in nature. From an applied science perspective there are implications for human health and agriculture.

GFG systems are well-understood genetically and in many cases biochemically, as well as being an important source of disease resistance for plant breeders. Therefore considerable effort has gone into studying their evolutionary dynamics in natural pathosystems. Much work has also gone into modelling and analysing the conditions under which long-term polymorphism persists or breaks down.

GFG systems are defined by resistance/susceptibility loci in plants and specific corresponding virulence/avirulence loci in parasites. Polymorphism in both these interacting genes is common and in many cases ancient in wild pathosystems. Conversely, in agriculture the introduction of a new resistance gene normally results in the loss of the matching parasite avirulence gene within a few years.

Simple genetic models of GFG-coevolution do not duplicate the long-term stable polymorphism known to exist in nature. Various more complex models do but are difficult to analyse. Recent work has shown a common factor in models capable of stable polymorphism is negative direct frequency-dependent selection on one or more of the four genotypes involved, meaning the genotype becomes less fit as it grows more common regardless of genotype frequencies in the other species. Negative direct frequency-dependent selection is not present in simplified models, but occurs in real-world pathosystems as a result of various ecological and epidemiological factors.

In this thesis I expand on previous work by demonstrating that realistic demography, specifically density-dependent regulation of parasite incidence, can generate negatively self-regulating stabilising pressure. This is loosely analogous to negative frequency-dependent selection and, similarly, can allow long-term stable polymorphism in hosts and parasites. I show density-dependent regulation of parasites allows stability in both non-spatial deterministic and spatial stochastic systems. I also explore how this stabilising factor interacts with other complicating biological details, primarily limited dispersal and resultant spatial structure in populations.

This introductory chapter contains:

An overview of the gene-for-gene model including definitions, coevolutionary dynamics and the terminology used in this thesis (Section 1.1).

A summary of the deterministic models of gene-for-gene coevolution that preceded this thesis and their main results (Section 1.2).

A brief description of spatial models in this area (Section 1.3).

A summary of the three results chapters and the general discussion chapter (Section 1.4).

1.1 - THE GENE-FOR-GENE MODEL

1.1.1 - Gene-for-gene resistance is ubiquitous in plants, well-understood genetically, binary in nature and thus easy to measure and of both theoretical, evolutionary, and practical, agricultural, importance

In plant pathology, disease resistance is typically divided into two categories. Partial resistance is usually controlled by multiple genes. It is non-race-specific, provides less than total protection and is durable in agriculture (St Clair 2010). Gene-for-gene (GFG) or race-specific resistance is associated with specific pairs of genes in host and parasite species. GFG resistance provides total or nearly total protection to the plant and a corresponding total or nearly total loss of fitness to the parasite. In nature, polymorphism in GFG alleles can persist for millions of years (Tian et al. 2003, Bakker et al. 2006).

Conversely in modern agricultural systems GFG-resistance is normally overcome within a few years - parasites evolve corresponding virulence genes and these rapidly go to fixation (Brown & Hovmoller 2002). Compared to partial resistance GFG resistance is better understood genetically, is easier to quantify and so measure in the field and is easier to select for breeders to introduce to cultivars. The same simpler genetics make it possible to measure allele frequencies in wild pathosystems (Thrall & Burdon 2000, Laine 2005, Laine & Hanski 2006). In some cases, genetic analysis has been used to calculate allele ages and thus how long polymorphism has existed (Tian et al. 2003, Bakker et al. 2006).

The above factors mean that GFG systems are a well-understood case of host-parasite coevolution and thus a natural focus for experimental and theoretical evolutionary biology studies, in addition to their agricultural importance. The dynamics of GFG resistance in natural and agricultural pathosystems have been the subject of intense research including both real-world studies (Thrall & Burdon 2000, 2003, Brown 2003A, Laine 2005, Laine & Hanski 2006, Thrall et al. 2012, Tack et al. 2012) and modelling studies (Jayakar 1970, Leonard 1977, 1994, Gandon et al. 1996, Sasaki 2000, Sasaki et al. 2002, Thrall & Burdon 2002, Salathe et al. 2005, Segarra 2005, Tellier & Brown 2007A, 2007B, 2009, 2011).

A related area of modelling covers spatial elements of host-parasite coevolution with genetics that are not necessarily GFG (Gandon 2002, Gandon & Michalakis 2002, Morgan et al. 2005, Nuismer 2005, Nuismer & Otto 2006, Gandon & Nuismer 2009). Major conclusions are that local adaptation occurs in the species with greater dispersal when dispersal is limited for both species (Gandon 2002) and that geographic mosaics of strong and weak coevolutionary pressure can drastically alter observed patterns of local adaptation, making analysis of real data-sets difficult (Nuismer 2006).

In addition to wild GFG systems, spatial and temporal dynamics of host-parasite coevolution have also been investigated in evolutionary microcosms. Study systems include bacteria and phage (Lenski & Levin 1985, Bohannon et al. 1999, 2000, Elena & Lenski 2003, Brockhurst et al. 2006, Poullain et al. 2008), water-fleas and their parasites (Ebert et al. 2008) and moth larva and viral parasites (Boots et al. 2007, 2009). Major conclusions are that limiting dispersal allows longer-term coexistence of multiple genotypes (Brockhurst et al. 2006) and dampens oscillations in species density (Boots et al. 2009). Coevolutionary microcosms, analogies between bacteria-phage and GFG genetics

and relevance to hypotheses about GFG coevolution are further discussed in Chapter 5 (Sections 5.3.2-5.3.3).

1.1.2 - GFG definition and vocabulary

The GFG model was first proposed by Flor (1955). In a GFG relationship between a host and a parasite species, there exists one or more sets of genes such that a specific parasite avirulence gene corresponds to a specific host resistance gene. The products of resistance genes recognise the products, or downstream results of the products, of avirulence genes. If a parasite has a specific avirulence gene and a plant the corresponding resistance gene, the plant will launch an immune response and the parasite will fail to colonise or have its success massively reduced. In all other cases the parasite will colonise successfully (Figure 1.1, next page). Successful and unsuccessful parasite colonisations result in compatible and incompatible reactions respectively. The biochemical mechanisms underlying GFG interaction are reviewed in Section 1.1.3.

While plants can have hundreds of resistance genes (Jones & Dangl 2006), most are specific in detecting the product of one single avirulence gene. Some resistance genes can detect more than one avirulence gene, for example RPM1 in *Arabidopsis* spp. (Mackey et al. 2003), but the one-to-one relationship applies generally (Dodds & Rathjen 2010) and is typically assumed by modellers simulating more than one pair of genes (Sasaki 2000, Sasaki et al. 2002, Salathe et al. 2005, Segarra 2005).

Both resistance and avirulence genes can have alleles that do not participate in the detection reaction. Resistance genes can have alleles called susceptibility genes that do not detect avirulence genes. Similarly, avirulence genes can have virulence alleles that are not detected by resistance genes. In wild pathosystems, polymorphism between these alleles in matching sets of genes is common and can persist for millions of years (Tian et al. 2003). Conversely in agriculture resistance genes are introduced by breeders, maintained at relatively high levels by breeders and virulence genes typically arise by mutation and overcome resistance within a few years (Brown & Hovmoller 2002).

In the GFG system the term virulence refers purely to an allele of a specific avirulence gene that escapes detection by a matching specific resistance gene. Virulence has a different meaning in animal disease literature, which is broadly the cost to the host of being

infected. More recent plant pathology papers sometimes adopt this convention and refer to the ability to infect hosts as pathogenicity or infectivity (Sacristan & Garcia-Arenal 2008), but I do not adopt this convention.

Throughout this thesis the notation is that *RES* = resistance (genes, genotype or individuals), *res* = susceptibility, *avr* = virulence and *AVR* = avirulence. *R* and *r* are the frequencies within the host populations of individuals with genotypes *RES* and *res* respectively. Similarly *a* and *A* are the frequencies within the parasite population of individuals with genotypes *avr* and *AVR* respectively. $R + r = 1$ and $a + A = 1$. This notation assumes the studying or modelling of only one resistance/susceptibility gene and one matching virulence/avirulence gene. The assumption of one gene with two alleles in each species applies to all the models and results described in this thesis.

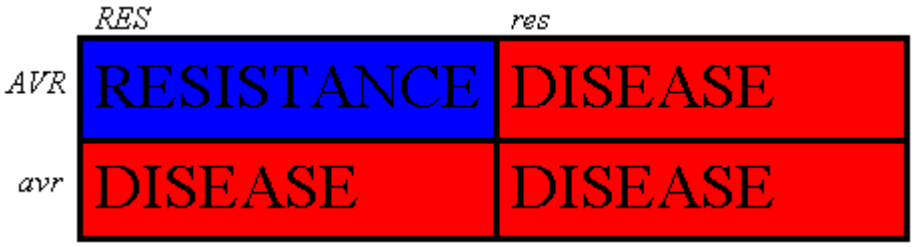

Figure 1.1 – Outcomes of gene-for-gene interactions. If a specific *AVR*-protein is detected by a corresponding specific *RES*-protein the infection is inhibited, resulting in an incompatible interaction. In all other cases – if the *AVR*-protein, the *RES*-protein or both are absent – the infection is not inhibited and a compatible interaction occurs.

1.1.3 - Biology of *RES*/*res* and *AVR*/*avr* genes

An *AVR*-protein is any parasite protein detected by a host *RES*-protein. Thus, any parasite protein that is detectable by the host could potentially become an *AVR*-protein. It is clearly disadvantageous for parasites to be detected by the host, so *AVR*-proteins must have functions that benefit parasite fitness in the absence of detection. Plant immune responses, including partial and GFG resistance and the biological function and diversity of *AVR*- and *RES*-proteins, are reviewed by Jones & Dangl (2006) and Dodds & Rathjen (2010).

Plant immune responses divide into two levels, PTI or PAMP-triggered immunity and ETI or effector-triggered immunity (Jones & Dangl 2006, Dodds & Rathjen 2010). PAMPs are pathogen-associated molecular patterns and many are highly conserved parasite-associated molecules such as bacterial flagellin, fungal chitin and bacterial elongation factor Tu

(Boller & Felix 2009). The host proteins that detect these are known as PAMP-recognising receptors or PRRs and are typically transmembrane proteins (Caplan et al. 2008, Monaghan & Zipfel 2012, Feng & Zhou 2012). PTI is sometimes regarded as an aspect of partial resistance but there is as yet little direct evidence to show that this is the case (Dodds & Rathjen 2010).

The PAMP molecules detected by the PTI mechanism are vital for the parasite, often as structural elements, and are necessarily highly conserved and highly expressed. Parasite infection strategies therefore rely on suppressing or overcoming partial resistance, rather than not triggering it.

In contrast to the parasite genes detected by the PTI system, which are widely conserved across species, parasite genes involved with GFG resistance vary greatly between and within species (Boller & Felix 2009, Dodds & Rathjen 2010, Feng & Zhou 2012). *AVR*-genes are best understood in bacteria. Parasitic bacteria typically express 15-30 *AVR*-proteins, also called effectors. These enter host tissue via a protein complex called the type-three secretion system. Many effectors suppress elements of host partial resistance. Others have less well-understood functions involving altering host metabolism. Oomycetes and fungi have larger secretomes of hundreds of effectors, which are less well understood than their bacterial counterparts. Effectors are normally expressed outside the parasite cells, meaning they can easily both impact host processes and be detected by *RES*-proteins.

Unlike detectors for the PTI system, many *RES*-proteins are expressed inside host cells. In contrast to the variety of *AVR*-genes many *RES*-genes code for two groups of protein (Afzal et al. 2008). Nucleotide-binding leucine-rich-repeat or NB-LRR proteins intracellular and detect parasite effectors that infiltrate host cells (Jones & Dangl 2006), while receptor-like kinase proteins or RLKs are trans-membrane proteins that can detect extracellular elicitors (Afzal et al. 2008). In both families a leucine-rich-repeat domain binds to an elicitor and an attached domain then triggers plant defence pathways (Afzal et al. 2008). Each *RES*-protein recognises one or more different *AVR*-factors. *RES*-genes are often clustered in the genome and close to elements that facilitate duplication, suggesting duplication provides "spare" *RES*-genes that can evolve new specificities (Holub 2001). This allows both rapid evolution and large numbers of *RES*-genes.

RES-proteins can interact with *AVR*-proteins in two ways, by directly binding to the parasite protein (direct interaction) or by responding to changes in a host protein targeted by the parasite (indirect interaction or guard hypothesis). Parasite effectors detected by direct interaction include *Magnaporthe grisea* AvrPita and flax rust *Melampsora lini* AvrL567 (Dodds et al. 2004) and AvrM (Canazariti et al. 2010) proteins. The *Arabidopsis thaliana* protein RIN4 is a well-studied example of indirect interactions. RIN4 is targeted by multiple parasite effectors including the *Pseudomonas syringia* effectors AvrB, AvrRPM1 and AvrRpt2 (Axtell & Staskawicz 2003, Mackey et al. 2003). These interactions are detected by the resistance protein RPM1.

Direct interaction should be easier for parasites to overcome than indirect interaction, as small sequence changes in effectors could theoretically avoid detection. Supporting this, flax-rust effectors and corresponding *RES*-genes show strong evidence of diversifying selection, which is not seen in the RIN4 protein sequence (Dodds & Rathjen 2010). Jones & Dangl (2006) suggested deletion mutations of specific *AVR*-genes as a possible mechanism to avoid resistance. This requires that the effector's role can be adequately complemented by other effectors (Jones & Dangl 2006). *AVR*-genes can exist in large families capable of rapid duplication, which would facilitate such redundancy of effectors. For example the *AVRk1-AVRa10* gene family in *Blumeria graminis* is associated with a LINE-1 retrotransposon element, which has led to rapid duplication and evolution of new specificities (Sacristan et al. 2009), while *Phytophtora sojae* and *P. ramorum* effectors are associated with deletion and duplication events (Jiang et al. 2006).

Alternatively a parasite's losing one effector and not targeting one host protein may reduce but not eliminate infection success, due to the multitude of effectors a parasite typically possesses. Whether an *AVR*-allele mutates to avoid detection or is lost altogether, the result is an *avr*-allele.

The defence triggered by *RES*-genes varies. In many cases *RES*-genes trigger localised cell death, termed the hypersensitive response. This is effective in halting infection by biotrophs and potentially hemibiotrophs, but not necrotrophs (Jones & Dangl 2006). Other *RES*-mediated plant responses proceed differently. For example, tomato resistance to *Cladisporium fulvum* involving Cf2 and Cf9 *RES*-genes does not require cell-death (Brading et al. 2000). In the case of some fungal parasites, the penetrating hyphal growth

slows and eventually stops. In these cases it is unclear whether the fungi are dead or merely constrained (Skamnioti & Ridout 2005).

To summarise, the products of *RES*- and *AVR*-genes interact in a disease-detection relationship. *res*- and *avr*- alleles are not active in this relationship. This accounts for the dominance orders *RES* > *res* and *AVR* > *avr* commonly used in modelling. However some authors assume co-dominance, for example Nuismer & Otto (2005).

1.1.4 - GFG systems are coevolutionary

Coevolution is a situation in which two or more species exert evolutionary pressure on each other, evolve in response to this pressure and change the pressure they exert on the other species as they evolve. A simple example is speed in predators and prey – as each species becomes faster, the selection pressure it imposes on the other species increases and the other species evolves to become faster in turn. This appears to be unidirectional pressure that only varies in intensity, as it is never advantageous to be slower with respect to selection pressure imposed by the other species. However there are fitness costs to being fast due to heightened energy and nutrition needs and increased vulnerability to injury. Thus the net direction of selection on species can vary. Overall selective pressure on each species to be faster will be positive or negative depending on the relative magnitudes of the benefits of being faster, which depend on the other coevolutionary species, and the costs of being faster, which do not.

In GFG-systems, host and parasite populations impose and react to varying selective pressures on one another. This is a natural consequence of the antagonistic nature of the host-parasite interaction and the interactions defined by the GFG model (Flor 1955, 1971, Brown & Tellier 2011). The cost to the host of being diseased means selection for host resistance is exerted by *AVR*-parasites, so the magnitude of the fitness advantage *RES*-hosts have over *res*-hosts depends on the frequency of *AVR*-parasites. Similarly the cost to the parasite of being detected means selection for parasite virulence is exerted by *RES*-hosts, so the magnitude of the fitness advantage *avr*-parasites have over *AVR*-parasites depends on the frequency of *RES*-hosts.

From the above, it seems that while selective pressures can vary in intensity they are unidirectional. This would mean that *RES*-hosts are always fitter than *res*-hosts and avr-parasites are always fitter than *AVR*-parasites, implying that *res*-hosts and *AVR*-parasites

always go extinct. However in wild pathosystems polymorphism is common and can persist for millions of years (Bakker et al 2006, Tian et al. 2003). This suggests there are intrinsic fitness costs to resistance in hosts and virulence in parasites, analogous to the costs of being faster mentioned above, that balance the costs of susceptibility and avirulence. Theories and evidence for intrinsic costs of *RES*- and *avr*-alleles are reviewed in Sections 1.18 and 1.19 respectively.

If intrinsic fitness costs for *RES*- and *avr*-genes exist, *RES*-hosts are only fitter than *res*-hosts when the average fitness penalty experienced by *res*-hosts because of infection by *AVR*-parasites exceeds the fixed fitness penalty intrinsic to *RES*-hosts. Likewise, *avr*-parasites are only fitter than *AVR*-parasites when the average fitness penalty experienced by *AVR*-parasites because of their inability to infect *RES*-hosts exceeds the fixed fitness penalty intrinsic to *avr*-parasites. Thus whether *RES*- or *res*-hosts are fitter depends on the frequency of *AVR*-hosts and the relative size of intrinsic and extrinsic fitness costs. Likewise, whether *avr*- or *AVR*-parasites are fitter depends on the frequency of *RES*-hosts and the magnitudes of costs. The effect of genotype frequencies in one species on genotype fitnesses in another is indirect frequency-dependent selection, described in Section 1.16.

In large populations, genotypes become more common when they are fitter than the opposing genotype and less common when they are less fit. As genotype frequencies in each species determine relative fitnesses of genotypes in the other species there are linked oscillations in genotype frequencies, described in Section 1.1.5.

Box 1.1 - Consequences of costs

Costs reduce reproductive fitness. Thus an individual paying a cost or costs totalling 0.05 will produce on average 95% as many seeds or spores as an individual not paying any costs. This could represent a 5% chance of dying before reproducing, a 5% reduction in seeds or spores produced or a combination of both. In population-based models it makes no difference, although in individual-based models it does and the modeller must decide which is the case. Throughout this thesis I have modelled individuals of lower fitness as producing fewer offspring rather than having a percentage chance of dying before reproducing.

In population-based models such as Tellier & Brown (2007A) and Models 1 and 2 (Chapter

2), one average fitness for each genotype is calculated based on the frequencies of genotypes in the other species and the magnitudes of costs. In individual-based models such as Models 3-5 (Chapters 3 and 4), each individual has its own fitness calculated based on its own genotype, the presence/absence and genotype of an individual of the other species and the magnitudes of costs.

1.1.5 - iFDS and ndFDS - indirect and negative direct frequency dependent selection

The term frequency-dependent selection has been used in two senses in coevolutionary literature. What is described in most evolutionary biology textbooks is direct frequency-dependent selection (Futuyma 1996, Hedrick 2009), where the fitness of individuals of a particular genotype or phenotype depends on their frequency within the population. However research articles on coevolution often discuss indirect frequency-dependent selection, where relative genotype or phenotype fitnesses in one species are affected by genotype frequencies in the other species (Chaboudez & Burdon 1995, Carius et al. 2001). Stable GFG polymorphism actually requires both, necessitating a clear distinction in discussion. To facilitate this distinction, the terms direct and indirect frequency-dependent selection were introduced by Tellier & Brown (2007A).

Direct frequency-dependent selection can be negative or positive, that is genotypes or phenotypes can have their fitness reduced or increased with increasing frequency within a species. As well as acting on individuals of one genotype within a species, direct frequency-dependent selection can act on individuals of one species within a community. Sources of negative direct frequency-dependent selection include disease pressure (Holt 1985, Kohler 2001), predator focal choice (Tinbergen 1951), competition for resources in plants where genotype affects resource partitioning (Ellstrand & Antonovics 1984) and mate-choice favouring rare phenotypes (Spiess 1968).

Most of my results are actually concerned with density-dependent selection on the parasite species rather than direct frequency-dependent selection on either parasite genotype. Density-dependent selection is similar to direct frequency-dependent selection. The difference is that density-dependent selection depends on and regulates the density of individuals of one species within a community, rather than the frequency of individuals of one genotype within a population (Hedrick 2009).

In GFG coevolution both indirect and negative direct frequency-dependent selection are required for stable polymorphism. Indirect frequency-dependent selection is necessary for the pathosystems to have an equilibrium where all genotypes are present. Negative direct frequency-dependent selection is necessary for a system's internal equilibrium point to be stable so long-term polymorphism can persist (Tellier & Brown 2007A). In diploids, heterozygote advantage can replace negative frequency-dependent selection as the factor stabilising polymorphism (Ye et al. 2003).

1.1.6 - Indirect frequency-dependent selection, iFDS - relative genotype fitnesses in each species are determined by genotype frequencies in the other species – caused by costs inherent to individuals or one-to-one genotype reactions

GFG coevolutionary systems experience indirect frequency-dependent selection, iFDS (Tellier & Brown 2007A), as follows. Whether *AVR*- or *avr*-parasites are, on average, fitter within a given population at a given time depends on the relative frequencies of *RES*- and *res*-hosts. Similarly, the relative average fitnesses of *RES*- and *res*-hosts depend on the relative frequencies of *avr*- and *AVR*-parasites. This iFDS in both species leads to linked oscillations in genotype frequencies. On a graph of *R* against *A*, genotype frequencies circle around an internal equilibrium point (IEP) where *RES*- and *res*-hosts and *avr*- and *AVR*-parasites have equal fitness and no net selection occurs (Figure 1.2, overleaf).

iFDS means the frequencies of genotypes in each species at the IEP depend on the relative magnitude of costs in the other species. Equilibrium *R* or $\hat{R}$ is equal to the cost of avirulence divided by the cost of being detected by *RES*-hosts. Similarly, equilibrium *A* or $\hat{A}$ is proportional to the cost of resistance divided by the cost of being infected.

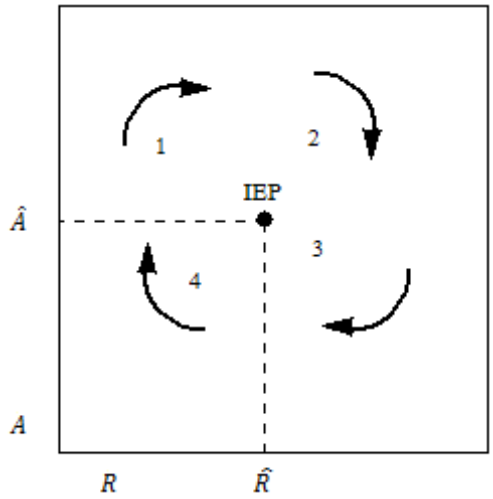

Figure 1.2 - Gene frequencies cycle around the internal equilibrium point or IEP. R = the frequency of a specific *RES*-gene in the host population and A = the frequency of the matching specific *AVR*-gene in the parasite population. $\hat{R}$ and $\hat{A}$ are the equilibrium frequencies at the IEP. At the IEP, $R = \hat{R}$ so *AVR*- and *avr*-parasites have the same fitness and $A = \hat{A}$ so *RES*- and *res*-hosts have the same fitness. When R is below $\hat{R}$ AVR-parasites are fitter than avr-parasites and A increases (1, 4), while when R is above $\hat{R}$ AVR-parasites are less fit than avr-parasites and A decreases (2, 3). Similarly when A is above $\hat{A}$ *RES*-hosts are fitter than res-hosts and R increases (1, 2) and when A is below $\hat{A}$ *RES*-hosts are less fit than res-hosts and R decreases (3, 4). These selection pressures lead to cycling around the IEP.

1.1.7 - Negative direct frequency-dependent selection, ndFDS – fitness of one or more genotypes decreases as its own frequency increases regardless of genotype frequencies in other species – caused by population-level ecological and epidemiological factors

Coevolution has different outcomes in wild and agricultural systems. In agriculture, introduced GFG resistance is typically overcome by the parasite within a few years because *avr*-parasite genotypes arise by mutation and spread to fixation in the parasite population (Brown & Hovmoller 2002). In nature, by contrast, polymorphism in *RES*/*res* and *avr*/*AVR* genes can persist for millions of years. Further, simple modelling studies (Jayakar 1970, Leonard 1977, 1994) do not lead to stable polymorphism. Many later modelling studies do lead to stable polymorphism but prior to Tellier & Brown (2007A) it was not clear what the underlying mechanisms required for stability were.

The iFDS defined above means that the relative fitnesses of genotypes in one species depend on the frequencies of genotypes in the other species. This leads to linked oscillations circling about the IEP, but not to stable polymorphism. The circles expand over time until genotype fixation occurs (Jayakar 1970, Leonard 1977, 1994), while stable

dynamics would have circles shrinking until the system reached its IEP. Stable polymorphism requires a source of ndFDS, so that one or both genotypes within one or both species become less fit as they grow more common. This dampens increases in genotypes when they are most common and thus allows linked oscillations to spiral inward rather than outward. This is shown in Figure 1.2, below.

ndFDS on genotypes is an emergent property of pathosystems. It is caused by ecological or epidemiological processes that alter selection pressures on genes as the environment changes, rather than being an intrinsic property of the genes themselves (Brown & Tellier 2007A). Brown & Tellier (2007A, 2009) showed that this underlying mechanism explains multiple factors known to lead to stability in models. This showed the common assumption that complex, interacting factors are required to generate stability is false, although because complex models are more likely to include one or more factors generating ndFDS complexity has been perceived as necessary.

Tack et al. (2012) reviewed 29 spatial studies of pathosystems and reported spatial variation in pathogenicity occurred in 28 of them. This is indicative of temporal variation in pathogenicity consistent with ndFDS (Chaboudez & Burdon 1995, Thrall et al. 2012).

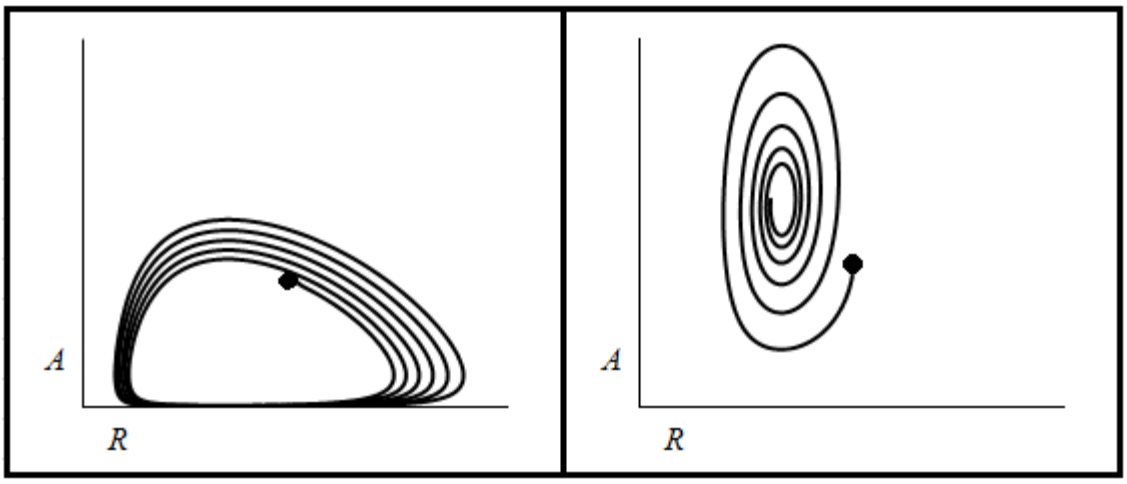

Figure 1.2 – Oscillations of gene-frequencies in coevolution with unstable and stable polymorphism of *RES*/*res* and *avr*/*AVR* genes. Initial frequencies are solid circles. Genotype frequencies circle clockwise over time. In the first panel polymorphism is unstable because ndFDS is either absent or not sufficiently strong to prevent the system spiralling outwards. In the second panel polymorphism is stable because ndFDS exists and is sufficiently strong. In an unstable system the IEP is an unstable equilibrium and circling frequencies move away from it over time, ending in genotype fixation. In a stable system the IEP is a stable equilibrium and circling frequencies move towards it over time and eventually converge on it. Classification of the IEP as stable or unstable is the purpose of the Jacobian analyses described in Section 1.110 and used in Chapter 2.

1.1.8 - Costs of *RES* and *avr* alleles – probable causes

The requirement for costs of *RES*- and *avr*-genes to allow stable polymorphism in GFG systems is explained in Section 1.13. Here I review possible biological explanations for these costs. Experimental evidence of costs is reviewed in Section 1.1.9.

In theory it is easy to see how costs for virulence and resistance could arise. For the parasite, an effector protein that has changed its structure or is no longer expressed may well be less able to perform its (presumably important) function. This is especially true for *AVR*-proteins which, as noted above, are often secreted into the host cell to facilitate infection and thus are both easily detected by hosts and important for infection. For the plant producing a specialised receptor protein would logically have a metabolic cost and thus a fitness cost, although this would probably be miniscule. More plausibly, it has been shown (Conrath et al. 2006, Vos et al. 2013) that *RES*-proteins can to some extent activate defensive pathways in the absence of *AVR*-proteins, either constitutively or in response to environmental stress. Resources wasted on such inappropriate defence could lead to reduced fitness.

1.1.9 – Costs of *RES* and *avr* alleles – evidence to date

Bergelson & Purrington (1996) provided a meta-analysis of eighty-eight experiments which were analysed to determines fitness costs in plants of resistance to herbicides, parasites and herbivores and reported an observed cost of resistance in 56% of the cases reviewed. As noted in Brown (2002) and by the authors themselves their use of separate experiments as identical data-points can lead to conclusions which are not necessarily accurate in detail. Also, it is possible this report suffers a bias due to negative results being less likely to be submitted as papers. Tian et al. (2003) reported a fitness-cost of 9% associated with one particular resistance gene that is believed to have co-existed with its susceptible allele for over nine million years, but this is a somewhat unusual result. To control other differences, the authors inserted the *RES*-allele into a susceptible plant line. Hosts with such a transplanted gene would not have a chance to develop compensatory mutations (below). Similarly, Kjaer (1990) reported a yield-reduction of approximately 4% for three different mutagen-induced *mlo* resistance genes – while these are not the same as *RES*-genes, the principle of hosts with recently introduced genes not having time

to gain compensatory mutations remains. Kjaer et al. reported that the costs could be reduced by re-assortment of background genes, which would occur over time due to natural selection in wild populations.

Laine & Barres (2013) recently reviewed studies of life-history trade-offs in parasites. They focussed on costs of virulence (14 studies, 9 fungal parasites, 4 viral, 1 oomycete) and found a majority of studies supported costs of virulence. They also found fitnesses of virulent and avirulent strains can be affected differently by abiotic factors such as temperature (Huang et al. 2010). Tack et al. (2012) reported that parasite populations are seldom dominated by universally virulent strains. This implies some kind of cost for virulence, perhaps a cost of having many virulence factors (Section 1.1.10). Thrall et al. (2003) found reduced spore production in strains of *Melampsora lini* with large numbers of identified virulence alleles, suggesting a cost of *avr*. Correspondingly, they reported a correlation between gene frequencies *R* and *a* at various sites. A review by Garcia-Arenal & Fraile (2013) found that viruses often experience significant fitness costs to both expanding host-ranges and avoiding *RES*-genes.

Experiments have found evidence for costs of virulence and resistance in many, but not in all, cases. Possible explanations follow. For both *avr*- and *RES*- genes, the magnitude of costs could be limited by factors relating to either the initial mutation or subsequent evolution. If multiple mutations can lead to virulence or resistance, those with the lowest cost would be favoured. After the evolution of virulence or resistance, compensatory mutations elsewhere in the genome that reduce the magnitude of the cost would be selected. Thus, natural selection could act to minimise or eliminate costs of *avr*- and *RES*-genes. Additionally, costs may only be evident when species are under other stresses or (for *RES*-genes) when pathways are inappropriately triggered. These factors, combined with natural biological variability which makes it hard to measure reductions in fitness of less than a few per cent, could explain the inconsistent evidence for costs.

1.1.10 - Other types of cost? Induced cost, extra loci cost, cost of uniformity, reduced competitiveness, heterozygote advantage

Simple, constitutive fitness costs of *avr*- and *RES*-genes are only one way to explain the existence of average fitness costs to individuals of a specific genotype. Considering the ambiguous evidence for them, there may be more complex factors at work than fixed

reproductive penalties. There is evidence that some *RES*-genes impose a fitness penalty only when expressed (Smedegaard-Petersen & Tolstrup 1985). This is known as a cost of induced resistance. Supporting this, many *RES*-genes are known to only be expressed in response to parasite elicitors (Brown 2003B, Vos et al. 2013) and constitutive expression is known to reduce fitness in the absence of parasites (Vos et al. 2013). A response known as priming prepares but does not fully express defence pathways (Vos et al. 2013, Conrath et al. 2006). This is less costly than constitutive defence and appears to be a compromise between constitutive defence and being unprepared for an attack, which would increase costs of disease. Priming is activated appropriately in response to pathogenic microbes and herbivores. However in some cases it is triggered inappropriately in response to benign microbes and mechanical wounding (Conrath et al. 2006). Tellier & Brown (2007B) reported that induced rather than constitutive costs of resistance reduce stable polymorphism. This is true if the cost is only paid when appropriately induced. However it does not apply to costs for inappropriately induced priming or resistance. If such inappropriate induction occurs, induced costs will manifest as a low average constitutive cost of resistance. Such a cost could easily be overlooked in studies that do not duplicate the conditions triggering inappropriate induction.

A rarer but more costly form of resistance is associated with mutations in metabolic pathways rather than *RES*-genes. Such loss-of-function mutations are typically recessive and are reviewed in Huckelhoven et al. (2013). They include loss of negative regulation of defences leading to constitutive defence expression, e.g. *HvCRK1* in barley (Rayapuram et al. 2012), and cell wall alterations, e.g. *pmr5* and *6* in *Arabidopsis*. In the latter case it is speculated that cell walls may lack factors powdery mildew needs to invade or that cells may produce different immune receptors (Vogel et al. 2004). Costs of constitutive defence are discussed above, while the cell-wall mutants *pmr5* and *6* have retarded growth.

Host-parasite interactions typically feature more than one GFG gene-pair. The way in which multiple *RES*- or *avr*- costs interact could affect stability, as well as making individual cost difficult to measure. Tellier and Brown (2007B) showed that synergy of costs of virulence and complementation of costs of resistance, i.e. increasing costs for each extra *avr*-allele and decreasing costs for each extra *RES*-allele, would both expand the stable parameter space for a system with a source of ndFDS.

There is some evidence for synergistic costs of virulence in parasites. Montarry et al. (2010) reported such costs in *Phytophthora infestans* isolates with between one and eleven *avr*-genes, while Wichmann & Bergelson (2003) presented data supporting such costs in *Xanthamonas axonpodis* pv. *vesicatoria*. Both studies examined multiple fitness components for parasites. More generally, a review by Frank (1992) noted that, although *avr*-alleles are much more common in populations than *RES*-alleles, relatively few parasite races have lost all their *AVR*-alleles. This supports the idea of common synergistic costs of virulence in parasites, as such costs would oppose the emergence of a universally virulent race.

Resistance genes often share loci whereas virulence genes do not (Flor 1971, Holub 2001). This may suggest there is a cumulative cost to the plant for each resistance locus (Brown 2003B). The pattern of multi-allelic resistance loci has been reported in flax (Flor 1971), common beans (Kelly 2004), *Arabidopsis thaliana* and barley (Brown 2003B). This loci-sharing property suggests there is a cost for each extra resistance locus but not necessarily for different alleles at the same locus. If each gene was on its own locus this would lead to synergy for costs of *RES*, destabilising coevolution (Tellier & Brown 2007). However locus-sharing could favour polymorphism between different RES-alleles. As one *RES*-allele becomes rare, the corresponding parasite *AVR*-genotype could become more common and give plants with the rare allele an advantage. Interestingly, if such a locus has many *RES*-genes and no *res*-genes and one of the corresponding *avr*-genes goes to fixation, the now-useless *RES*-gene will be recorded as a *res*-gene by any scientist investigating the system. If such a *RES*-gene persists for any length of time and maintains its fitness cost, this may further complicate measuring costs of resistance.

It is also possible that *avr*-alleles do not have a fixed cost but a potential cost depending on circumstances, i.e. environmentally dependent costs. Skamnioti & Ridout (2005) reported that many bacterial *AVR*-proteins may have a role in interfering with competing bacteria. The cost of not having such *AVR*-alleles would only be paid when these genotypes were actively in competition with other micro-organisms.

Heterozygote advantage or over-dominance is another possible cause of stable polymorphism in host-parasite coevolution (Lewontin 1958). If hosts that are heterozygous for resistance genes enjoy increased fitness, possibly due to heterozygote

advantage in associated genes, polymorphism in resistance genes could be preserved by selection of heterozygotes.

1.1.11 - Multi-lines and the advantage of host heterogeneity

Wolfe (1980) reported that multi-lines, mixtures of different cultivars planted together, produce higher biomass and higher grain yield than single-race crops. Wolfe suggested this was because disease spread between plants is less effective when the plants are less similar, an idea validated by Chin & Wolfe (1984A, 1984B). Thus there is a cost to the plant-population of uniformity, or a cost to each individual plant that depends on how close to the typical plant in that population it is. In populations where *RES*-genes are common, this could manifest as a cost of *RES*-genes. As a particular variety of plant starts to dominate the population selective pressure against it will increase. This could maintain equilibrium in the plant genotypes and thus the parasite genotypes.

In agriculture, using multi-lines slows the emergence of universally virulent parasite races (Wolfe 1980; Mundt 2002). However the authors warned that over several years a so-called super-race, universally virulent on all cultivars in the multi-line, can emerge. For this reason, they suggested changing the mix of cultivars regularly. Huang et al. (1994) reported that the average number of *avr*-alleles in *Blumeria graminis* f. sp. *hordei* was higher in mixed than in pure strands of spring barley but that even after two months of selection in thoroughly mixed strands less than 15% of parasites were universally virulent. This matches the observation in Frank (1992) that few parasites loss all their *AVR*-alleles and supports the idea of synergistic costs (Section 1.1.10).

Groth (1976) mathematically investigated the risk of a super-race emerging by modelling the relationship between the *avr*-genes and the relative fitness of parasites with different numbers of *avr*-gens competing in multi-line host populations exhibiting different single resistance genes. Groth reported that, for any non-zero cost of virulence, it is possible to have sufficient components in a multi-line that the universally virulent parasite will not be fitter than all other parasite races so will not go to dominance or fixation. However, the required number of cultivars may be very large. Sasaki (2000) also suggested that mixing large numbers of cultivars with a single resistance gene each may make it harder for super-parasites to evolve. Since host gene-mixtures in agriculture are controlled by the breeders rather than by nature, these papers suggest a practical way to restrict parasite virulence –

and, conversely, indicate that some agricultural strategies promote the emergence of super-races. However, since resistance to disease in general is rarely a plant-breeder's top priority and resistance to a specific disease is almost never so (Brown 2002), this is never deliberately done in practice.

1.1.12 – Coevolutionary outcomes - arms race or trench warfare?

There are two theories about how polymorphism at GFG loci is maintained (Holub 2001). The arms-race model suggests that new alleles arise by mutation (or immigration) and relatively quickly go to extinction or fixation. Which one a particular allele will go to depends on the genotype and response of the other species, as well as any fitness-costs it imposes. This implies that polymorphism reported at any given time is a snap-shot of a short-term event. It also implies that such events are common enough for such snap-shots to be observed frequently. The trench-warfare model suggests that once parasites have evolved to overcome a particular resistance allele it will decline to a low frequency in the population without going extinct, then increase in frequency once the pathogen population moves away from expressing the associated virulence allele at such high frequency. Thus, gene frequencies will either continue cycling in this manner indefinitely or reach a quasi-stable point. Tian et al (2003) supported the latter view by providing evidence that at least one *RES*/*res* polymorphism has existed for nine million years.

These two models epitomise the difference between unstable and quasi-stable coevolutionary cycles. While trench-warfare dynamics are common in nature (Bergelson et al. 2001), something like arms-race dynamics tend to dominate in modern agricultural pathosystems (Brown & Hovmoller 2002). The latter is not true coevolution as in agriculture the hosts are not subject to natural selection. Rather, farmers control levels of resistance year after year, so once avirulent genes occur by mutation parasites rapidly evolve to fixation in *a*. Conversely in wild pathosystems host populations are subject to natural selection, allowing true coevolution to occur. This could result in arms-race or trench-warfare dynamics and frequently results in the latter.

1.2 – MODELLING COEVOLUTION USING DETERMINISTIC POPULATION-BASED MODELS

1.2.1 - Simple models are deterministic, population-based and non-spatial

Much effort has gone into modelling the population dynamics of gene-for-gene pathosystems. Early models assumed populations had no spatial structure, no differences between individuals save for the genotype of interest and no chance events such as mutation, immigration or genetic drift (Jayakar 1970, Leonard 1977, 1994). No genetic drift makes the implicit assumption that populations are infinite, as otherwise there would be chance fluctuations in genotypes. In these early models, frequencies of genotypes therefore changed between generations in a purely fitness-dependent way.

There are advantages to such simple models. They are easy to design, describe and understand and as simple deterministic systems they are amenable to mathematical analysis by techniques from linear algebra (Tellier & Brown 2007A, Model 1 in Chapter 2). However they do not reflect spatial structure or chance events.

Early models without costs for resistance or virulence failed to maintain long-term polymorphism (Jayakar 1970). Many models developed since do successfully maintain polymorphism. However these models use differing mechanisms and there is no consensus on which are biologically accurate.

1.2.2 - Limit cycles

In coevolutionary models, a limit cycle is a situation where genotype frequencies over time describe a closed loop around the IEP and the cycle does not overall move towards either the IEP or the trivial equilibrium. There are two different types of limit cycle, found in deterministic and stochastic models.

In deterministic models a limit cycle represents a set of genotype frequencies that effectively form an equilibrium – frequencies cycle repeatedly and do not move towards or away from the IEP. Jayakar (1970) and Leonard (1977) reported such limit cycles. Like any equilibrium, limit cycles can be stable or unstable. A stable limit cycle means gene-frequencies will move from any point in the system towards that cycle, making the IEP and

the four trivial equilibria locally unstable. An unstable limit cycle means gene-frequencies at any point in the system will move away from the limit-cycle, making all five equilibria locally stable. In either case the limit cycle is a boundary, between the local stability/instability neighbourhood of the IEP and that of the four trivial equilibria.

More recent papers (Segarra 2002, Sasaki 2005) do not report deterministic limit cycles. They appear to be artefacts due to rounding errors in early models (Kirby & Burdon 1997). Crucially, without limit cycles there are no local neighbourhoods and the stability or instability of the IEP covers the entire phase-space. This means gene frequencies at any given time-point, excluding actual fixation, are irrelevant to whether the system be stable or not. The stability of the entire system can thus be predicted from the stability of its IEP.

In stochastic models with a stable IEP, small perturbations in genotype frequency due to chance events drive genotype frequencies away from the IEP. Frequencies then circle the IEP, with selection driving them towards it and stochasticity driving them away from it. Cycles tend to have roughly constant magnitude, presumably because the strength of ndFDS varies with distance from the IEP and at a certain distance it will balance chance perturbations. Tellier & Brown (2007A) observed this type of limit cycle in stochastic versions of their deterministically stable models, resulting in quasi-stable cycling around the IEP.

These stochastic deviations from stability are a specific case of a more general phenomenon. Any finite population will undergo stochastic deviations from fitness-driven predicted changes in frequency, possibly leading to genotype fixation. Such stochasticity will obscure the population dynamics driving coevolution. All else being equal, stochastic effects will decline with increasing population size. This is well-known for genetic drift, genotype loss in small populations and founder effects (Hedrick 2009). Stochastic genotype loss and the factors affecting it are discussed in Chapter 3, in the context of my Model 3.

1.2.3 – The problem with mutation

Several of the models discussed below (Sasaki 2000, Thrall & Burdon 2002, Salathe et al. 2005, Segarra 2005) included mutation and reported long-term polymorphism. Mutation in models confuses the issue of stable polymorphism as follows. Mutation between *RES*

and *res* or *avr* and *AVR* favours the rarer allele because the more common allele, being more common, is more likely to mutate to its opposite. This is true in both deterministic and stochastic models. In nature, new alleles introduced by mutation are distinguishable from old alleles maintained by selection. However in simple models such a distinction is not made and mutation reinforces populations of the rarer genotype in each species.

Mutation can still be a source of ndFDS in models. While the new alleles induced by mutation do not directly replenish the rarer genotype, the existence of mutation still attenuates the fitness of genotypes as they grow more common. This ndFDS is fairly weak unless mutation rates are set extremely high, so is unlikely to maintain stability by itself. However it might do so if added to other factors, for example limited slowing coevolution (Section 1.3.2). Even in these cases it is difficult to interpret whether old alleles are maintained in "stable" systems including mutation that do not differentiate between old alleles and those created by mutation.

This lack of clarity means models including mutation are more relevant to arms-race dynamics and frequent, overlapping transitory polymorphisms than to trench warfare dynamics and long-term or ancient polymorphisms. I am interested in long-term stability of polymorphism and thus trench-warfare rather than arms-race dynamics. Therefore, models in this thesis do not include mutation.

1.2.4 - Early models had unstable IEPs

Jayakar (1970) described a simple deterministic model without costs of resistance or virulence and went on to add variations for diploid species, possible survival of parasites that did not immediately find a host and a three-locus system with partial resistance and partial virulence. This model included automatic host death from the parasite, which is not a feature of most foliar crop diseases, but Jayakar also tested variations without this assumption. In all cases it was shown that the IEP, if it existed, was unstable. This is because the model has gene-for-gene interactions as a source of iFDS, leading to cycling genotype frequencies, but has no source of ndFDS so such cycles will inevitably be unstable.

Leonard (1977), analysed at greater length in Leonard (1993), described a simple deterministic model that differed from Jayakar's in having costs of resistance and virulence

as well as delayed rather than immediate genetic feed-back between hosts and parasites. In this model, delayed feedback can be a source of ndFDS (Tellier & Brown 2009, discussed in Section 1.2.5). Again the IEP is unstable, but Leonard (1993) stated it is possible for unstable limit cycles to exist so that genotype frequencies inside the limit cycles cannot reach the outside of phase-space and go to fixation. Leonard (1993) defines this as biological stability. This is an intriguing model, but it relies on limit cycles (Section 1.2.3, above) and only works with costs of resistance and/or virulence that are higher than most people would consider realistic (Bergelson & Purrington 1996).

1.2.5 - Later more complex models often had stable IEPs but many relied on mutation, which is a weak source of ndFDS

Subsequent modelling work has thus focussed on trying to identify factors that allow stability with low or non-existent costs of resistance and virulence. Sasaki (2000), Salathe et al. (2005) and Segarra (2005) all modelled multi-locus interactions with the possibility of mutation and found stable or quasi-stable cycles only when mutation supplemented rare genotypes. Segarra additionally reported that delayed as opposed to immediate feedback could lead to stable cycles, as mentioned in Leonard (1977). Sasaki also identified conditions under which *A* and *r* go to fixation instead of the more common *a* and *R*. These were linked to high costs of virulence and resistance and a cost to *AVR*-parasites of being detected below 1.

All these models have mutation as a source of ndFDS. While mutation is a theoretically possible source of ndFDS it is both generally weak and problematic to analyse, as discussed in Section 1.2.3. Delayed feedback as in Segarra (2005) is an attenuation of natural selection in time and is thus another source of ndFDS (Tellier & Brown 2009). Unlike mutation, such delayed feedback in nature would preserve existing polymorphisms rather than creating new alleles and is a plausible stabilising factor.

1.2.6 - ndFDS allows stability

For some time, it was thought that models required complexity and multiple interacting factors to generate stable polymorphism. Tellier & Brown (2007A, 2009, 2011) found that several biologically realistic additions to the basic model including multiple parasite generations per host generation and auto-infection (2007A), seed-banks and perenniality in

the host (2009) and subdivided populations in a spatially heterogeneous environment (2011) can lead to stable polymorphisms. Seed-banks cause delayed feedback, also reported by Segarra (2005) as a stabilising factor.

Crucially, this work provides a simple condition for stability – the ndFDS discussed in Section 1.1.7 – that can be created by a variety of biological processes. A key point is that all these processes are generated by the behaviour of the population as a whole, rather than being properties of individual organisms or genes. Another key point is that polymorphism can be stabilised by a single ndFDS-generating factor, provided that ndFDS is strong enough. It is not necessary to have complex systems with multiple interacting factors to promote stability.

1.3 – MODELLING COEVOLUTION USING SPATIAL MODELS

1.3.1 – Populations have spatial structure and many models include spatial dynamics

Both real-world observations and modelling experiments indicate that the spatial structure of plant and parasite populations can have an important effect on gene dynamics. Burdon & Thrall (1999) argued that events such as local extinction and re-colonisation, gene-flow between sub-populations and the extremely rapid changes of allele frequency that occur in very small sub-populations can have important effects on the total population. In this case, biologically realistic behaviour is best modelled within a meta-population framework.

Pathosystems that have been studied as metapopulations and in which significant spatial structuring has been observed include wild flax *Linum marginale* with rust fungus *Melampsora lini* (Thrall & Burdon 2000, Burdon et al. 2002) and *Plantago lanceolata* with *Podosphaera plantaginis* (Laine 2005, Laine & Hanski 2006). A recent review (Tack et al. 2012) of animal and plant pathosystems found variation in genotype frequencies between demes in metapopulations was significant in 28 of 29 studies. Coevolutionary models featuring spatial structure include Gandon et al. (1996), Damgaard (1999), Thrall & Burdon (2002), Sasaki et al. (2002), Gandon (2002), Gandon & Michalakis (2002), Nuismer (2006), Gandon & Nuismer (2009), Tellier & Brown (2011) and Blanquart et al. (2012).

Reported stability in spatial models can be broken down into three types.

Firstly, while numerous papers have reported stability in spatial models, such models are inherently difficult to analyse and can take a long time to reach fixation even when unstable. Brown & Tellier (2011) described such systems as self-slowing rather than self-stabilising (Section 1.3.2).

Secondly, some spatial models include stabilising factors that could have been modelled in a non-spatial context (Section 1.3.3).

Finally, a specifically spatial source of stability is oscillation damping caused by linkage between two or more demes with different environments and thus different intrinsic oscillation periods (Section 1.3.4).

1.3.2 – Spatial models can be self-slowing, exhibiting prolonged transient polymorphism, rather than truly stable. Spatial models are inherently difficult to analyse and many include mutation as a source of ndFDS, making distinguishing the two behaviours difficult

Brown & Tellier (2011) described self-slowing behaviour as anything which makes an unstable model or system take a long time to reach fixation but does not actually make the system stable. Spatial models tend to generate self-slowing behaviour because separate demes are partially isolated and it takes time for genotype frequency changes in one deme to affect the rest of the system. Spatial models are often difficult to analyse mathematically, so distinguishing self-slowing from stable behaviour is difficult. However some spatial models are better-analysed, for example the simplified two-patch metapopulation model in Tellier & Brown (2011).

Many spatial models also include mutation as a source of ndFDS, leading to weak ndFDS that is difficult to interpret (Section 1.2.3). This weak ndFDS combine with self-slowing behaviour such as limited dispersal and lack of analytical tractability to confuse the issue further. In particular, many modelling studies that varied dispersal distances and featured low mutation rates reported limited dispersal by itself can lead to stability. Spatial models including both mutation and likely self-slowing factors are described in the next three paragraphs.

Thrall & Burdon (2002) described a series of simulations on a 100x100 grid of virtual patches using a five-locus GFG system. They modelled immigration, local extinction and re-colonisation. Mutation was the source of new genotypes. They reported significant between-patch variation, depending on dispersal distances, together with maintenance of numerous genotypes for each. This was achieved without costs for resistance or virulence.

Sasaki (2002) reported that in a large group of modelled host-parasite populations linked by limited migration across a range of dispersal distances and rates, "pacemaker islands" spontaneously emerge which interact with the other populations and cause stable limit cycles. This is different from the oscillation damping discussed below (Tellier & Brown 2011) because it does not rely on environmental heterogeneity. In situations with more than one pair of loci, the environment breaks up into smaller patches, each one with its own pacemaker and oscillating between two opposite patterns of alleles in each species.

A review by Briggs & Hoopes (2004) reported that spatial systems can exhibit statistical stability, in which different demes are at different points in their coevolutionary cycles and overall global genotypes do not change. Tellier & Brown (2011) tested this and found it to be another case of self-slowing rather than true stability. Briggs & Hoopes also reported that stability could result from immigration uncoupled from local dynamics, which Tellier & Brown (2011) found could be a source of stability if it represented environmental heterogeneity (Section 1.3.4), and non-linear averaging, which requires extremely specific conditions and may be a modelling artefact.

Overall, because spatial models often include self-slowing factors and are not amenable to stability analysis, distinguishing truly stable from unstable systems and identifying the factors that generate stability can be difficult. Ultimately spatial models, like non-spatial models, need a biologically realistic source of ndFDS. Such sources could be simple factors that could be modelled in a non-spatial context (Section 1.3.3) or effects of the spatial structure, most obviously oscillation damping owing to gene-transfer between different environments (Section 1.3.4).

Mutation also appears in a related area, namely the models described in Gandon et al. (1996), Gandon (2002), Gandon & Michalakis (2002), Nuismer (2006) and Nuismer & Gandon (2009). These models all focussed on local adaptation of parasites or hosts, rather

than long-term stability or instability. To facilitate this, their models included mutation to create and maintain genetic variation. Their papers did not discuss long-term stability and their measures of local adaptation were not impeded by the difficulty of analysing stability in spatial models. Gandon & Michalakis (2002) reported that higher migration and mutation rates both favoured local adaptation, whereas shorter generation times did not. Gandon (2002) expanded on this by reporting that higher specificity and virulence both increased the advantage of the coevolutionary partner already ahead in local adaptation. Nuismer (2006) showed that spatial mosaics of selection coefficients can change both the direction and magnitude of local adaptation. Gandon & Nuismer (2009) showed that genetic drift can increase parasite local adaptation.

1.3.3 – Spatial models can incorporate essentially non-spatial sources of ndFDS

Damgaard (1999) reported stable polymorphism in a pathosystem of 100 demes linked by migration, where individual demes had a chance of extinction based on their average fitness. The biological justification for this group-selection was that unfit populations may be ravaged by secondary parasites against which only partial resistance exists. However the epidemiology of the secondary disease, which inevitably acts at the level of individual hosts, was not modelled explicitly. Variation in the parasite increased with deme extinction rate and genetic drift but decreased with the cost of infection and migration rate, as a higher migration rate makes the system behave more like a single panmictic population. When using a random extinction method or a single population, virulence was almost fixed at one and resistance had an intermediate value (not zero, because this model did not assume costs of fitness or resistance). Thus group-selection on heavily-parasitized patches, leading to ndFDS on avirulent parasites, led to stability in this model. Like most spatial models prior to Tellier & Brown (2007), Damgaard (1999) did not analyse results in terms of ndFDS at the time.

Another example of an essentially non-spatial source of ndFDS in a spatial model is Frank (1993), who reported that density-dependent regulation on infecting spores allowed polymorphism. My work in Chapters 2 and 3 demonstrates that such regulation is a stabilising factor, in non-spatial and spatial models respectively.

All of the above are essentially non-spatial factors, in that they can be expressed in non-spatial models as well as spatial models. The only specific factor known to generate

ndFDS and to be absolutely reliant on spatial structure is oscillation damping between heterogeneous environments, discussed below.

1.3.4 – Heterogeneous spatial models can generate ndFDS by damping between linked demes with different periods of oscillation

Tellier & Brown (2011) reported stability in models consisting of linked, heterogeneous demes. Demes varied in costs of resistance, virulence or infection and stability arose provided differences between demes were sufficiently pronounced. Biologically, differences in costs could be caused by different environments. Different environments and thus costs can lead to different periods of oscillations in genotype. Linking environments with different periods of oscillation can lead to damping in oscillations in all environments (Barr-Eli 1993). This result supported Briggs (2004), who noted that spatial or temporal variability combined with limited dispersal of parasites, hosts or both can lead to stable equilibriums.

1.3.5 – Summary

Overall, spatial models appear more likely to generate stable polymorphism than non-spatial models. However, the difficulties in analysing spatial models and the frequent inclusion of mutation in these models make recognising genuine stability difficult. It appears that many spatial models are self-slowing, where initial heterogeneity in genotype distribution and limited dispersal favours oscillation cycles that take a long time to spiral outwards, rather than truly self-stabilising (Brown & Tellier 2011).

The two classes of spatial models that are truly stable are those that include essentially non-spatial sources of ndFDS and those that generate ndFDS through environmental heterogeneity. In the former case ndFDS can arise as an artefact of how models handle mutation (Thrall & Burdon 2002, Sasaki 2002) or, more usefully, be caused by biologically plausible factors (Damgaard 1999, Franks 1993). The latter case is exemplified by the work of Tellier & Brown (2011).

1.4 – RESULTS IN THIS THESIS

In this thesis I aimed to introduce more realistic demography to the theory of host-parasite coevolution, supplementing Tellier & Brown's mathematically rigorous ndFDS framework with more varied ecological and epidemiological detail. My starting point was the observation that disease incidence is neither total nor constant, in both wild (Laine 2005, Burdon 1999) and agricultural (Workneh 1999) pathosystems. This led to Models 1 and 2 which include temporal and spatial variation respectively in disease incidence. Similarly, the wide-spread recognition that spatial processes effect coevolutionary dynamics (Laine & Hanski 2006) led to Model 3 and its focus on limited dispersal distances. Most previous spatial host-parasite coevolution models used metapopulation frameworks (Section 1.3). I used an individual-based stochastic system, with small populations, a single GFG-interaction and no immigration or mutation, to simplify analysis. Models 4 and 5 added the complicating biological factors of variable host density and a spatially heterogeneous second parasite, respectively, to the spatial and individual-based Model 3.

This thesis contains three results chapters.

Chapter 2 describes a deterministic population-based model (Model 1) similar to those in Tellier & Brown (2007A) with the alteration that parasite incidence is less than total and can vary between generations in a manner based on the logistic growth equation (Murray 1989). This variation is analogous to ndFDS on the two parasite genotypes and stabilises coevolution across a range of parameters. For any given set of costs there is a range of parasite basic reproductive rates that allow stable polymorphism. With realistic costs this occurs when parasite basic reproductive rate is at a level that results in low to intermediate disease incidence. In a related model (Model 2) there are two demes which differ only in their fixed but different disease incidences. This is variation of Tellier & Brown's (2011) model of demes with different environmental parameters (hence different costs) that are linked by gene flow. As in Tellier & Brown's model, stability occurs across a wide range of rates of gene flow, provided conditions within the demes are sufficiently dissimilar. I explain this in terms of amplitude death due to dissimilar oscillation periods in the two demes, which is a stabilising factor analogous to ndFDS (Bar-Eli 1985, Aronson 1990). Models 1 and 2 show that variable disease incidence in time and space, respectively, can lead to ndFDS or analogous stabilising factors and thus to stable polymorphism. More

broadly, they suggest that the number of ecological and epidemiological factors generating ndFDS is large and it is likely several such factors act on any given pathosystem.

Chapter 3 describes a spatially explicit, individual-based model (Model 3) in which variable parasite incidence is the stabilising factor in Model 1. The effects of spatial structure caused by limited dispersal distances were investigated, using larger populations to study how equilibrium genotype frequencies vary and smaller populations to study how average time to stochastic genotype loss varies with varying parasite and host dispersal. Limited dispersal leads to altered equilibrium genotype frequencies and longer, lower-amplitude oscillations in frequencies, decreasing the chances of stochastic genotype loss. I propose that in finite real-world populations, mechanisms which oppose stochastic genotype loss frequently occur in conjunction with ndFDS-causing mechanisms which oppose deterministic genotype loss. Together these two classes of mechanism make genotype loss in real, stochastically fluctuating populations less likely.

Chapter 4 describes variations of Model 3 featuring variable host incidence (Model 4) and a second parasite (Model 5). The results can be interpreted in the context of ndFDS and stochastic genotype loss developed in Chapters 2 and 3.

Chapter 5 discusses how hypotheses from Models 1-5 could be tested in biological systems and the advantages of experimental coevolutionary microcosms, particularly bacteria and phage owing to their short generation times and interesting genetics, for performing such tests. It also discusses the possibility of increasing density and connectivity in crops driving abrupt switches from wild-type trench-warfare to agricultural-type arms-race dynamics.

CHAPTER 2 – VARIABLE DISEASE INCIDENCE IN TWO DETERMINISTIC MODELS

OVERVIEW

Polymorphism at corresponding host-immunity and parasite-antigenicity loci is common and can persist for millions of years. The gene-for-gene (GFG) model of immune interaction between plants and their parasites is a well-understood case of this coevolutionary relationship and has been extensively studied and modelled. Maintaining polymorphism in interacting host and parasite genes requires a source of negative direct frequency-dependent selection (ndFDS). This selection means that as a genotype grows more common its relative fitness decreases. Many ecological processes can cause such selection.

Models of GFG interactions investigating stable polymorphism have mostly assumed complete disease incidence, defined here as the proportion of a given plant population exposed to a given parasite species. However in reality incidence is known to vary greatly in both wild and agricultural plant-parasite associations. Here I present two models simulating less than total disease incidence. The first is a single-patch model in which incidence can vary over time in a density- and fitness-dependent manner. The second model uses two patches linked by gene-flow, with fixed but different disease incidences in each patch.

Both models can lead to stable polymorphism at resistance/susceptibility loci in plants and virulence/avirulence loci in parasites. In the first model variable disease incidence causes virulent and avirulent parasites to act as partially independent populations, both subject to density-dependent self-regulation. Density-dependent regulation on the avirulent parasite stabilises coevolution, an effect analogous to negative direct frequency-dependent selection. In the second model each patch shows oscillations in gene frequency with a specific period. Linking the patches can result in oscillation damping and so lead to stable

polymorphisms. These mechanisms are not mutually exclusive and individually or together they can be used to describe various real-world situations.

2.1 – INTRODUCTION

2.1.1 – Polymorphism in host immunity genes and corresponding parasite avirulence genes is common, can persist for millions of years and is of both theoretical interest and epidemiological and agricultural importance

Resistance to parasites is a target for natural selection in virtually all species. Similarly, the ability to overcome such resistance is a target for natural selection in parasites. As discussed in Chapter 1, there is allelic diversity in host genes controlling parasite recognition, e.g. the MHC in vertebrates (Apanius et al. 1997, Hill 2001) and the gene-for-gene system in plants (Stahl et al. 1999, Holub 2001, Tian et al. 2003). Similarly there is allelic diversity in the parasite proteins detected by host defence mechanisms. Coevolutionary dynamics between the two species maintain or fail to maintain this polymorphism.

Known interactions between host recognition and parasite avoidance genes include the MHC system in mammals and the gene-for-gene (GFG) system in plants. In this thesis I focus on the latter because of its well-understood genetics, its ubiquity in the plant kingdom and its importance to agriculture. The GFG theory has been extant for more than fifty years (Flor 1955, 1971, Thompson & Burdon 1992). It states that a GFG relationship exists between a plant species and a parasite species if the plant has one or more resistance/susceptibility (*RES*/*res*) loci, each corresponding to a single virulence/avirulence (*avr*/*AVR*) locus in the parasite. If one or more plant *RES*-alleles match specific parasite *AVR*-alleles, the plant will detect the infection and resistance will be induced, otherwise the infection will not be detected and will proceed.

2.1.2 – stable polymorphism requires both iFDS and ndFDS

Long-term polymorphism of *RES*/*res* and *avr*/*AVR* genotypes requires both indirect and negative direct frequency-dependent selection, iFDS and ndFDS, as discussed in Chapter 1 (Section 1.1.5-1.1.7). To recap, iFDS means relative genotype fitnesses in one species

depend on relative genotype frequencies in the other species. If there are intrinsic costs of resistance and virulence, such that *res*-hosts out-compete *RES*-hosts when *AVR*-parasites are scarce and *AVR*-parasites out-compete *avr*-parasites when *RES*-hosts are scarce, this leads to cycling around the system's internal equilibrium point (IEP).

However, iFDS alone does not lead to stable polymorphism. Over time the oscillations diverge and genotypes go extinct (Leonard 1977, 1994). As well as iFDS, stable polymorphism requires negative direct frequency-dependent selection (ndFDS) - the net fitness of one or more of the genotypes involved in the host-parasite system must decline as its own frequency increases, regardless of genotype frequencies in the other species (Tellier & Brown 2007A). This ndFDS is essential for stable polymorphism and its existence and strength relative to other factors determine whether a given system is stable or not. Diverse ecological and epidemiological factors can generate ndFDS, for example repeated cycles of auto-infection of diseased hosts within one season (Tellier & Brown 2007A) and perenniality or seed banks in hosts (Tellier & Brown 2009). Linking patches with dissimilar periods of gene-frequency oscillation can dampen oscillations in both patches. This stabilises coevolution, an effect analogous to ndFDS (Tellier & Brown 2011).

2.1.3 – Variable disease incidence has been ignored in most modelling studies of GFG coevolution but is capable of stabilising polymorphism

The studies cited above all assume uniform (total) disease prevalence. Real population densities vary over time and between environments as a result of numerous biotic and abiotic factors. Species with linked population dynamics often experience linked, regular oscillations in density. This occurs in two-species interactions including predator-prey cycles, such as the snow-shoe hare and lynx (reviewed in Krebs et al. 2001), herbivore-food and host-parasite systems. Laine (2005) and Burdon & Thrall (1999) are examples of parasite density fluctuations in nature, while Workneh et al. (1999) is an example from an agricultural parasite. Parasite density can be measured as disease incidence, the fraction of hosts infected, rather than actual spatial density of the parasite. All organisms experience reduced population growth as the population nears carrying capacity and parasites at high incidence are no exception. I term this density-limited reproduction, or DLR.

In host-parasite systems, DLR can stabilise otherwise unstable genotype polymorphisms (May & Anderson 1983). The simplifying assumption of full incidence has been used in many GFG models (Groth 1976, Leonard 1977, Damgaard 1999, Sasaki 2000, Sasaki et al. 2002, Segarra 2005, Tellier & Brown 2007A, 2009, 2011). Variable incidence had been modelled in both GFG (Thrall & Burdon 2005, Salathe et al. 2005) and non-GFG systems (Frank 1993, Gandon 2002), but these models include several ecological and epidemiological factors and the conditions under which variable incidence alone can stabilise balanced polymorphism are unclear.

In this chapter I demonstrate that temporal or spatial variation in disease incidence, or parasite density, can stabilise balanced polymorphism in model populations of GFG parasites and hosts. In a single population where variable disease incidence is regulated by density-limited reproduction (DLR), the fitness of both parasite genotypes decreases as the total incidence of parasite increases. This negative, density-dependent regulation is similar to ndFDS on the fitness of both parasite genotypes. Like ndFDS, it is capable of allowing long-term stable polymorphism. I solve analytically to find the combinations of costs and parasite growth-rates that allow stability. In a model in which the population is sub-divided with fixed but different levels of disease in each deme, gene-flow between demes dampens oscillations and leads to stability. Both these mechanisms have wide biological applicability for plant-parasite interactions.

2.2 - MODELS AND METHODS

2.2.1 – Disease incidence can vary over time or over space, modelled separately. Both models use recurrence equations to track genotype frequencies

If disease incidence is less than total, what factors affect it? Incidence could depend on parasite population dynamics or on environmental conditions (or more realistically on both). The former assumes the number and fitness of parasite "colonies" (all the parasites on one host, modelled as one parasite individual) in one generation determine the number of colonies in the next generation. The latter assumes parasite density is controlled by environmental factors and is thus roughly constant and independent of density in earlier generations. These cases are modelled separately. Model 1 is a non-spatial model in which disease incidence varies over generations as mean parasite fitness changes. Mean

parasite fitness is based on genotype frequencies and costs *b* and *c* (Table 1). Disease incidence changes each generation according to a standard logistic growth term (Murray 1989) where basic reproductive rate is scaled by this mean parasite fitness. Model 2 is a spatial model in which two demes have fixed but different disease incidences and there is gene-flow between them. Each patch experiences selection on genotypes according to its own internal conditions before final frequencies are affected by the migration of genetic material between patches. Thus Model 1 describes disease incidence varying in time and Model 2 describes it varying in space. In both cases stable polymorphism of *RES*/*res* and *av*r/*AVR* genes can result.

I make simplifying assumptions, as is standard in population genetics models. Plant populations are assumed to have infinite size and constant density. As only one *RES*-locus and one corresponding *AVR*-locus are modelled, species are assumed to be haploid and asexual. The frequency of mutation is assumed to be negligible (zero). Hosts and parasites are assumed to have simultaneous, discrete generations and parasite spores infect hosts randomly. There is no horizontal transmission of parasites and no effect of spatial structure other than the gene flow modelled explicitly in Model 2. Each plant encounters at most one parasite spore, which always infects successfully unless it is an *AVR*-spore on a *RES*-plant. Finally all individual hosts and successful parasite infections are assumed to be the same size, i.e. all successful infections inflict the same cost on and gain the same resources from their hosts and there is no variation in reproductive success of either species apart from that caused by GFG fitness costs.

Tables 2.1 and 2.2 are on the next three pages. Table 2.1 defines the variables (part 1) and constants (part 2) used in Models 1 and 2. Table 2.2 gives the fitnesses of genotypes of each species in each possible interaction. These fitnesses are the same in Models 1 and 2.

Variable	Description
	Models 1 & 2
R	frequency of resistance allele (*RES*) in plant population in current generation
r	frequency of susceptibility allele (*res*) in plant population ($R + r = 1$)
A	frequency of avirulence allele (*AVR*) in parasite population in current generation
a	frequency of virulence allele (*avr*) in parasite population ($A + a = 1$)
$\widehat{\boldsymbol{R}}$, $\widehat{\boldsymbol{A}}$…	equilibrium values of R, A etc.
$\boldsymbol{\alpha}$	log of R/r
$\boldsymbol{\rho}$	log of A/a
$\Delta\boldsymbol{\alpha}$, $\Delta\boldsymbol{\rho}$	change in $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ between the current and next generations
$\overline{\boldsymbol{W_h}}$, $\overline{\boldsymbol{W_p}}$	mean fitnesses of host and parasite populations
	Model 1 only
M	variable disease incidence (the fraction of plants encountering the parasite)
$\boldsymbol{\mu}$	log of M
$\Delta\boldsymbol{\mu}$	change in $\boldsymbol{\mu}$ between the current and next generations
R', A'…	values of R, A etc. in the next generation
	Model 2 only
$\boldsymbol{R_i}$, $\boldsymbol{A_i}$…	genotype frequencies of *RES*-hosts, *AVR*-parasites etc. in deme i
$\boldsymbol{R_i}^{\#}$, $\boldsymbol{A_i}^{\#}$…	values of $\boldsymbol{R_i}$, $\boldsymbol{A_i}$ etc. in the next generation after selection within demes but before gene flow between demes
$\boldsymbol{R_i}'$, $\boldsymbol{A_i}'$...	values of $\boldsymbol{R_i}$, $\boldsymbol{A_i}$ etc. in the next generation after both selection within demes and gene-flow between demes

Table 2.1 part 1 – Variables in Models 1 and 2. Table 2.1 part 2, model constants, is overleaf.

Constant	Description
	Models 1 and 2
b	fitness cost to the parasite of having the *avr*-allele
c	fitness cost to an *AVR*-parasite of a failed attempt to infect a *RES*-plant
u	fitness cost to the plant of having the *RES*-allele
s	fitness cost to the plant of being diseased -not paid by *RES*-hosts with *AVR*-parasites
	Model 1 only
y	parasite's intrinsic rate of increase
	Model 2 only
m_i	fixed disease incidence in population i (i = 1 or 2)
h_{ij}	fraction of seeds transferred from population i to population j each generation (i = 1 or 2; j = the other)
g_{ij}	fraction of spores transferred from population i to population j each generation

Table 2.1 part 2 – Constants in Models 1 and 2. Costs are as follows. b is the cost to the parasite of possessing the *avr*-allele and represents the *avr*-genotype's reduction in growth and reproduction relative to *AVR*-genotype colonies on susceptible hosts. Biologically this is due to the reduced efficiency of the protein coded by *avr*, which is a side-effect of its being less detectable. c is the cost to *AVR*-genotype parasites of being detected by *RES*-genotype plants. In nature this cost is very often 1, or total failure of the parasite to survive the host defence. u is the cost to hosts of possessing the *RES*-allele and represents the *RES*-genotype's reduced basic rate of reproduction relative to *res*-genotypes in the absence of *AVR*-parasites. Biologically it is unclear why this occurs and if it is ubiquitous (Chapter 1, Sections 1.1.8-1.1.9). s is the cost to plants of being successfully infected. s is paid by all plants infected with *avr*-parasites and by *res*-plants infected with *AVR*-parasites but not by *RES*-plants infected with *AVR*-parasites because the GFG relationship means such infections are detected and stopped before they do appreciable damage.

	***RES*-host *AVR*-parasite**	***RES*-host *avr*-parasite**	***RES*-host no parasite**	***res*-host *AVR*-parasite**	***res*-host *avr*-parasite**	***res*-host no parasite**
Host fitness	1 - *u*	(1 - *u*)(1 - *s*)	1 - *u*	1 - *s*	1 – *s*	1
Parasite fitness	1 - *c*	1 - *b*	-	1	1 – *b*	-

Table 2.2 – Fitness outcomes for each genotype of both species for all possible interactions. In Model 1 each genotype is modelled as a population with a single averaged fitness value, while in Model 2 average fitnesses are calculated for each genotype within each deme. Average fitnesses are calculated each generation based on the fitness values for and relative frequencies of each interacting genotype. For example, the average fitness of *RES* hosts at a given generation is the summed weighted fitnesses of *RES*-hosts with no parasite, *RES*-hosts with *AVR*-parasites and *RES*-hosts with *avr*-parasites.

Recurrence equations are used to model between-generation changes in *R*, *r*, *A* and *a*. While the two genotypes in a species have different fitnesses the ratio between them changes every generation. As the relative fitness of genotypes in one species depends on the frequency of genotypes in the other species, the changes in frequency are linked and lead to a cycling of frequencies over time. Similarly disease incidence *M* in Model 1 is modelled by a recurrence equation and varies over time in a way dependent on average parasite fitness, which is linked to frequencies of genotypes in both species.

2.2.2 - Model 1 is a single-patch non-spatial model that simulates changes over generations in disease incidence *M*, as well as genotype frequencies *R* and *A*

Incidence changes each generation according to Eqn. 2.3, making the equilibrium value $\hat{M}$ dependent on the parasite's intrinsic rate of increase *y* and the average parasite fitness $\overline{W_p}$. Simultaneously the ratios *R*/*r* and *A*/*a* change each generation according to Eqn.s 2.1 and 2.2, making equilibrium values $\hat{R}$ and $\hat{A}$ dependant on relative costs. The recurrence equations for the three variables, representing changes in gene-frequency and incidence from one generation to the next, are as follows.

For parasite *A*/*a* ratio

$$A' = \frac{A(1 - Rc)}{\overline{W_p}}$$

$$a' = \frac{a(1-b)}{\overline{W_p}}$$

Eqn. 2.1

$$\frac{A'}{a'} = \frac{A}{a} \cdot \frac{1-Rc}{1-b}$$

For plant *R/r* ratio

$$R' = \frac{R(1-u)(1-saM)}{\overline{W_h}}$$

$$r' = \frac{r(1-sM)}{\overline{W_h}}$$

Eqn. 2.2

$$\frac{R'}{r'} = \frac{R}{r} \cdot \frac{(1-u)(1-saM)}{1-sM}$$

For parasite incidence *M*

Eqn. 2.3

$$M' = My\overline{W_p}(1-M)$$

To simplify analysis expressions are translated into log form using the variables α (the log of the ratio *A*:*a*), ρ (the log of the ratio *R*:*r*) and μ (the log of *M*)

$$\alpha' = \alpha + \log(1-Rc\,) - \log(1-b)$$

$$\rho' = \rho + \log(1-u) + \log(1-saM) - \log(1-sM)$$

$$\mu' = \mu + \log y + \log \overline{W_p} + \log(1-M)$$

Log form expressions are then rephrased in terms of the differences Δ in α, ρ and μ in the next generation

Eqn. 2.1a

$$\Delta\alpha = \log(1 - Rc) - \log(1 - b)$$

Eqn. 2.2a

$$\Delta\rho = \log(1 - u) + \log(1 - saM) - \log(1 - sM)$$

Eqn. 2.3a

$$\Delta\mu = \log y + \log \overline{W_p} + \log(1 - M)$$

Mean host and parasite fitnesses $\overline{W_h}$ and $\overline{W_p}$ are the weighted means of the fitness of each genotype (Table 2) and have the values

$$\overline{W_h} = R(1 - u)(1 - saM) + r(1 - sM)$$

$$\overline{W_p} = a(1 - b) + A(1 - Rc)$$

The equations for $\Delta\alpha$ and $\Delta\rho$ follow Tellier & Brown (2007A), adjusted for variable parasite incidence *M*. The equation for $\Delta\mu$ is the logistic growth equation (Murray 1989), which models density-dependent population growth by multiplying the current density *M* by the mean growth rate $y\overline{W_p}$ and a density-dependent term that reduces growth as a limiting density is approached (1-*M*). This last term represents the increasing difficulty for new spores to establish as hosts are saturated with disease. This is density limited reproduction (DLR) and is the source of stability in Model 1. DLR is equivalent to ndFDS, save that it acts on the total incidence of the parasite population rather than on a genotype frequency within a population.

2.2.3 - Model 2 - two patches with different disease incidences, linked by gene-flow

Model 2 investigates the effects of fixed, different levels of disease incidence in two patches linked by migration of plant seeds and parasite spores. Fixed incidence removes the intrinsic stabilising DLR, so each patch is unstable by itself. However it has been demonstrated (Tellier & Brown 2011) that two or more patches that have no internal

source of ndFDS (or analogous DLR) and are thus inherently unstable can stabilise each other by small amounts of gene-exchange, if the environments differ so that one or more of the costs differs between patches. This is because the different costs lead to different periods for the gene-frequency oscillations in each patch. Linking systems with dissimilar oscillation periods can cause harmonic damping, leading to amplitude death of oscillations and stabilising both systems at points near their respective IEPs (Aronson et al. 1990, Barr-Eli 1985).

With linked patches, any difference in dynamics that leads to different frequencies can cause stability in this way. In Tellier & Brown (2011), it was assumed environmental conditions lead to different fitness costs in different patches. Here different environmental conditions affect not costs but rather disease incidence, which is the only factor to differ between patches, leading to different periods and thus oscillation damping and stability.

Each generation is modelled in two stages. This is a selection-reproduction-migration system. Selection and reproduction are modelled as one stage because as only one gene is involved explicit modelling of sexual reproduction is not required. Migration is modelled as the second stage.

Each patch first makes spores and seeds independently. Within each patch the genotypes experience selective pressure and their frequencies change according to their relative fitnesses, as in Model 1 but with fixed disease incidence m_i (Equations 2.4 and 2.5, below).

After fitness-dependent selection, genetic material moves between patches. The transfer parameters h_{ij} and g_{ij} measure the fraction of seeds and spores, respectively, moving from patch i to patch j (Equations 2.6 and 2.7, below).

The relative sizes of the patch 1 and patch 2 seed and spore populations are assumed to be proportional to the relative adult populations of the two demes multiplied by the reproductive fitnesses of each population($\overline{W_{pi}}$, $\overline{W_{pj}}$, $\overline{W_{hi}}$ and $\overline{W_{hj}}$). Host populations are the same for all results in this chapter, while parasite populations are proportional to m_i and m_j. When there is gene flow between patches, the relative contribution of genotype frequencies from each population is a function of these relative population sizes and the fractions of seeds or spores transferred ($h_{ij,}$ h_{ji}, g_{ij} and g_{ji}).

In Model 2, $R_i^{\#}$ is next generation's *RES*-frequency in patch *i* after selection and seed- and spore-production within the patch but before movement of seeds and spores between patches. R_i' is the frequency in the next generation after both selection and movement.

For plant *R*/*r* ratio and parasite *A*/*a* ratio in patch *i* before mixing

Eqn 2.4

$$\frac{R_i^{\#}}{r_i^{\#}} = \frac{R_i}{r_i}\frac{(1-u)(1-sa_im_i)}{1-sm_i}$$

Eqn 2.5

$$\frac{A_i^{\#}}{a_i^{\#}} = \frac{A_i}{a_i}\frac{1-R_ic}{1-b}$$

For plant *R(r)* frequency and parasite *A(a)* frequency in patch *i* after mixing

Eqn 2.6

$$R_i' = \frac{R_i^{\#}\overline{W_{hl}}(1-h_{ij}) + R_j^{\#}\overline{W_{hJ}}h_{ji}}{\overline{W_{hl}}(1-h_{ij}) + \overline{W_{hJ}}h_{ji}}$$

Eqn 2.7

$$A_i' = \frac{A_i^{\#}\overline{W_{pl}}(1-g_{ij}) + A_j^{\#}\overline{W_{pJ}}g_{ji}(m_j/m_i)}{\overline{W_{pl}}(1-g_{ij}) + \overline{W_{pJ}}g_{ji}(m_i/m_j)}$$

The fitness-dependent selection equations 2.4 and 2.5 are equivalent to Model 1 equations 2.1 and 2.2 respectively. Equation 2.6 states that, for hosts, the post-mixing genotype frequency R_i in patch *i* is based on the weighted contributions of post-selection, pre-mixing host genotype frequencies from both patch *i* (native material) and patch *j* (immigrating material). Weighting is based on the transfer parameters h_{ij} and h_{ji} and the relative average fitnesses of host populations, $\overline{W_{hl}}$ and $\overline{W_{hl}}$, as fitness determines how many seeds hosts produce. Similarly Equation 2.7 states that, for parasites, the post-mixing genotype frequency A_i in patch *i* is based on the weighted contributions of post-selection, pre-mixing parasite genotype frequencies from both patch *i* (native material) and patch *j* (immigrating material). Weighting is based on the transfer parameters g_{ij} and g_{ji}, the relative average

fitnesses of parasite populations $\overline{W_{pi}}$ and $\overline{W_{pj}}$ and the relative disease incidences m_i and m_j.

The equations for genotype frequencies *r* and *a* can be found by replacing *R* with *r* and *A* with *a* throughout equations 2.6 and 2.7 respectively. The equations for genotype frequencies in patch *j* are obtained from equations 2.6 and 2.7 by exchanging the subscripts *i* and *j* throughout.

2.3 - RESULTS

2.3.1 - Model 1 behaviour includes two types of stable polymorphism and is determined by parasite intrinsic growth rate *y* relative to costs

Model 1 systems exhibits one of five types of behaviour depending on the value of the parasite's intrinsic rate of reproduction *y* relative to the other constants. These are named Types 1 to 5 and defined in Table 2.3, below. They include two types of stable polymorphism. **Type 3 polymorphism** is the case where parasite incidence is low, *AVR* is fixed and *RES*/*res* polymorphism is maintained. **Type 4 polymorphism** is the case where parasite incidence is higher and there is stable, balanced polymorphism for both species. Both polymorphisms are maintained by DLR on the *AVR*-parasite population, which dampens oscillations in *MA* and thus in *R*.

Type and range of y	Behaviour	Equilibria
Type 1 $\mathbf{0 \le y \le 1}$	Parasite incidence declines to 0 Without parasites *RES*-hosts go extinct (cost of resistance u)	$\hat{A} = N/A$ $\hat{R} = 0$ $\hat{M} = 0$
Type 2 $\mathbf{1 < y \le \frac{s}{s-u}}$	Low incidence of *AVR*-parasites *AVR*-parasite incidence low enough that *RES*-hosts go extinct Without *RES*-hosts *avr*-parasites go extinct (cost of avr b) DLR stabilises *M* only	$\hat{A} = 1$ $\hat{R} = 0$ $\hat{M} = 1 - y^{-1}$
Type 3 $\mathbf{\frac{s}{s-u} < y \le \frac{s}{(s-u)(1-b)}}$	Low (but higher than Type 2) incidence of *AVR*-parasites *AVR*-parasite incidence higher so *RES*/res polymorphism *RES*-hosts rare enough that *avr*-parasites go extinct DLR sufficient to stabilise *RES*/*res* polymorphism	$\hat{A} = 1$ $\hat{R} = \frac{1 - \left(y(1 - us^{-1})\right)^{-1}}{c}$ $\hat{M} = us^{-1}$
Type 4 $\mathbf{\frac{s}{(s-u)(1-b)} < y \le J}$	Low to intermediate incidence of *AVR*-parasites *AVR*-parasite incidence higher so *RES/res* polymorphism (stable) *RES*-hosts more common so *avr*/*AVR* polymorphism (stable) DLR sufficient to stabilise system	$\hat{A} = \frac{u\left(1 - s\hat{M}\right)}{s\hat{M}(1-u)}$ $\hat{R} = \frac{b}{c}$ $\hat{M} = 1 - \frac{1}{y\overline{W_p}} = 1 - \frac{1}{y(1-b)}$
Type 5 $\mathbf{J < y}$	Intermediate to high incidence of *AVR*-parasites *AVR*-parasite incidence higher so *RES/res* polymorphism (unstable) *RES*-hosts more common so *avr*/*AVR* polymorphism (unstable) DLR insufficient to stabilise system	As Type 4

Table 2.3 – Possible behaviours of Model 1 according to values of y relative to costs. Distinguishing Type 4 and 5 behaviour requires Jacobian analysis, as the boundary value J is not solvable.

Type 1 and 2 behaviours are trivial. In Type 1 systems, the parasite basic reproductive rate is less than 1 so the parasite invariably declines to extinction. In Type 2 systems, the parasite persists but – as its own density limits its fitness – it persists at a level so low *RES*-hosts will always be out-competed by *res*-hosts. Types 3-5 all have IEPs with

polymorphism in one or both species and are discussed below. For this discussion I define *RES*-fit, *res*-fit, *avr*-fit and *AVR*-fit as the average fitnesses of each genotype.

Figure 2.1, below, shows genotype and disease incidence oscillations over time in Type 3-5 systems. In Type 3 systems average disease incidence $\widehat{M}$ is very low, average *RES*-frequency $\hat{R}$ is low, average *AVR*-frequency $\hat{A}$ goes to fixation and oscillations die down very rapidly. In Type 4 systems $\widehat{M}$ and $\hat{R}$ are higher, polymorphism persists in the parasite leading to intermediate $\hat{A}$ and oscillations take longer to die down. In Type 5 systems $\widehat{M}$ is higher, $\hat{R}$ is the same as in Type 4 systems, $\hat{A}$ is lower because $\widehat{M}\hat{A}$ is the same and $\widehat{M}$ is higher (Section 2.32) and oscillations diverge over time leading to fixation. Figure 2.2, overleaf, shows converging and diverging cycles in gene-frequency for a Type 4 system with a stable IEP and a Type 5 system with an unstable IEP.

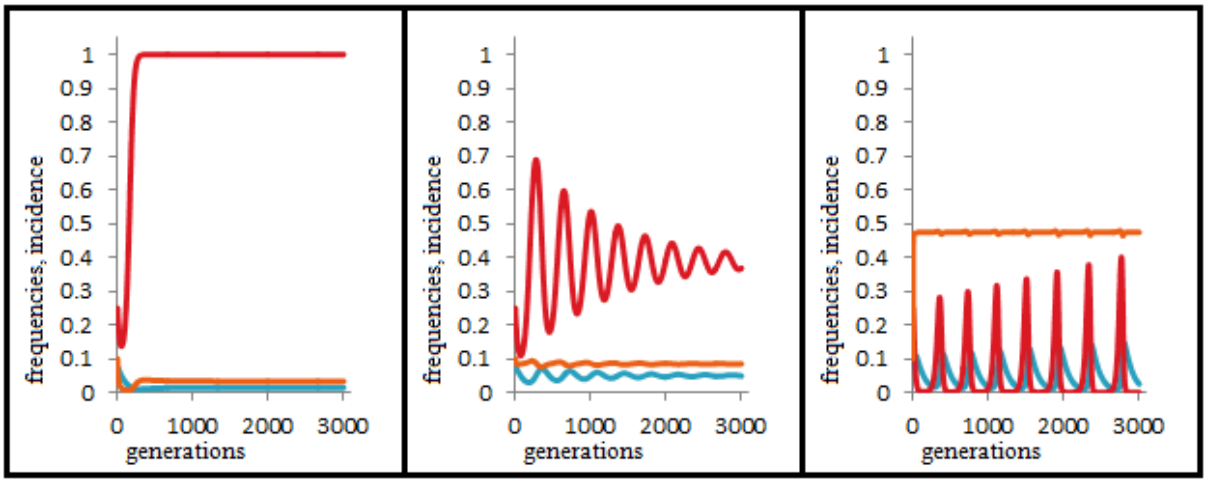

Figure 2.1 – Genotype frequency and disease incidence oscillations in Type 3, 4 and 5 systems. Type 3 system is panel A (y=1.05), Type 4 is panel B (y=1.15) and Type 5 is panel C (y=2). x-axis are generations and y-axis are genotype frequencies and disease incidence. *R* is blue, *A* is red and *M* is orange. Costs are u=0.01, s=0.3, b=0.05 and c=1. Oscillations are dampened almost immediately in the Type 3 system, dampened over thousands of generations in the Type 4 system and expand over time in the Type 5 system. Types 3 and 4 are stable polymorphism, Type 5 is unstable polymorphism.

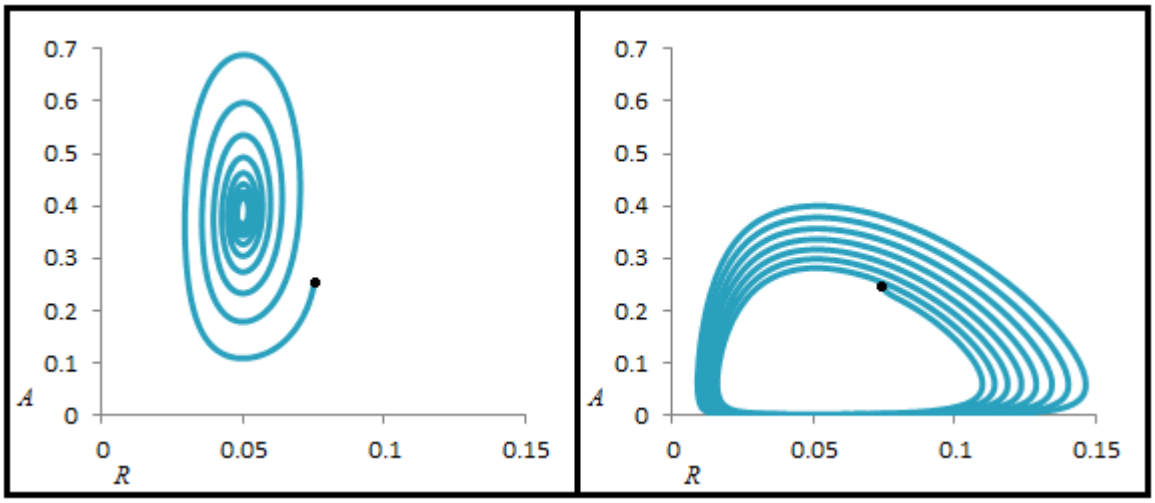

Figure 2.2 – Linked oscillations of *R* and *A* in stable (first panel, Type 4, *y*=1.15) and unstable (second panel, Type 5, *y*=2). X-axes are *R* and y-axes are *A*. Costs are *u*=0.01, *s*=0.3, *b*=0.05 and *c*=1. *y* is 1.15 in the first panel and 2 in the second panel. Type 4 systems spiral inwards towards an IEP, while Type 5 systems spiral outwards towards fixation. Initial frequencies are marked with black circles.

2.3.2 – Increasing parasite basic growth rate *y* increases average parasite incidence $\widehat{M}$, leading first to *RES*/res polymorphism and then to *avr*/*AVR* polymorphism

In Type 3 systems all parasites are *AVR* at equilibrium so $\widehat{M} = \widehat{M}\hat{A}$, i.e. total parasite incidence is equal to *AVR*-parasite incidence. *MA* and *R* regulate each other, dampening oscillations in each other and moving the system towards the Type 3 IEP, as follows. When $MA = \widehat{M}$, *RES*-fit = *res*-fit (eqn. 2.2, $Ms = u$ and $saM = 0$) so host genotypes have equal fitness and *R* doesn't change. If $MA > \widehat{M}$, *RES*-fit > res-fit (eqn. 1.2, $Ms > u$) and *R* will increase. Conversely if $MA < \widehat{M}$, *RES*-fit < res-fit (eqn. 2.2, $Ms < u$) and *R* will decrease. Similarly if $R = \hat{R}$ *MA* doesn't change, if $R > \hat{R}$ *MA* decreases and if $R < \hat{R}$ *MA* increases (eqn. 2.3, $W_p = 1 - Rc$). Together these dependencies mean any movement away from equilibrium in *R* (conversely *MA*) leads to *MA* (*R*) changing in a way that reverses the change in *R* (*MA*), resulting in a stable equilibrium at which there is polymorphism for *RES*/*res* and *AVR* is fixed. In Type 3 systems *avr*-parasites always go extinct. This is because the density-dependent reduction in parasite fitness, density-limited reproduction or DLR, is driven by and reduces the total incidence of both parasite genotypes. Thus if there are any *avr*-parasites at all this DLR will drive *MA* below $\widehat{M}$, lowering the fitness of *RES*-hosts and causing *R* to fall which in turn increases the fitness of *AVR*-parasites and causes *a* to decline. In Type 3 systems, *a* will not stop declining until it is extinct.

This equilibrium explains the Type 2/3 and Type 3/4 system boundaries. If $y < s(s-u)^{-1}$ *MA* cannot reach us^{-1} (eqn. 2.3) so *RES*-hosts will always be out-competed by *res*-hosts and go extinct (Type 2) (eqn. 2.2). Conversely, if $y > s(s-u)^{-1}(1-b)^{-1}$ $\widehat{M}$ will reach a higher level than us^{-1} even when $R = bc^{-1}$ (eqn. 2.3). As $MA > us^{-1}$ triggers an increase in *R* that reduces *MA* (eqn. 2.1) and $R > bc^{-1}$ triggers a decrease in *MA* that reduces *R* (eqn. 2.2), this increase in $\widehat{M}$ means *avr*-parasites are present at the IEP ($\widehat{M}\hat{a} = \widehat{M} - \widehat{M}\hat{A}$) and the system is Type 4 or Type 5.

2.3.3 - Stability of Type 4/5 systems must be determined by a Jacobian matrix

Type 4 and 5 systems have the same equilibria, so there is no simple boundary term as there is for Type 2/3 and Type 3/4 boundaries. For a set of recurrence equations with an IEP, the stability of the IEP can be determined by constructing a Jacobian matrix and finding the matrix's eigenvalues (Roughgarden 1979). A Jacobian matrix is an *n* by *n* matrix of all the recurrence equations partially differentiated by all the variables. For Model 1 this is a matrix of the equations for $\Delta\alpha$, $\Delta\rho$ and $\Delta\mu$ (Equations 2.1a-2.3a) partially differentiated by α, ρ and μ (Figure 2.3, below).

$$\begin{bmatrix} \frac{d\Delta\alpha}{d\alpha} & \frac{d\Delta\rho}{d\alpha} & \frac{d\Delta\mu}{d\alpha} \\ \frac{d\Delta\alpha}{d\rho} & \frac{d\Delta\rho}{d\rho} & \frac{d\Delta\mu}{d\rho} \\ \frac{d\Delta\alpha}{d\mu} & \frac{d\Delta\rho}{d\mu} & \frac{d\Delta\mu}{d\mu} \end{bmatrix} = \begin{bmatrix} 0 & \frac{\hat{A}(1-\hat{A})s\widehat{M}}{1-s\widehat{M}+s\hat{A}\widehat{M}} & \frac{\hat{A}(1-\hat{A})(b-\hat{R}c)}{1-b+\hat{A}b-\hat{A}\hat{R}c} \\ \hat{R}(1-\hat{R})\frac{-c}{1-\hat{R}c} & 0 & \frac{\hat{R}(1-\hat{R})(-\hat{A}c)}{1-b+\hat{A}b-\hat{A}\hat{R}c} \\ 0 & \frac{s\hat{A}\widehat{M}}{(1-s\hat{A}\widehat{M})(1-s\widehat{M})} & \frac{-\widehat{M}}{1-\widehat{M}} \end{bmatrix}$$

Figure 2.3: Jacobian matrix for the equation set in Model 1

Variables are taken at their equilibrium value for the Type 4/5 IEP to give a matrix of simple numbers. A necessary but not sufficient condition for stability is that the trace, which is the sum of the matrix's diagonal elements or each recurrence function differentiated with regard to its own variable, must be negative. This is true for all Type 4/5 systems owing to the $\frac{d\Delta\mu}{d\mu}$ partial differential. This represents the DLR on disease incidence *M* due to the $(1-M)$ term. For the system to be stable (Type 4) the eigenvalues of the matrix must lie within a circle of radius 1 on the complex plane, otherwise the system is unstable (Type 5). Higher values of *y* lead to eigenvalues outside this range.

2.3.4 – Gene-flow between patches can stabilise polymorphism in Model 2

Without inter-patch gene-flow, the patches in Model 2 are unstable. An isolated patch with constant disease incidence has no source of stability (ndFDS or DLR) in this model. Polymorphism in such a patch is unstable assuming $m \neq us^{-1}$. If $m < us^{-1}$ low incidence causes *AVR*-parasites and *res*-hosts to become fixed, similar to Type 2 systems in Model 1. If $m > us^{-1}$ there are diverging, unstable oscillations of gene frequencies, similar to Type 5 systems in Model 1 (data not shown).

If $m = us^{-1}$ exactly there is stable polymorphism for *RES*/*res* while *AVR* is fixed, similar to Type 3 systems in Model 1. However, this result is not biologically interesting because it seems unlikely that external factors would hold disease incidence to exactly this level.

2.3.5 - One-directional gene-flow can cause stability in the receiving patch in the specific case $m \leq us^{-1}$ in the source-patch and $m > us^{-1}$ in the sink-patch

One-directional gene-flow from a population with $m \leq us^{-1}$(where $\hat{R} = 0$ and $\hat{A} = 1$) can cause stable polymorphism for both *RES*/*res* and *avr*/*AVR* in a second population with $m > us^{-1}$. The influx of genes from the first population dampens genotype oscillations in the second population. This result is biologically interesting because it could apply to real-world situations, specifically long-range dispersal of fungal spores, pollen and wind-dispersed seeds. At large scales prevailing wind directions could overwhelm local fluctuations in wind direction and make seed or spore transfer one-directional. At large scales it is also more likely populations would have radically different environments and hence disease incidences. Therefore, one-directional transfer of seeds or spores from a lightly infected or disease-free population ($m \leq us^{-1}$) to a more heavily infected population ($m > us^{-1}$) could be a very common situation.

If there are multiple host sub-populations some will generally be disease-free at any given time (Laine 2005). Thus seeds or pollen travelling from disease-free to diseased populations could be a common occurrence at smaller spatial scales, where the difference in disease levels is due to stochasticity rather than environmental difference. Although the smaller scale means such transfers may well be reciprocal (Section 2.3.6), they do not need to be to generate stability. This result is interesting for showing oscillation damping, the

stabilising factor in two-directional gene exchange (Tellier & Brown 2011), can occur with one-way gene-flow, albeit under more limited conditions than with two-way gene-flow.

Low to high disease incidence is the only case where one-directional transfer stabilises polymorphism in the receiving patch. If both patches have incidence below us^{-1} both go to fixation for *AVR* and *res*. If the source patch has disease incidence above us^{-1} it imposes its oscillation period, increasing oscillation magnitude and eventual genotype fixation on the sink-patch. This occurs even with very low rates of transfer of seeds, spores, or both.

2.3.6 - Two-directional gene-flow dampens oscillations and leads to stable polymorphism in both patches across a wide range of parameters

As described in methods, bi-directional gene-flow can lead to stability by damping oscillations in genotype frequency. Figure 2.4, below, is an example of this.

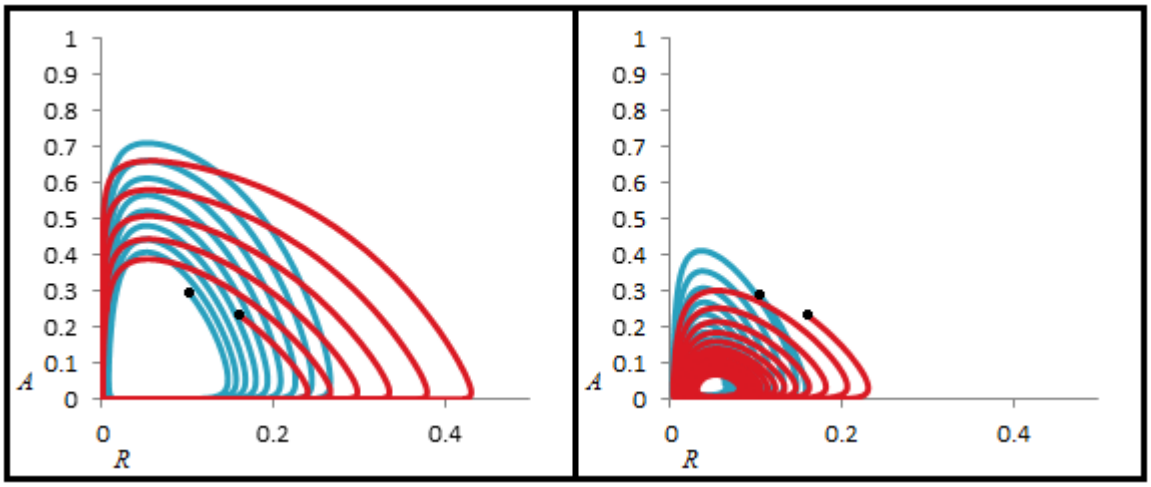

Figure 2.4 – Unlinked oscillations (panel A, spiralling outwards towards fixation) and linked oscillations (panel B, spiralling inwards towards IEPs) in gene frequency in two patches with differing disease incidence. X-axes are *R* and y-axes are *A*. Costs are u=0.01, s=0.3, b=0.05 and c=1. Disease incidences are m_1=0.5 and m_2=1. Oscillations in patch 1 are blue and in patch 2 are red. Initial frequencies are marked with black circles.

Oscillation damping occurs over a wide range of parameter spaces, some of which are briefly described below. Stable ranges are much less limited than for one-directional stability and include considerable regions of parameter space where both populations have disease incidences above us^{-1} (Figure 2.6, all panels although less so in the bottom left-hand panel).

For two-directional gene-flow, what the observed stable parameter spaces have in common is a notable difference in incidence levels *m*1 and *m*2 and (generally) wider ranges of other parameters as this difference increases. For these results stable is defined as the complete damping of oscillations within 500,000 generations preserving polymorphism in one or more of the four populations (hosts and parasites in two patches) in the system.

The effects of several parameters on stable polymorphisms resulting from gene-exchange were investigated. In the following experiments gene-exchange was always symmetrical (g_{ij}=g_{ji}, h_{ij}=h_{ji} and g_{ij}=h_{ij} unless g_{ij}=0 or h_{ij}=0)

The most significant results from simulations are presented graphically in Figures 2.5 (below) and 2.6 (page after next).

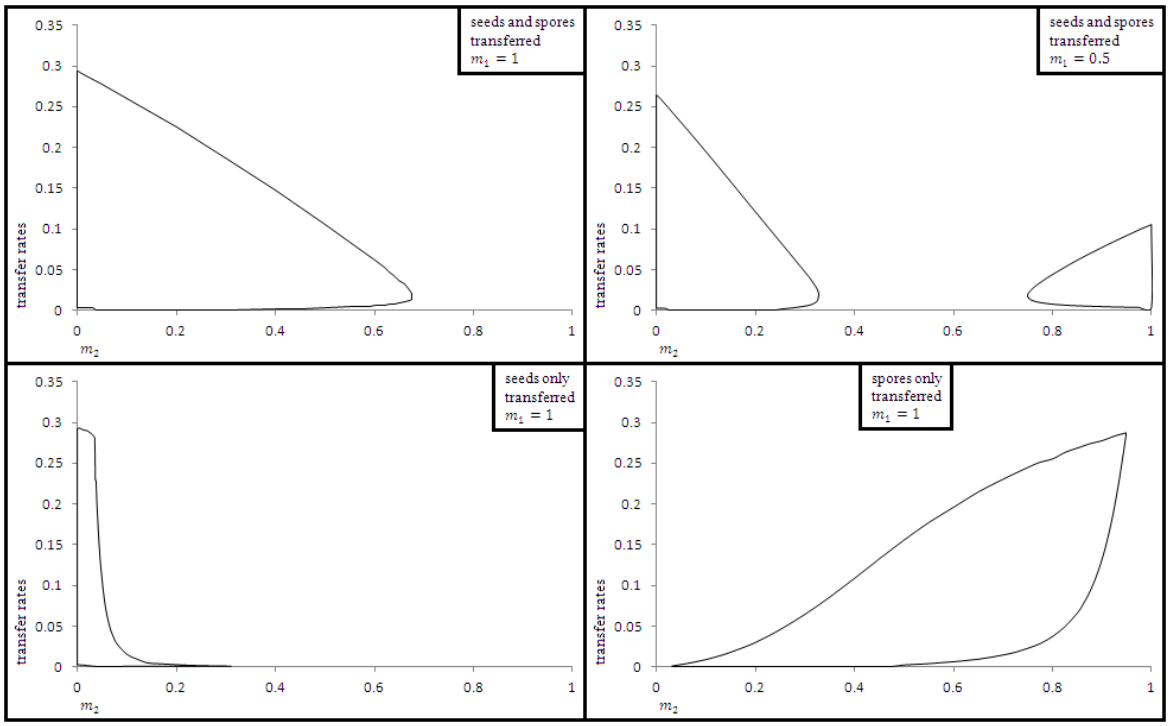

Figure 2.5 - Zones of stable polymorphism in terms of patch 2 disease incidence m_2 (x-axes) and rate of seed/spore transfer between patches (y-axes, specified in legends). Patch 1 disease incidence m_1 = 1 in all cases except the top right-hand panel, where m_1 = 0.5. All transfers of seeds, spores and both are two-way and symmetrical, that is $g_{12} = g_{21}$ and $h_{12} = h_{21}$. Costs are *u*=0.01, *s*=0.3, *b*=0.05 and *c*=1. Greater difference between disease incidences in the two patch increases the range of transfer rates over which stable polymorphism occurs, in all cases except transfer of spores only.

Figure 2.5 shows zones of stable polymorphism for a fixed set of costs. Figures are phase-spaces, varying in patch 2 disease incidence m_2 and transfer parameters h_{ij} and/or g_{ij} (legends) while holding other parameters constant. The top two panels are cases where seed and spore transmission rates are the same. There are zones of stable polymorphism whenever patch 2 disease incidence m_2 is sufficiently different from patch 1 incidence m_1 (1 in the first panel, 0.5 in the second panel) and these stable zones grow wider with respect to transfer rates as the difference between incidences increase. The lower left- and right-hand panels show the stable zones when only seeds and spores respectively are exchanged. When only seeds are exchanged, the stable zone shrinks more rapidly as m_1 approaches m_2 than it does with the exchange of both seeds and spores. When only spores are exchanged, surprisingly, the trend of shrinking stable zones with increasing similarity of incidences is broken. Instead the stable zone grows wider as m_2 approaches m_1 until the incidences are quite similar, then rapidly shrinks and vanishes before they converge.

Different values of fitness costs *b*, *c*, *u* and *s* in patches linked by gene exchange can lead to stability (Tellier & Brown 2011). The mechanism is the same as for different incidences - linked patches have different frequencies and thus dampen and stabilise oscillations in each other. Changing fitness costs by the same amount in both patches will not lead to the differences in oscillation period that allow stability. However changing costs in both patches alters the period of both patches and, when disease incidence levels and thus oscillation periods are already different, changing costs does alter the stable zones of transfer rates and differences in incidence. This is shown in Figure 2.6, overleaf.

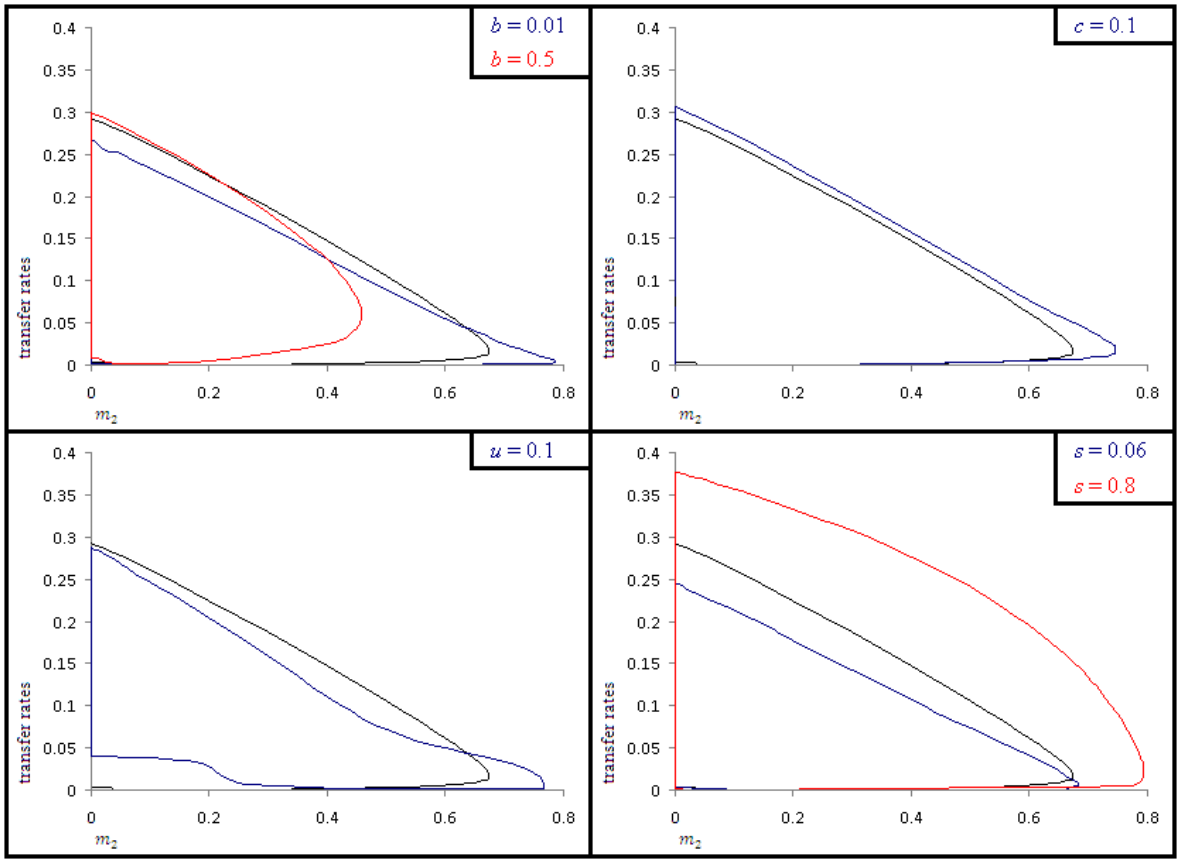

Figure 2.6 - Effects of varying costs on zones of stable polymorphism. x-axes are patch 2 disease incidence m_2 and y-axes are transmission rates for seeds g_{ij} and spores h_{ij} . All transfer-rates are two-way, symmetrical and of both seeds and spores ($g_{12} = g_{21} = h_{12} = h_{21}$). The black-edged stable zone in all panels is for a standard set of costs (b = 0.05 c = 1 u = 0.01 s = 0.3). One cost is varied in each panel. In the upper left-hand panel b = 0.01 is the blue-edged stable zone and b = 0.5 is the red-edged stable zone, in the upper right-hand panel c = 0.1 is the blue-edged stable zone, in the lower left-hand panel u = 0.1 is the blue-edged stable zone and in the lower right-hand panel s = 0.06 is the blue-edged stable zone and s = 0.8 is the red-edged stable zone.

Figure 2.6, above shows that stable zones are wider with respect to both transfer rates and similarity of incidences for decreased cost of detection c (upper right-hand panel) and increased cost of disease s (lower right-hand panel). They are wider with respect to similarity of incidences but narrower with respect to transfer rate for increasing cost of resistance u (lower left-hand panel) and decreasing cost of virulence b (upper left-hand panel).

2.4 - DISCUSSION

2.4.1 - Overview

Stable polymorphism requires negative direct frequency-dependent selection (ndFDS), or an analogous factor such as density-limited reproduction (DLR), generated by processes which attenuate natural selection in time or space. Model 1 is an example of attenuation in time as parasites experience DLR. Model 2 is an example of attenuation in space as linked cycles dampen one another. Both mechanisms are capable of generating stable polymorphism. Importantly, the two mechanisms are not mutually exclusive. They are very likely to occur in combination with each other and with other stabilising factors, especially in wild as opposed to agricultural systems, and they could well reinforce one another and extend the conditions under which stable polymorphism is possible. The dynamics important to both models could be caused by a variety of real-world processes.

A key point is that both Model 1 and Model 2 predictions match the observed dichotomy between long-term stable polymorphism in wild pathosystems and rapid fixation of genotypes in agricultural pathosystems. Agricultural plants are grown at high densities, increasing the parasite's within-patch growth-rate. They are also grown in locations and treated in ways which favour the plants, reducing differences between patches.

2.4.2 - Hypotheses from Models 1 and 2 and biological evidence

Model 1 predicts polymorphism will be most common and stable when disease incidence is low to intermediate. Generally wild pathosystems are known to have lower disease incidence than agricultural crops (Laine 2005, Burdon & Thrall 1999) and are more likely to preserve polymorphism (Stahl et al. 1999, Holub 2001, Tian et al. 2003). A modelling study by Packer et al. (2003) found host-parasite cycles were more stable when host density was regulated by predation, which would effectively reduce parasite reproductive rate and thus incidence. Model 4 in Chapter 4 addresses negative direct regulation of host density in GFG pathosystems.

Model 2 predicts polymorphism will be most common and stable when patches with different levels of disease incidence, due to biotic, abiotic or stochastic factors, are linked by limited dispersal of seeds or spores or both. Many plant diseases have reproductive

success and thus disease incidence affected by climate, causing incidence to vary between regions. A review of wheat diseases (Juroszek & Tiedemann 2013) stated many have varying severity with climate and climate change will have different effects on different diseases. This suggests that in many pathosystems costs and incidences can vary between regions, allowing oscillation damping as seen in Tellier & Brown (2011) and Model 2.

Model 1 suggests balanced polymorphism in both species (Type 4 stability) requires a certain level of intermediate disease incidence. Low-density hosts, poor environmental conditions for parasites and less effective reproduction or dispersal of parasites are all factors that may lead to disease incidence being lower than this level. Model 1 Type 2-3 results suggest such low disease incidence will lead to low disease incidence, *AVR* fixed and *RES* respectively absent or present at very low levels. If a plant parasite is well-studied it is likely fairly successful. It is also likely to be an agricultural parasite, which because of high host density means much the same thing. Thus, pathosystems with lower disease incidence and Type 2 or 3 dynamics may be occurring across the world without attracting a great deal of notice. Long-term studies of GFG dynamics in wild pathosystems (Burdon & Thrall 1999, Laine 2005) typically report polymorphism for both species but, again, these diseases are successful and high-incidence enough to be conducive to study. Wild pathosystems with lower disease incidence within patches and lower disease presence at a metapopulation scale may give different results.

In cases where pathosystem metapopulations are characterised by repeated extinction-recolonisation events (Laine 2005, Thrall & Burdon 2002), both Model 1 and Model 2 dynamics occur. Disease incidence will vary both within patches due to population growth over time and between patches due to the extinction-recolonisation. Such systems are common in nature and Models 1 and 2 suggest this wide range of incidences in both time and space would act to stabilise polymorphism. Variable incidences caused by extinction-recolonisation dynamics are the exact opposite of agricultural situations, where high crop densities across large areas enable parasites to establish and reach high incidences in most locations in most years.

Populations of the same pathosystem in different locations can differ in disease incidence (Workneh et al. 1999). The parasite's effective rate of production could be affected by the environment, the host's density or both. (Such differences could also promote temporal variability for prevalence within patches, underlining the point that mechanisms in Model

1 and Model 2 may often occur simultaneously.) Limited gene-flow between effectively panmictic patches, as described in Model 2, appears to be a reasonable description of several natural plant-parasite systems (Burdon & Thrall 1999, Laine 2005). Because linked patches can stabilise each other with transfer of seeds, spores or both and because some seeds and spores can travel great distances and link very different environments, this could be a very important mechanism for maintaining polymorphism (Thrall & Burdon 2000, Laine 2005).

Further hypotheses based on this work are:

- In closely related pathosystems, or geographically isolated cases of the same pathosystem, the likelihood of polymorphism in immune/antigenicity genes corresponds to the average parasite incidence, which may depend on factors such as local climate and host-density as well as parasite reproductive ability.

- Long-term stable polymorphism is more likely when the host-parasite association occurs across a range of environments with different characteristic levels of disease incidence in each, assuming a small amount of transfer between environments.

- Hosts which have low levels of disease incidence will not be under strong selection to develop resistance. If resistance occurs it will be at very low frequencies and parasite adaptation to overcome such resistance will consequently be very rare.

- In cases of radically increasing host-density, such as agricultural crops or animals or human populations, there will be a somewhat abrupt switch from stable to unstable dynamics – this may be detectable by molecular phylogenies.

- Similarly, increasing mobility and transport between distant populations will increase transfer rates and somewhat abruptly destabilise polymorphism.

The increased susceptibility of large, dense, highly-connected host-populations to both parasite invasion in general and parasite evolution to overcome resistance in particular doubtless has implications for biodiversity. Moreover, it has important implications for both humans and the plants and animals they depend on in today's increasingly crowded and connected world.

CHAPTER 3 – INDIVIDUAL-BASED SPATIAL MODELLING WITH VARIABLE DISPERSAL

OVERVIEW

Polymorphism at corresponding host-immunity and parasite-antigenicity loci is common and can persist for millions of years. The gene-for-gene (GFG) model of immune interaction between plants and their parasites is a well-understood case of this coevolutionary relationship and has been extensively studied and modelled. Maintaining polymorphism in interacting host and parasite genes requires a source of negative direct frequency-dependent selection (ndFDS) or an analogous selective pressure such as density-limited reproduction (DLR). This selection means that as a genotype grows more common its relative fitness decreases. Many ecological processes can cause such selection. Although much of this theory has been developed with non-spatial deterministic models, spatial and stochastic processes influence dynamics in real host-parasite systems.

Here, I extend a non-spatial model with variable disease incidence as the stabilising factor (Chapter 2) to a spatially explicit individual-based model. Dispersal distances for plant seeds and parasite spores are variable. Lowering these distances leads to stronger local adaptation and thus higher incidence for avirulent parasites, thus also to higher incidences for resistant hosts. Oscillation cycles are longer and less pronounced in this situation because low dispersal weakens selective pressure and slows the effective rate of natural selection. As avirulent parasites and resistant hosts are the genotypes commonly lost at low frequencies in their oscillations, both these factors make genotype loss less likely.

I distinguish between deterministic and stochastic genotype loss. The former is caused by no or insufficient ndFDS leading to divergent genotype frequency oscillations, is mathematically tractable and is well-addressed in previous chapters and by earlier modellers. The latter is caused by random fluctuations in limited populations with overall stable dynamics and becomes more or less likely depending on population size and how much time oscillations spend near the boundaries. This stochastic genotype loss is what is reduced by limited dispersal.

3.1 - INTRODUCTION

3.1.1 –GFG model and stability of coevolution

Gene-for-gene (GFG) interactions are an extensively studied, well-understood and agriculturally important case of interacting host-immunity and parasite-antigenicity genes. The GFG model was first proposed in the 1950s (Flor 1955) and is now known to underlie reactions between most plant species and fungal, bacterial and viral parasites (Dangl & Jones 2001, Chisholm et al. 2006, Stukenbrock & McDonald 2008, Dodds & Rathjen 2010). In the GFG model, the product of a plant's resistance (*RES*) gene detects the product or downstream result of a parasite's avirulence (*AVR*) gene. This detection leads to a defence response by the plant. *RES*-genes have susceptible (*res*) alleles which do not detect avirulence genes and *AVR*-genes have virulence (*avr*) alleles which are not detected by resistance genes. More details are found in Chapter 1 (Section 1.1).

Polymorphism at these loci is common and can persist for millions of years (Stahl et al. 1999, Holub 2001). Stable polymorphism requires both indirect and negative direct frequency-dependent selection, iFDS and ndFDS (Chapter 1, Sections 1.1.5-1.1.7) (Tellier & Brown 2007A).

In Chapter 2, I showed that variable disease incidence can stabilise polymorphism in a non-spatial system (Model 1). Parasites experience density-limited reproduction (DLR), reducing the fitness of both *avr*- and *AVR*-genotypes as total parasite incidence increases. This is similar to the stabilising ndFDS generated by numerous ecological and epidemiological factors (Tellier & Brown 2007A, 2009) except that it operates on the total density of the parasite population rather than the relative frequencies of genotypes. Here, I study this stabilising mechanism in a spatially explicit model.

3.1.2 - Spatial modelling – dispersal and local extinction occur in real populations

In theoretical population biology, deterministic non-spatial models are easier to create and analyse than stochastic spatial models and have led to advances in many fields of population ecology. However, real populations have spatial structure. Processes such as local extinction and re-colonisation change the dynamics of populations, both in nature (Burdon & Thrall 1999, Laine & Hanski 2006) and in models (Damgaard 1999, Thrall &

Burdon 2002). These processes are important to our understanding of host-parasite coevolution, as well as applied epidemiology.

It has been suggested that spatial structure by itself can stabilise resistance-antigenicity polymorphism (Sasaki et al. 2002, Thrall & Burdon 2002). It has also been suggested that finite population processes such as genetic drift can stabilise polymorphism (Salathe et al. 2005). However these models rely on mutation to replenish genotype polymorphism, which is problematic as discussed in Chapter 1 (Section 1.2.3). Mathematical analysis shows that spatial structure can slow but not prevent fixation in a system without ndFDS or an equivalent stabilising factor (Tellier & Brown 2007A), although this transient polymorphism can last for a very long time as in Sasaki's model.

Spatial structure by itself does not stabilise polymorphism. However structure and dispersal can be part of some ndFDS-generating mechanisms. Most obviously spatial structure allows the variation in costs, disease incidence or possibly other parameters between or across environments that causes oscillation damping, which in turn generates ndFDS (Tellier & Brown 2011, Model 2 in Chapter 2). Alternatively, spatial processes may alter the dynamics in models with essentially non-spatial sources of stability (Daamgard 1999, Thrall & Burdon 2002).

In this chapter Model 1 (Chapter 2) is developed into a spatial, individual-based model (Model 3) and the effects of this spatial dimension are analysed. In particular I focus on limited dispersal distances for host seeds and parasite spores. Limited dispersal does not by itself generate stability. However in a finite-population system which already includes the stabilising DLR , limited dispersal can make genotype loss less likely. This is because limited dispersal allows local adaptation of *AVR*-parasites, increasing the equilibrium frequency of *AVR*-parasites and thus of *RES*-hosts, and reduces the strength of selection, thus reducing the frequency and amplitude of genotype frequency oscillations. As *RES* and *AVR* are the genotypes that typically go extinct, increasing their average frequencies and reducing the rate and amplitude of oscillations in these frequencies makes random loss of these genotypes less likely. While not as mathematically tractable as the deterministic genotype loss discussed in Chapter 2, such stochastic genotype loss and the factors reducing it may be important in the real world.

3.2 - MODEL AND METHODS

3.2.1 – NetLogo – a spatial, agent-based modelling package

NetLogo is a freeware modelling package that models large numbers of autonomous agents in a spatially explicit environment (Wilensky 1999). It supports two types of agents, patches and turtles. My models only use patches, which are 1-by-1 unit squares that compose the spatially explicit simulation. Each patch has its own status, with its own values of whatever individual variables the user specifies. Turtles, by contrast, are mobile, point-like agents. The "world" (arena) consists of a two-dimensional array of patches, looped vertically and horizontally. Arena size can be controlled.

3.2.2 - Model 3 - model summary

The model is a square array of patches. Each patch contains an individual host and may or may not contain a parasite. Individuals reproduce and die in annual time-steps (one generation per year for both species). The fitness costs are as in Model 1 but, instead of each genotype having an average fitness, each individual has its own fitness based on its genotype and the presence/absence and genotype of the other species. Each generation each individual produces seeds or spores, in numbers which depend on its fitness, and these disperse a random-exponential distance. One seed and a maximum of one spore establish in each patch. The current occupants die and are replaced by the seed and (if present) the spore, which grow to adulthood. Modelling terms are specified in Table 3.1, while fitnesses are specified in Table 3.2. Both are found at the end of Section 3.2.3.

3.2.3 – Model 3 - model description

Each patch always contains one plant, a simplifying assumption meaning that plant population size and density are constant and not affected by the parasite. Each patch (i.e. each host) may or may not contain a parasite. Plants can be *RES* or *res*. As in Model 1, it is assumed that all the parasites on a single host come from one spore so the parasites on each host are wholly *avr* or *AVR*. There are fitness costs *u* of being *RES* and *b* of being *avr*. *u* and *b* are constitutive, paid regardless of the genotype or presence/absence of the other species. There is a cost *s* to the host of being diseased when the parasite's infection is successful, paid by *res*-plants with any parasite and *RES*-plants with *avr*-parasites but

not by *RES*-plants with *AVR*-parasites. There is a cost *c* to the parasite of being detected so the infection is unsuccessful, paid by *AVR*-parasites on *RES*-hosts.

Fitness is calculated separately for each individual from the above costs (Table 3.2, below). Fitness scores determine the absolute (for parasites) or relative (for hosts) number of seeds or spores produced. Individuals take it in turns, in random order, to disperse all their seeds or spores. The first seed to arrive in a patch always establishes and exclude all other seeds. To simplify the model while retaining its essential elements and to reduce computation, it is assumed that the first spore to land on each host establishes, grows to an entire plant's worth of parasites and prevents any subsequent spores establishing. It is further assumed that the number of spores produced by colonies is typically in single figures. "Individual" spores are thus a simplification analogous to "individual" parasite colonies.

Dispersal parameters are β_h for hosts and β_p for parasites. Dispersal distances are randomly generated from an exponential function with a mean of the relevant β. Seeds and spores disperse from the centre of a patch and if β is sufficiently low they are more likely to auto-infect that patch than to move outside. Dispersal direction is uniformly random. In this chapter, β without a subscript generally refers to both dispersal parameters with equal values. For example, the effects of increasing both β_h and β_p and keeping their values the same would be referred to as the effect of increasing β.

Hosts and parasites are modelled as haploid, thus seeds and spores inherits parental *RES*/*res* or *avr*/*AVR* genotypes. There are assumed to be no other genetic or random differences between individuals which affect fitness. Thus all plants are the same "size" so they occupy one patch, have the same relative seed-number except for the effects of costs *u* and *s* and provide the same resources for parasites. Similarly all successful parasite infections are the same "size" so occupy one host, have the same spore-number except for the effects of costs *b* and *c* and random rounding and do the same damage to the host unless infection is prevented by the GFG-relationship. The model parameters and variables are summarised in Table 3.1. Host and parasite fitnesses are summarised in Table 3.2. Figure 3.1 shows examples of simulations of Model 3 in progress which contrast low and high dispersal parameters. The typical values are those used in the simulations described in this chapter. Parameter choices are discussed in Section 3.3.4.

Name	Description	Typical values
	Global constants	
u	Fitness cost to plant of having a resistance gene	0.01
s	Fitness cost to plant of being diseased, paid by *res*-plants with any parasite and *RES*-plants with *avr*-parasites	0.3
b	Fitness cost to parasite of being *avr*	0.05
c	Fitness cost to parasite of being detected, paid by *AVR*-parasites on *RES*-plants	1
y	Parasite basic reproductive number	1.25
$\boldsymbol{\beta_h}$, $\boldsymbol{\beta_p}$	Scale parameter (mean) of exponentially distributed random dispersal distances for plant and parasite respectively	0.5-25
	Size of NetLogo arena	63-by-63 or 255-by-255
	Individual variables	
$\boldsymbol{W_h}$	Fitness of individual plant, see Table 2	≤ 1
$\boldsymbol{W_p}$	Fitness of individual parasite, see Table 2	≤ 1
$\boldsymbol{\sigma_p}$	How many spores a given parasite colony produces, rounded up or down probabilistically to an integer	yW_p
	Global variables - discussed in Box 3.1	
R	Fraction of hosts *RES* at current time-step	
r	Fraction of hosts *res* at current time-step; $R + r = 1$	
A	Fraction of parasites *AVR* at current time-step	
a	Fraction of parasites *avr* at current time-step; $A + a = 1$	
M	Disease incidence (fraction of patches with parasites) at current time-step	
MA, Ma	Incidence (fraction of patches) of *AVR*-parasites and *avr*-parasites respectively at current time-step	

Table 3.1 - Model-wide constants and individual-based variables for Model 3. Costs and parasite basic reproductive number are described and discussed in Chapter 2. The typical values for model-wide constants are those used in most of the simulations described in this chapter (details under Experiments). The typical values for individual variables are functions of the constants, generated by the model processes.

	***RES*-host *AVR*-parasite**	***RES*-host *avr*-parasite**	***RES*-host no parasite**	***res*-host *AVR*-parasite**	***res*-host *avr*-parasite**	***res*-host no parasite**
W_h	1 - u	(1 - u)(1 - s)	1 – u	1 – s	1 – s	1
W_p	1 - c	1 – b	-	1	1 – b	-

Table 3.2 - Fitness outcomes of all possible plant-parasite interactions. These apply to individual hosts and parasites rather than being averaged as in Models 1 and 2 (Chapter 2).

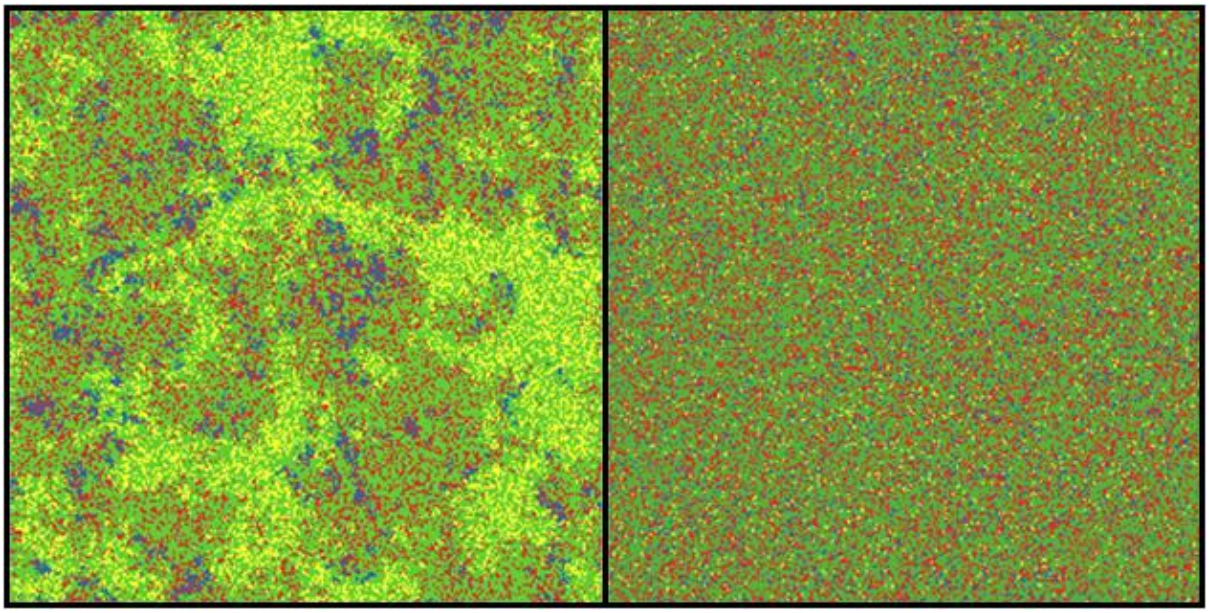

Figure 3.1 - Contrast of low-dispersal (β = 1) and high-dispersal (β = 15) systems. Blue patches are *RES*-plants, green patches are *res*-plants, red dots are *avr*-parasites and yellow dots are *AVR*-parasites. In the low-β system local clustering of genotypes occurs, whereas in the high β system dispersal prevents local clustering.

Initial conditions are not hugely important, as the model very quickly assumes equilibrium, but they are set as follows. In each patch a plant will be present and a parasite may be present. Initial frequencies are indicated by the subscript 0. Probabilities of a given patch being set with *RES*-hosts, parasites present and *AVR*-genotype parasites respectively are: $R_0 = b/c$, $M_0 = 1-\{u(1-b)\}^{-1}$ and $A_0 = (u\text{-}usM_0)(s\ M_0(1\text{-}u))^{-1}$. These values are based on the Type 4/5 IEP of the non-spatial Model 2, described in Table 2.3, Chapter 2. The model thus has no initial non-random spatial structure.

Overview – processes and order: The model proceeds in annual time-steps, each year being one generation for both host and parasite. Both species die over winter and are replaced by seeds and spores, generated and dispersed that year, which grow to adulthood

the following year. Seeds and spores always germinate the year after dispersal and never lie dormant, so there is no long-term seed- or spore-bank. Individuals' fitnesses are determined by their own genotype and interactions with the other species and they produce seeds or spores in numbers based on that fitness. Individuals, in random order, disperse their seeds and spores. All adult individuals die, ready for the next generation to grow. In more detail, these processes operate as follows.

Growth: After the end of the previous generation and the annual die-back of hosts and parasites, patches are empty except for the seeds and spores distributed last generation. There will be one seed and a maximum of one spore in each patch. All seeds and spores grow into adult plants and parasites. As the seeds and spores have become adult individuals, patches now contain no seeds or spores.

Fitness: For each patch host and parasite fitnesses are calculated according to genotypes and interactions, as in Table 3.2.

Spore generation: Each parasite "individual" produces a number of spores equal to its fitness multiplied by parasite basic reproductive number *y* and rounded up or down probabilistically. Thus 1.1 has a 0.1 chance of being rounded to 2 and a 0.9 chance of being rounded to 1, while 0 or 1 will always be rounded to 0 or 1 respectively.

Spore dispersal: Each parasite in turn, in random order, disperses all its spores. Each spore is dispersed in a random direction and a random exponential distance with mean β_p. If a spore lands in a patch without a spore it becomes that patch's spore. A patch may have one or no spores. Subsequent spores landing in a patch which already has a spore fail to establish.

Seed dispersal: Each patch selects a potential donor plant from a patch a random direction and a random-exponential distance with mean β_h away. To account for relative host fitness, a random number between 0 and 1 is generated. If this is less than the potential donor plant's fitness, the potential donor provides the seed for the patch. Otherwise, the process is repeated until a potential donor passes the fitness test and provides a seed. Each patch receives one seed, so all patches will contain one adult plant. Spore- and seed-dispersal are equivalent but, because patches always have plants but do not always have parasites, each individual patch takes in precisely one seed while

individual parasites disperse their spores to some but not all patches. Fitness also has subtly different effects in the two species. In parasites it affects the absolute number of spores, while in plants it determines the chance of successfully providing a seed at each opportunity. Distances are measured from the centre of the patch dispersing spores or searching for seeds and for low values of β_h and β_p patches are more likely receive seeds or spore from themselves than from elsewhere. As the arena loops horizontally and vertically, there is no chance a spore will disperse out of the arena or a patch will look for a seed from outside the arena.

Data acquisition: Every generation the programme records current and average values for genotype frequencies *R*, *r*, *a* and *A* and disease incidence *M*. The programme also records whether any genotypes have gone extinct.

End of annual time-step: Everything dies, leaving seeds and spores ready to become hosts and parasites next generation.

Box 3.1 - Genotype frequencies, disease incidence, averages and notation

As in Model 1 *R*, *r*, *a* and *A* are respectively the frequencies of *RES* and *res* genotypes within the host population and *avr* and *AVR* genotypes within the parasite population. $R + r$ and $a + A$ always equal 1. *M* is parasite incidence, defined here as the frequency of parasites in the system or the proportion of patches with parasites in the arena. Parasite incidence is calculated relative to the number of patches in the arena, not the number of hosts; in Model 3 the two are the same, but in Model 4 (Chapter 4), host incidence also varies. *Ma* and *MA* are therefore the frequencies of virulent and avirulent parasites within the arena, rather than within the parasite population. In relation to both parasites affecting hosts and the likelihood of random extinction, the key parasite frequency is typically *MA* rather than *A*, *a* or *Ma*.

Average genotype frequencies refer to values averaged over at least 2000 generations in large arenas. Actual time varied widely because highly dispersing systems reached equilibrium more rapidly. Averages are measured from the beginning of the simulation, without a burn-in period. This is not an issue, even though measured average *a* and thus *Ma* and *M* are higher than the predicted equilibrium values from Model 1 used to set initial conditions, because the system reaches equilibrium extremely quickly and any long-term

impact on averages is negligible. Time courses demonstrating this can be seen in Figure 3.5.

For clarity, the terms average *R*, average *A*, average *M* etc. refer to measured average frequencies and incidences from Model 3 and other spatial models, while the notation $\hat{R}$ etc. refers to predicted equilibrium frequencies from Model 1 or similar deterministic models. In many cases these are similar, but the notation and meaning are distinct.

3.3 - SIMULATIONS

3.3.1 - Dispersal parameters β_h and β_p were varied while other parameters were mostly kept constant

To investigate the effects of varying dispersal and the resulting spatial structure on coevolutionary outcomes, dispersal parameters were varied. β_h and β_p were varied both together and separately, in simulation sets 3.1 and 3.2 respectively. Costs and parasite intrinsic rate of reproduction were set to [u=0.01 s=0.3 b=0.05 c=1 y=1.25]. These parameters were kept constant to focus on the effects of varying dispersal and emerging spatial structures. The costs are biologically plausible. Costs to the host of resistance (u) and to the parasite of virulence (b) are often difficult to detect, may be environmentally labile (Tian et al. 2003, Brown 2003B) and are generally expected to be low (Bergelson & Purrington 1996, Thrall & Burdon 2003). Cost to the host of infection (s) can vary from almost nothing to total loss of fitness and 30% represents an intermediate value. The cost to the parasite of being detected (c) is generally modelled as total (Tellier & Brown 2007A). For these costs, a parasite growth-rate y of 1.25 generates low to intermediate average disease incidence (15.8% in Model 1 and 30% + in Model 3; Figure 3.2, below) and polymorphism in both species. The presentation of the results of simulations focuses on the existence and stability of polymorphism and the host and parasite gene frequencies. Different values of y and slightly different costs are used in simulation set 3.3, described below.

Initial investigations suggested that, for the costs and arena sizes used, the most significant variations in behaviour with changing β_h and β_p were found when dispersal parameters varied between 1 and 4. Thus these values were focussed on in all sets of experiments. It

seems likely that in much larger arenas changes would occur over a wider ranges of β_h and β_p, because dispersal is only local or limited relative to arena size. Other than arena size, there is no reason to think varying any other parameter (*b*, *c*, *u*, *s*, *y*) would affect the range over which dispersal alter the system's behaviour.

3.3.2 - Average frequencies were recorded in large arenas, average stability in smaller arenas

Any finite population modelled as a stochastic system will eventually lose genotypes through random processes, even if stabilising ndFDS or DLR means that an analogous infinite, fully deterministic population model is stable. In finite-population models with strong ndFDS, stochastic genotype loss is overwhelmingly likely to take a very long time in sufficiently large populations (possibly millions of generations), while stochasticity overwhelms any tendency to stability in sufficiently small populations. Thus 255-by-255 arenas with one replicate were used to find average genotype frequencies and disease incidences, because the longevity and reduced stochasticity of large systems meant that average values had time to emerge over at least 2000 generations. Conversely measuring the strength of stability requires intermediate-sized systems, such that genotype fixation occurs in a reasonable length of time but this length of time is significantly influenced by how intrinsically stable the system is and is not overwhelmed by stochastic effects. Thus 63-by-63 arenas with at least 30 repeats were used to measure average time to fixation (TTF) as a measure of stability. A cut-off of 1500 generations was used to prevent the slowly-dispersing systems with higher stability taking an inordinate amount of time to run.

3.3.3 - Parasite growth-rate was varied to study effect of dispersal on stable zones, with stability measured as average TTF in small arenas

A third set of simulations investigated whether varying dispersal parameters affected stable zones. Stable zones are defined here as the range of parasite basic reproductive rate *y* over which stable polymorphism is observed for a given set of costs. In Model 1 (Chapter 2) it is possible to predict whether polymorphism is stable or unstable from the value of intrinsic parasite reproductive rate *y* relative to costs *u*, *s*, *b* and *c* and the transition between stability and instability is abrupt. In a finite population model, this abrupt transition is replaced by a gradual change in measures of stability as *y* changes relative to costs.

It was predicted that the values of *y* over which these changes occur, and therefore the limits of different types of stability, would be influenced by varying β_h and β_p. In particular, I predicted that low dispersal would increase the maximum value of *y* for any given set of costs [*u s b c*]) at which stable polymorphism was observed. The maximum *y* for stable polymorphism was considered to be the value of *y* at which average TTF started to decline or reached an arbitrary intermediate value defined as the limit of the stable zone. This is because of self-exclusion; low dispersal means more parasite spores land on their parent patch and are excluded by sibling spores. Such sibling competition lowers the strength of natural selection.

3.3.4 – Simulation sets

Simulations 3.1 - Varying β (β_h and β_p together)

Dispersal parameters for the host and parasite were equal, at the following values: 0.5, 1, 2, 3, 4, 6, 8, 10, 15, 20 and 25. As described in Section 3.32 63-by-63 arenas with 40 replicates were used to measure stability (average TTF) and 255-by-255 arenas run for at least 2000 generations were used to measure average genotype frequencies and the period and magnitude of oscillations in these frequencies.

Simulations 3.2 - varying β_h and β_p independently

To investigate differences and interactions between the effects of these parameters, all combinations of β_h = [1 2 3 4 5] and β_p = [1 2 3 4 5] were used. As in Simulations 3.1 63-by-63 arenas with 40 replicates were used to measure stability, while 255-by-255 arenas run for at least 2000 generations were used to measure average genotype frequencies and the period and magnitude of oscillations in those frequencies. Simulations 3.1 and 3.2 both used costs [u = 0.01 s = 0.3 b = 0.05 c = 1], as discussed in Section 3.2.

Simulations 3.3 - effect of β_h and β_p on stable zones

In the deterministic Model 1, stability depends on the value of parasite basic growth-rate *y* relative to costs. The boundary values of *y* between different behaviours are analytically solvable (Table 2.3, Chapter 1). In the stochastic Model 3, time to fixation or TTF is used

as a measure of stability. Simulations 3.3 investigate how average TTF changes with changing *y* in a spatial individual-based model, whether these changes conform to Model 1 boundaries and whether this relationship is altered by varying dispersal parameters. Thus *y* was incrementally varied for different combinations of dispersal parameters. Two different sets of costs were used. Costs A were the common costs used in Simulations 3.1 and 3.2. Costs B are a different, generally higher, set of costs to investigate whether this affected the relationship between stable zones and dispersal. Both Type 3 and Type 4 stability were recorded (Section 3.4.5).

y was set at 0.02 intervals between 1.02 and 1.4 and at 0.05 intervals between 1.4 and 1.5. These values were chosen to focus on the predicted boundaries between stable and unstable behaviour from Model 1. Costs and predicted boundaries are given in Table 3.3, below. Stability was measured as average TTF in 63-by-63 arenas with 30 repeats. Simulations were limited to 4000 generations, to prevent especially stable values of *y* from taking an inordinate amount of time to run. For each *y* and set of costs, all combinations of β_h=[1 3 5] and β_p=[1 3 5] were run. Two measurements were taken from each repeat - the time when any one of the four genotypes went extinct, which estimates Type 4 stability, and the time when one of *RES*, *res* and *AVR* went extinct, measuring Type 3 stability.

	u	*s*	*b*	*c*	unstable/Type 3 boundary *y*	Type 3/4 boundary *y*	Type 4/unstable boundary *y*
Costs A	0.01	0.3	0.05	1	1.0345	1.0889	1.2868
Costs B	0.02	0.7	0.1	1	1.0294	1.1438	1.3324

Table 3.3 – Cost sets A and B and predicted (Model 1) boundary values of *y* between different behaviours.

3.4 - RESULTS

3.4.1 - Results of Simulations 3.1, varying β (β_h and β_p together) - low dispersal increases average *A* and *R*, decreases frequency and amplitude of oscillation cycles and increases stability

Average avirulence *A* decreased sharply as β increased from 1 to 4 and decreased very slightly as β increased further (Figure 3.2). Average virulence *a* mirrored this, increasing sharply then slightly as β increased. Average disease incidence *M* decreased to a lesser extent with increasing β. Average resistance *R* decreased from 0.121 to 0.044 as β increased (Figure 3.3). All curves were asymptotic with respect to increasing β.

At higher β, average *MA* and *R* approached the equilibrium values $\hat{M}\hat{A}$ and $\hat{R}$ from Model 1. This makes sense as higher β should make Model 3 behave more like a non-spatial system. Average *Ma*, however, remained above $\hat{M}\hat{a}$ from Model 1 regardless of β. Thus, average *M* and *A* were respectively higher and lower than predicted $\hat{M}$ and $\hat{A}$.

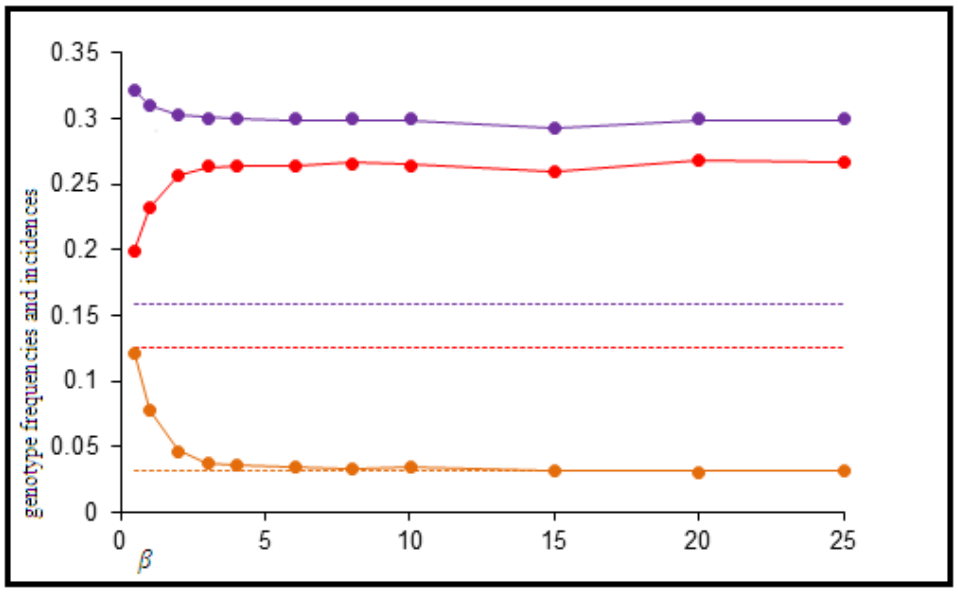

Figure 3.2 – Average values of *MA* (orange), *Ma* (red) and total disease incidence *M* (purple) against increasing dispersal parameters β. Dotted lines are the predicted levels from the non-spatial Model 1 (Chapter 2). x-axis is both β and y-axis is frequency. Average *MA* is higher when dispersal is low and asymptotically declines to the Model 1 value as dispersal increases. Average *Ma* mirrors this by increasing with dispersal but is always above Model 1 levels. Average *M* is very slightly higher at low dispersal.

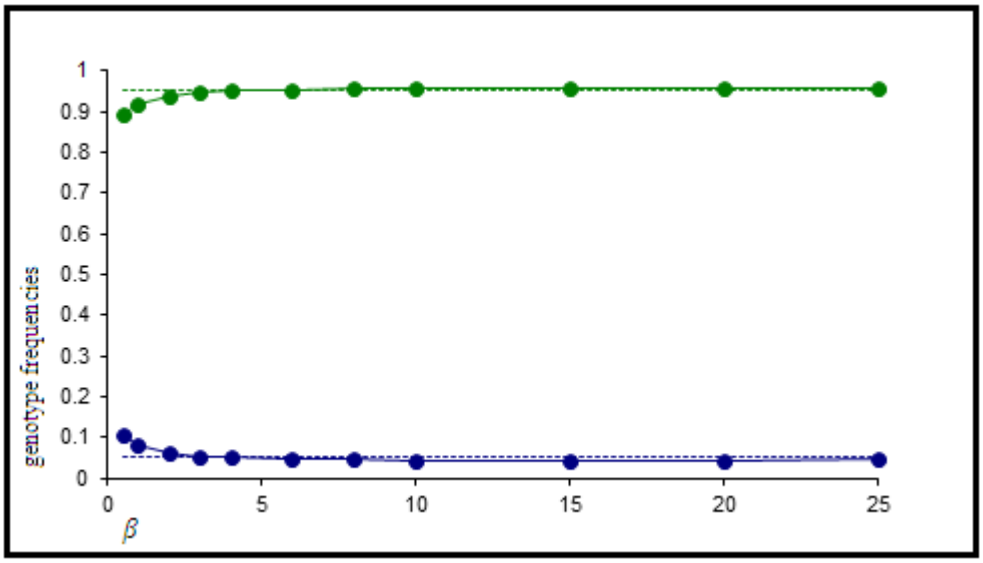

Figure 3.3 – Average values of *R* (blue) and *r* (green) against increasing β. Dotted lines are the predicted levels from Model 1. x-axis is both β and y-axis is frequency. Average R is higher when dispersal is low and asymptotically declines to the Model 1 value as dispersal increases.

As β increases, average cycle length for oscillations decreases asymptotically from 1519 at β=0.5 to 307 at β=25 (figure 3.4, below), while the average amplitude of oscillations increases (figure 3.5, overleaf).

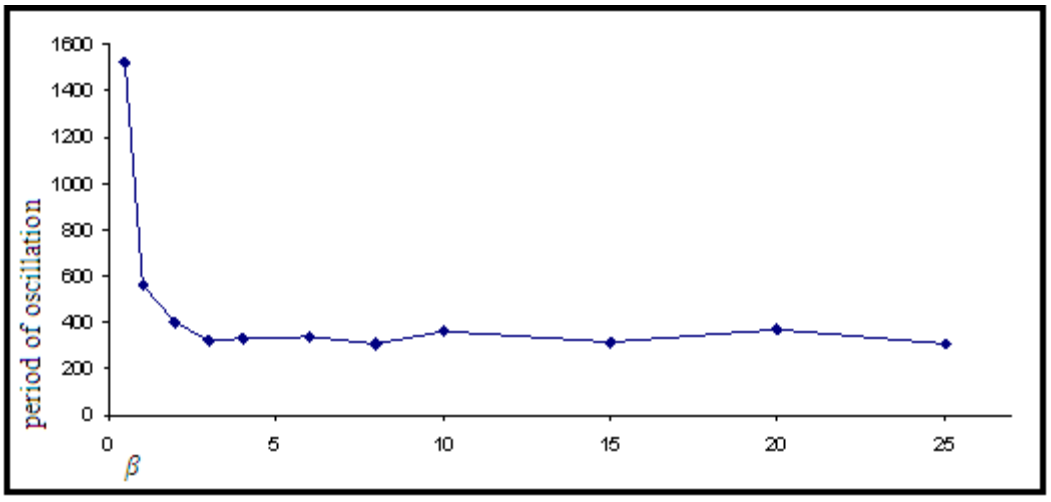

Figure 3.4– Average period of oscillations in genotype frequency. x-axis is both β, y-axis is generations. Average period declines asymptotically with increasing dispersal.

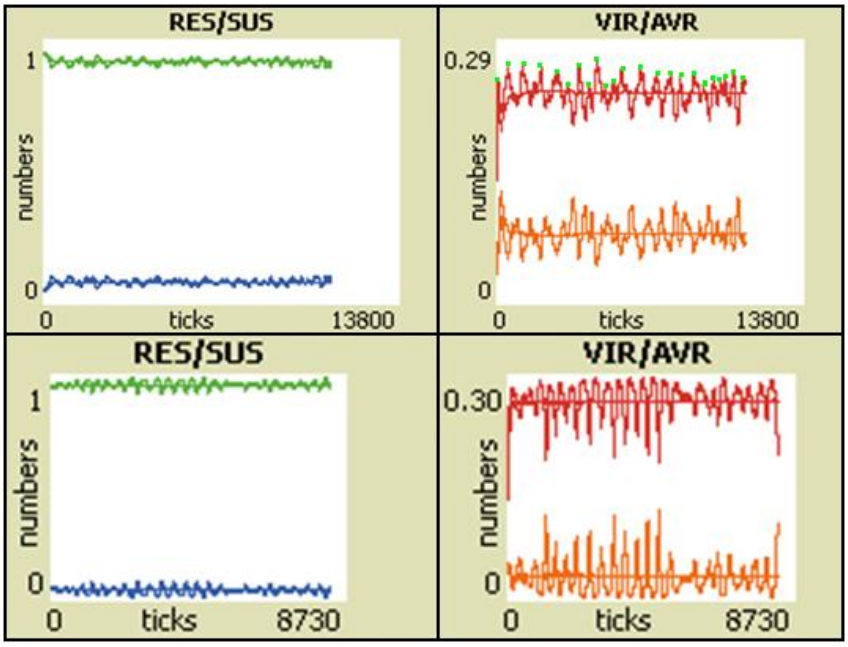

Figure 3.5 – Screenshots from Model 2 showing current and average genotype frequencies (y-axis, labelled numbers) against time (x-axis, labelled ticks) for low and high dispersal (respectively top row, β=1, and bottom row, β=20). Screenshots are from large-arena simulations. Current genotype frequencies are oscillating lines, while average genotype frequencies are smooth lines. All genotype frequencies are measured relative to the arena size, thus host frequencies always add up to 1 and parasite frequencies do not. The frequency of resistant hosts, *R*, is shown in blue and that of susceptible hosts, *r*, in green. The frequency of avirulent hosts relative to the arena size, *MA*, is shown in orange and that of virulent hosts, *Ma*, in red. Oscillations in genotype frequencies are counted manually from peaks in *MA* (green squares in top graph). At higher β, oscillations visibly increase in both frequency and amplitude.

As β increases the stability of polymorphism, measured as average TTF in small arenas, decreases from around 1500 at β=0.5 to approximately 300 by β=4 (figure 3.6). Again, the curve is asymptotic with respect to increasing β.

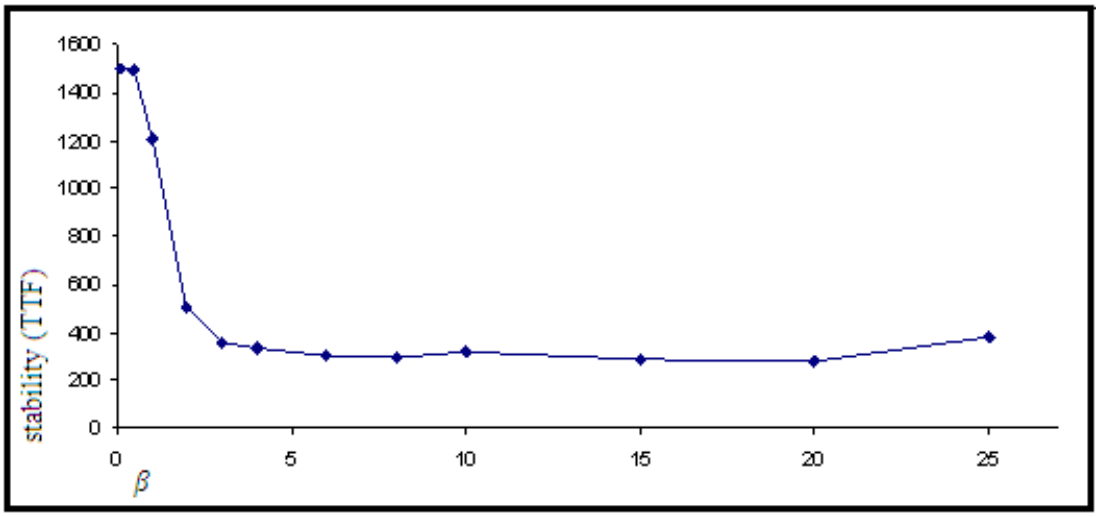

Figure 3.6 – Average stability. x-axis is both β, y-axis is time in generations until loss of polymorphism. Average TTF declines asymptotically with increasing β, similar to average *R* and *MA*.

3.4.2 – Results of Simulations 3.1 explained – altered average genotype frequencies

The increased average *A* and *R* at low β is caused by local adaptation and a resulting increased fitness of *AVR*-parasites, as follows. Low values of β lead to local aggregations of genotypes (Figure 3.1). Regions are dominated by either *AVR*- or *avr*-parasites and either *RES*- or *res*-plants, rather than a mixture of both. *AVR*-parasites cannot reproduce on *RES* hosts, so are mostly found in regions dominated by *res*-hosts and free of *RES*-hosts. Thus on average any given *AVR*-parasite is further away from the nearest *RES*-plant than would be the case with a random distribution of *AVR*-parasites. Therefore *AVR*-spores are more likely to land on *res*-plants than the value of *R* would suggest. Consequently *AVR*-parasites have higher average fitness than in a non-spatial model with the same *R*. This additional fitness alters the balance between *AVR*- and *avr*-genotypes, leading to higher average *A* and lower average *a*. I term this "extra" fitness for *AVR*-parasites.

AVR-parasites occupy space in regions of *res*-hosts at a slightly higher density than *avr*-parasites in regions of either host genotype, owing to the *avr*-parasites' reduced basic rate of reproduction. Thus higher *A* at low β corresponds to slightly higher total disease incidence *M*. Higher average *A* causes higher average *M*, as there are more *AVR*-parasites relative to the number of patches in the model.

Increased average *MA* at low β leads to increased average *R*. *RES*-plants are fitter than *res*-plants in regions with *AVR*-parasites, as *res*-plants suffer the cost of disease *s* from successful *AVR*-infection and *RES*-plants do not. Thus *RES*-plants invade regions of *AVR*-parasites. Conversely in regions with *avr*-parasites (or no parasites) *RES*-plants have no fitness advantage over *res*-plants and are at a disadvantage because of the cost of resistance *u*, so they are displaced by *res*-plants. Thus the increased *A* at low dispersal means there are more *AVR* areas for *RES*-plants to invade and leads to higher *R*. However, when *RES*-plants invade they very quickly wipe out the local *AVR*-parasites and are in turn displaced by *res*-plants. Therefore *RES*-plants only benefit from *AVR*-regions transiently and average *R* increases less than average *A*.

As β increases, local aggregations of genotypes break down and *AVR*-parasites no longer benefit from the increased probability of landing on *res*-hosts. The average fitness of *AVR*-parasites decreases and the *avr*/*AVR* balance shifts in favour of *avr*-parasites, increasing

average *a* at the expense of average *A*. The reduction in *AVR*-parasites reduces the fitness of *RES*-plants relative to *res*-plants because *RES*-plants only have a fitness advantage over res-plants in regions of high *AVR* incidence. This decreases average *R*. As *avr*-parasites produce fewer spores, reduced average *A* also reduces average disease incidence. High dispersal makes the system closer to a panmictic model and at high β the system's average *R*, *r* and *MA* (but not *Ma*, *M* or *A*) match the IEP from the non-spatial Model 1. *Ma* is in all cases higher than predicted from the panmictic Model 1. This is presumably due to the spores from widely-dispersed parasites being less likely to land on the same host and exclude one another.

3.4.3 – Results of Simulations 3.1 explained – oscillation dynamics

Increased β leads to shorter periods and increased amplitude of oscillations in genotype frequencies. This can be explained as follows. High β leads to high dispersal, which means separate regions are more strongly connected. When β is low, it takes many generations for descendants of organisms in one location to travel across the arena. To an extent regions will have independent oscillation dynamics which will tend to average out at a global scale, favouring unchanging or minimally changing genotype frequencies. Global dynamics will still emerge but global oscillations will propagate more slowly through the arena, reducing oscillation frequency, and will do less to suppress all the local oscillation dynamics, reducing oscillation amplitude. As β increases and propagules spread faster, global dynamics experience increased frequency and amplitude.

Tellier & Brown (2011) provided an equation for oscillation period in terms of costs *b*, *c*, *u* and *s*. To use this equation with my models it would have to be altered to account for variable incidence, probably by multiplying *s* by $\widehat{M}$ or average *M* (for Models 1 and 3 respectively). For Model 3, a term or terms reflecting dispersal in both species would also be required.

Low dispersal extends oscillation periods. Low dispersal also leads to self-shading, that is seeds and spores landing on or close to their parent patch and competing mainly with siblings, which blunts competition between genotypes within each species.

3.4.4 – Results of Simulations 3.1 explained – stability measured as average TTF

Increased stability at low β corresponds to both altered genotype frequencies and altered oscillation dynamics. Logically, this makes sense as there appear to be two ways that increasing dispersal could reduce average stability. Increased β both reduces the equilibrium values of *R* and *MA*, the frequencies of the rarer genotypes which typically go extinct, and increases the frequency and amplitude of oscillations. Loss of genotypes is a stochastic event which is most likely to occur when genotype frequencies are close to zero, so lower equilibria and more frequent and extreme oscillations could both make such loss more likely. However results from Simulations 3.2 (Section 3.4.5, below) suggest that reduced equilibrium values of *R* and *MA*, rather than altered oscillation dynamics, is the main factor reducing stability.

3.4.5 – Results of Simulations 3.2, varying β_h and β_p separately - low β_h and to a lesser extent β_p close to β_h increase average *A*, *R* and stability, while increasing either β_h or β_p increases frequency and amplitude of oscillations

Average genotype frequencies are affected differently by varying β_h and β_p (figure 3.7, overleaf). With increasing β_h, average *R* and *A* decrease. These trends follow an asymptote, coming close to their limit by $\beta_h = 4$. Thus the effects of varying β_h while holding β_p constant are similar in direction, but slightly less in magnitude, to the effects of varying β.

In contrast the effects of varying β_p are smaller in magnitude and more complex, depending on the relative values of β_h and β_p. For any given value of β_h, the highest average values of *R* and *A* tend to occur when β_p is on or near this value. Increasing the difference between β_p and β_h, rather than increasing or decreasing β_p *per se*, tends to result in lower average *R* and *A*. Thus increasing the difference between β_p and β_h has similar, although less pronounced, effects to increasing both β or β_h.

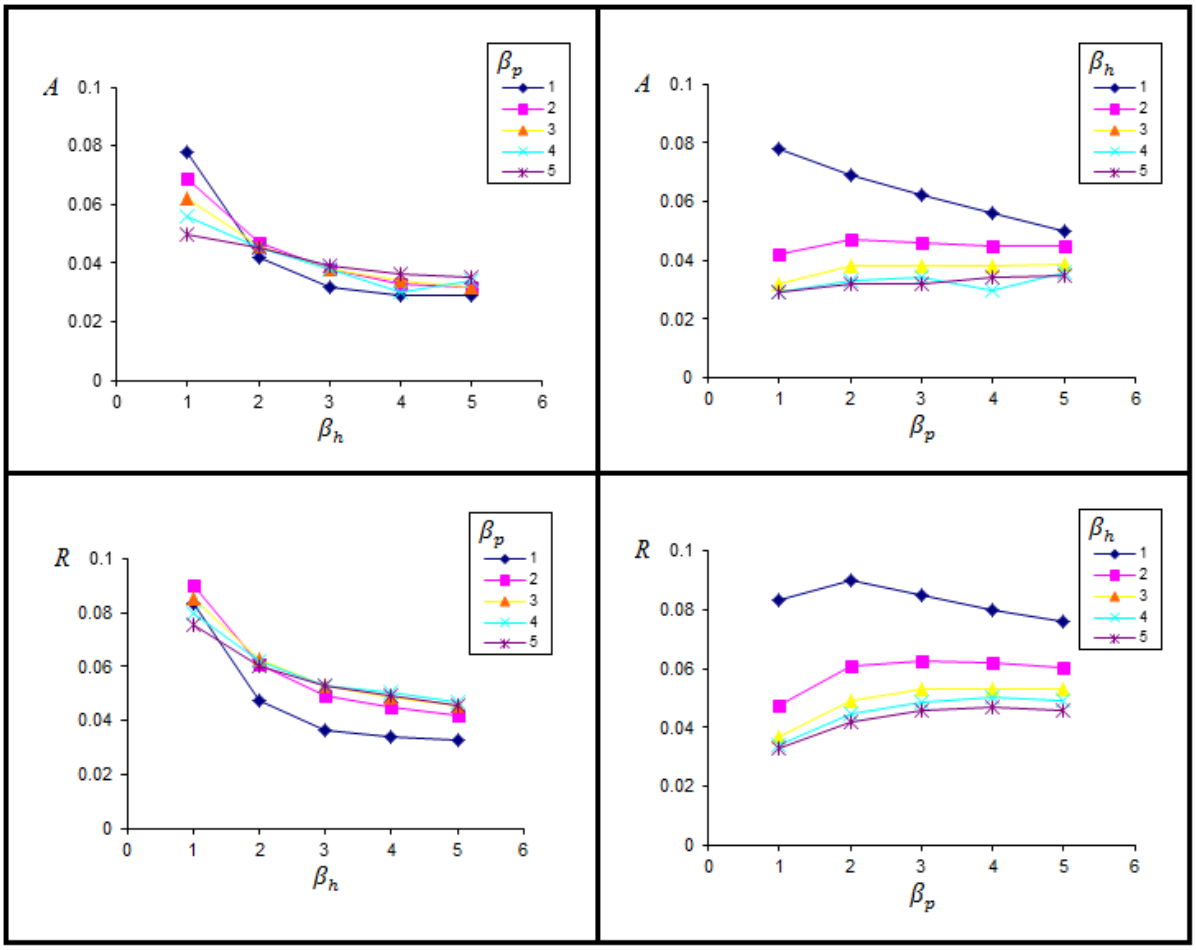

Figure 3.7 – Effect of varying β_h and β_p separately on average genotype frequencies *A* and *R*. x-axis are β_h (panels A and C, series are β_p) or β_p (panels B and D, series are β_h). y-axis are *A* (panels A and B) or *R* (panels C and D). Increasing β_h causes average *A* and *R* to decline asymptotically, similar to increasing both β (Section 3.41). The effect of increasing β_p depends on the value of β_h as described above.

Average oscillation period decreases in response to increasing either β_h or β_p (figure 3.8, overleaf). It approaches an asymptotic limit, much as it does with increasing both β, although the decrease for only one β is slightly less pronounced.

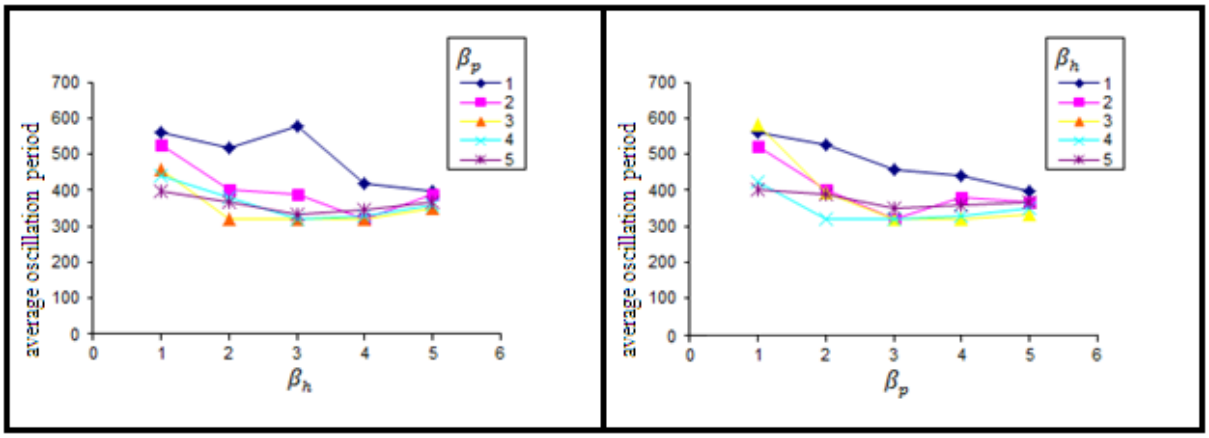

Figure 3.8 – Effect of varying dispersal distances on average period of oscillation in generations. x-axis is β_h (first panel, series are β_p) or β_p (second panel, series are β_h). y-axis is average period of oscillation Increasing β_h causes average period of oscillation to decline asymptotically, similar to increasing both β (Section 3.4.1). Increasing β_p has a similar but less pronounced affect.

Stability (average TTF) is affected differently by varying β_h and β_p (figure 3.9, overleaf). Increasing β_h decreases average TTF to an asymptotic limit regardless of β_p. In contrast changing β_p, whether up or down, to increase the difference from β_h decreases TTF. Thus increasing β_h is analogous to increasing both β and the effect is almost as strong, while increasing the difference between β_p and β_h is also analogous but much weaker.

A key point is that the effects on stability of varying β_h and β_p separately follow the same pattern as the effects on average genotype frequencies *R* and *A*, rather than the effects on oscillation length. This suggests increased stability at low dispersal has more to do with altered genotype frequencies than altered oscillation dynamics.

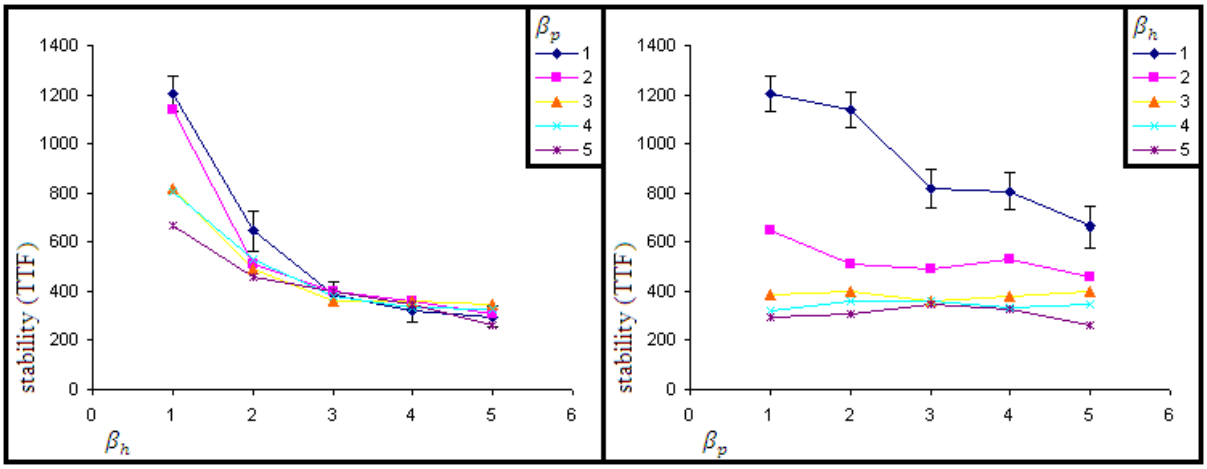

Figure 3.9 – Effect of varying dispersal distances on average stability measured as TTF in generations. Error bars are standard errors, shown on $\beta_p = 1$ (Panel A) and $\beta_h = 1$ (Panel B) only. Increasing β_h causes average stability to decline asymptotically, similar to increasing both β. Increasing β_p causes a decline in stability when $\beta_h = 1$ but no clear effect otherwise.

3.4.6 – Results of Simulations 3.2 explained – local adaptation of *AVR*-parasites and thus increased average *A*, *R* and stability depends on β_h being low but on β_p being close to β_h rather than low

As reported in Section 3.4.2, the altered genotype frequencies at low dispersal levels in Simulations 3.1 are due to increased fitness of *AVR*-parasites as their spores are less likely to land on *RES*-plants than would be the case in a highly dispersing system with the same global genotype frequencies. Simulations 3.2 show that increasing β_h always breaks down this pattern, just like increasing both β. However, the effects of changing β_p are not so simple and depend on β_h. If β_p is higher than β_h, spores disperse further than seeds and lose some of their local-adaptation advantage. If β_p is lower than β_h than spores are also less likely to land on the offspring of their parents' host than if spores had the same dispersal profile as those seeds. Again, if spores are less likely to land on the offspring of the plant their parental parasite parasitized then the local adaptation experienced by *AVR*-spores is less likely to persist. Thus for any β_h *AVR*-parasites will enjoy the greatest local adaptation, with the greatest consequent increase in average *A* and *R*, when $\beta_p = \beta_h$. Therefore making β_p more different from β_h has a similar, negative, effect on average *A*, *R* and incidence and thus on system stability as increasing β_h or both β.

This effect of increasing the difference between β_p and β_h is weaker than the effect of increasing β_h. Presumably at high β_h even similar β_p only slightly increases the chances

of *AVR*-spores landing on *res*-hosts so the advantage to the parasite of having similar rather than dissimilar dispersal is limited. Conversely at low β_h hosts will always be clustered and *AVR*-spores that disperse beyond clusters of *res*-hosts will always fail to reproduce. Thus so there will be no multi-generation build-up of locally maladapted *AVR*-spores so the impact on average frequencies of over-dispersing *AVR*-spores will be limited.

In contrast to the effect on average genotype frequencies, similarity of β_p to β_h has no impact on the period or magnitude of oscillations in those frequencies. This is because increasing either dispersal parameter will always increase system connectivity, thus shortening period and increasing magnitude of oscillations, regardless of the other parameter.

3.4.7 –Results of Simulations 3.3, effect of varying β_h and β_p on size of stable zones – low dispersal leads to increased average TTF within stable zones but does not alter the width of stable zones

The results of Simulations 3.3 include both Type 3 and Type 4 stability, as defined in Chapter 2 (Section 2.3.1, Table 2.3). A brief recap follows. Model 1 and Model 3 systems change their behaviour, moving from Type 1 to Type 5, as the parasite basic rate of increase *y* increases relative to other parameters. Type 3 stability has low *y* and thus low disease incidence, fixed *AVR* in parasites and polymorphism for *RES*/*res* in hosts. As *y* increases Type 3 behaviour is replaced by Type 4 behaviour. Type 4 stability has higher but still low to intermediate *y* and disease incidence, with polymorphism in both species.

Outside the zones of Type 3 and 4 behaviour, polymorphism is not maintained. At the higher boundary value of *y* Type 4 behaviour is replaced by Type 5 behaviour, which has the same equilibrium but an unstable IEP and diverging oscillations rather than a stable IEP and converging oscillations. At the lower boundary value of *y* Type 3 behaviour is replaced by Type 2 behaviour, which has fixed *A* but such low *M* that *RES*-hosts are always out-competed by *res*-hosts and go extinct.

In the deterministic Model 1 the changes between Types 2, 3, 4 and 5 systems are abrupt. In the stochastic Model 3, where stability is measured as TTF rather than a binary yes/no, changes are more gradual. Still, for both sets of costs and all dispersal parameters, the following is observed. With increasing *y* there is a sequential shift through rapid loss of

RES-hosts and *avr*-parasites (Type 2 unstable behaviour), to Type 3 stability, to Type 4 stability, to expanding oscillations and genotype loss (Type 5 unstable behaviour). This is shown in Figure 3.10 (Costs A, below) and 3.11 (Costs B, overleaf). Predicted boundaries between Types 2, 3, 4 and 5 behaviour are marked by red lines. A key result is that, while both the overall strengths of stability and the relative strengths of Type 3 and 4 stability (TTF) vary with dispersal, the widths of stable zones (range of *y*) do not.

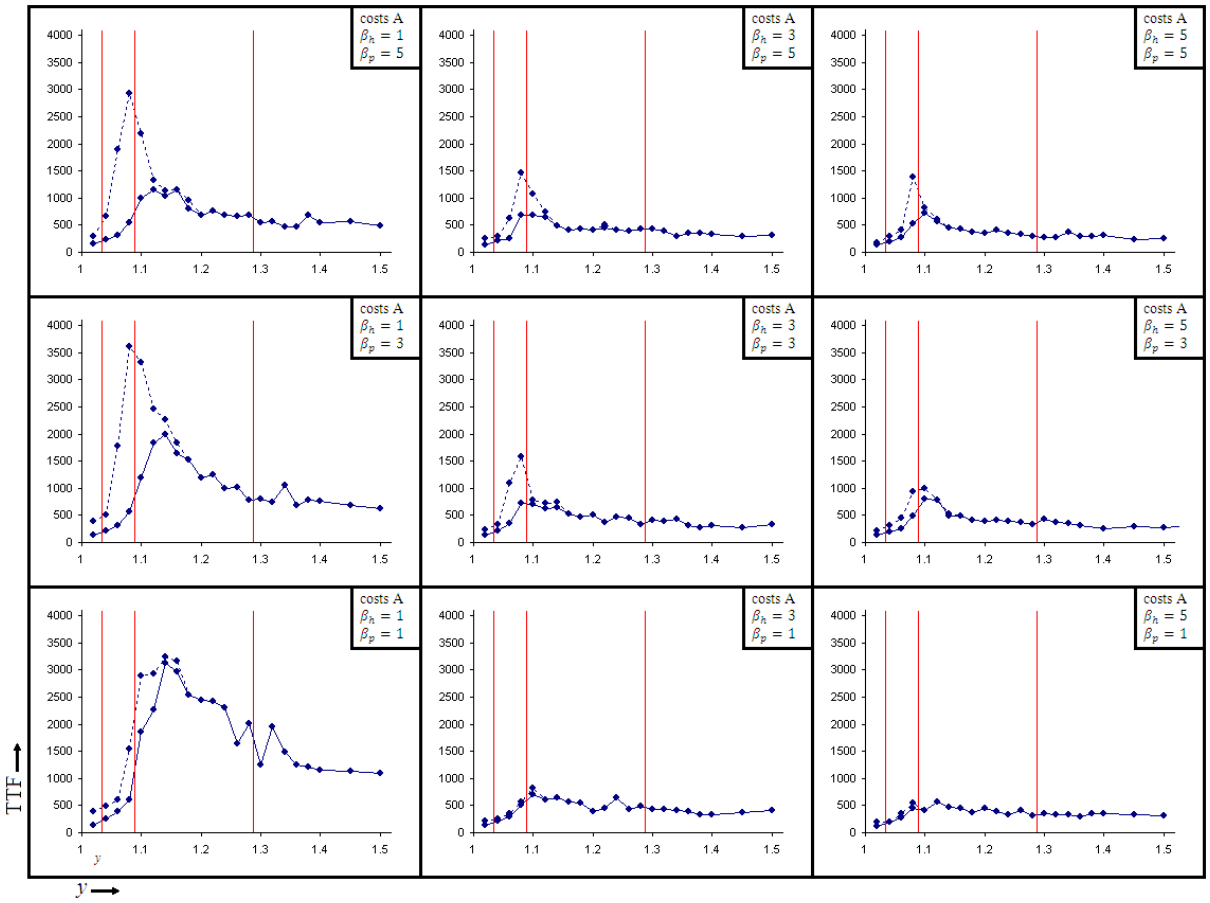

Figure 3.10 – Stable zones against parasite reproductive rate for differing dispersal parameters, costs A (u = 0.01 s = 0.3 b = 0.05 c = 1). y-axes are stability (TTF) and x-axes are parasite basic reproductive rate y. Red lines are predicted lower boundaries for Type 3 polymorphism (at y = 1.0345), between Type 3 and Type 4 polymorphism (y = 1.0889) and upper boundaries of type 4 polymorphism (y = 1.2868) in the fully deterministic Model 1 (Chapter 2). Solid line is type 4 polymorphism and dashed line is type 3 polymorphism. Type 4 polymorphism is most stable when β_h and β_p are low, with both having a similar effect. Type 3 polymorphism occurs and increases in stability when β_p equals or exceeds β_h.

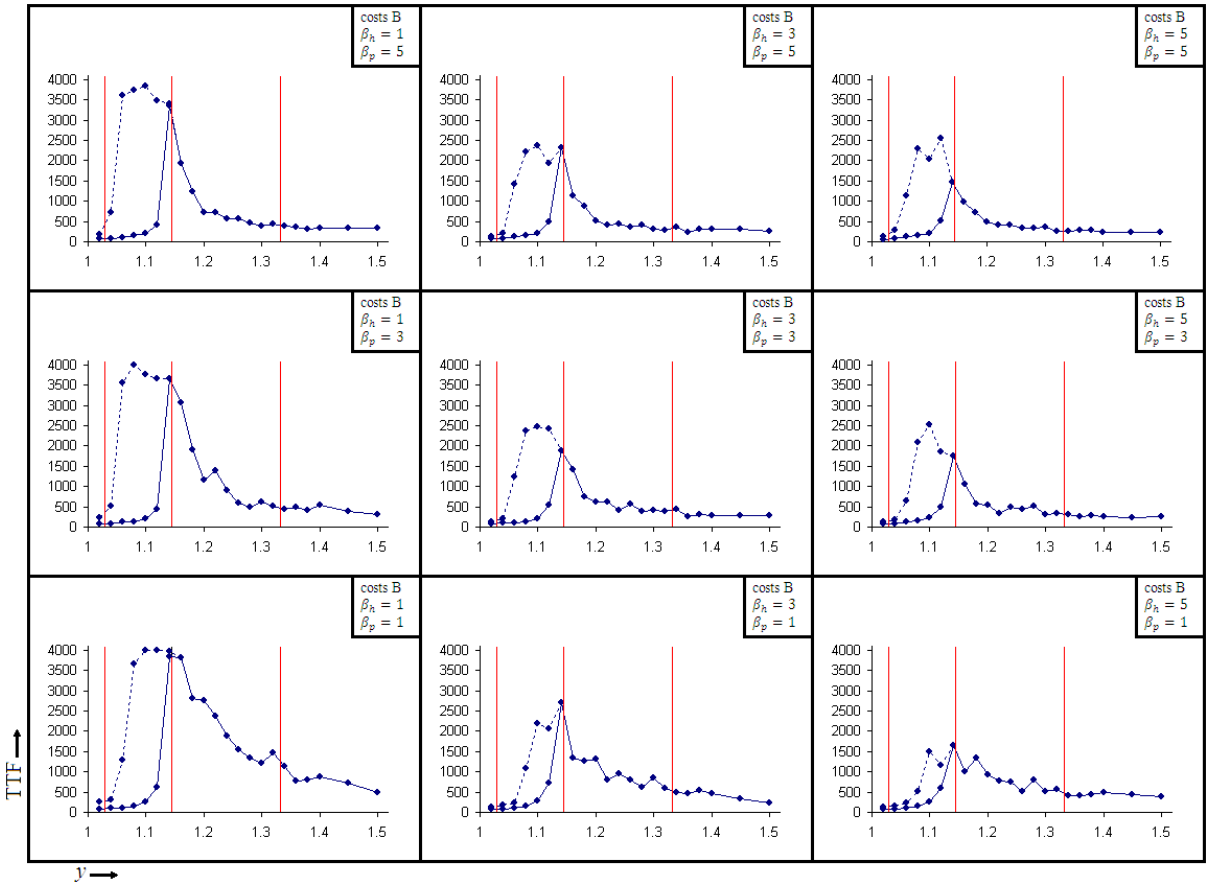

Figure 3.11 – Stable zones against parasite reproductive rate for differing dispersal parameters, costs B (u = 0.02 s = 0.7 b = 0.1 c = 1). y-axes are stability (TTF) and x-axes are parasite basic reproductive rate y. Red lines are predicted lower boundaries for Type 3 polymorphism (at y = 1.0294), between Type 3 and Type 4 polymorphism (y = 1.1438) and upper boundaries of Type 4 polymorphism (y = 1.3324) in the fully deterministic Model 1 (Chapter 2). Solid line is type 4 polymorphism and dashed line is type 3 polymorphism. Type 4 polymorphism is most stable when β_h and β_p are low, with β_h having a much stronger effect. Type 3 polymorphism always occurs but is most stable when β_p equals or exceeds β_h.

3.4.8 – Results of Simulations 3.3 - stability is highest well within the predicted stable zones, near the predicted Type 3/4 border, regardless of dispersal parameters

Types 3 and 4 stability both occur inside the predicted stable zones from Model 1. Type 3 stability begins to be observed around the Type 2/3 boundary but only reaches high levels near the upper bound of the predicted Type 3 zone. Similarly Type 4 stability declines to low levels well before y reaches the predicted Type 4/5 boundary, although TTF continues to decline very slightly through and after this boundary. Generally both types of stability show their highest TTF close to the predicted Type 3/4 boundary. The alteration from abrupt to gradual shifts in stability is expected in an individual-based model, although it is

interesting that stability (heightened TTF) occurs almost entirely within the predicted zones rather than crossing over the predicted boundaries more.

Maximum TTF for both Type 3 and Type 4 stability is higher for Costs B than for Costs A. This is presumably because Costs B lead to a wider predicted Type 3 stable zone. Thus for Costs B the predicted Type 3/4 boundary and the peak stability around that boundary are further from unstable zones. For Costs A maximum Type 3 stability is observed at or just before the predicted Type 3/4 boundary and maximum Type 4 stability is observed after that boundary, whereas for Costs B maximum Type 3 stability is observed before the predicted Type 3/4 boundary and maximum Type 4 stability is observed at or after that boundary. This is presumably due to the wider predicted Type 3 stable zone for Costs B, which means the higher end of the Type 3 region is further away from the unstable Type 2 region so higher stability can occur within the Type 3 region.

3.4.9 – Results of Simulations 3.3 - maximum Type 4 stability declines with increasing β_h, while maximum Type 3 stability exceeds maximum Type 4 stability unless β_p=1

Maximum Type 4 stability (maximum average TTF) declines with increasing β_h regardless of β_p and declines with increasing β_p when β_h = 1. This occurs for both sets of costs. There is no clear pattern effect on stability of increasing β_p when β_h = 3 or β_h = 5. The greatest decrease in maximum stability occurs when β_h increases from 1 to 3. These results are equivalent to those from Simulation Set 3.2. The ways β_h itself and β_p relative to β_h affect stability are discussed in Section 3.4.6, as are the reasons the effect of β_h is more pronounced.

Maximum Type 3 stability (maximum average TTF) exceeds maximum average Type 4 stability for all sets of costs and combinations of dispersal parameters where $\beta_p > 1$. This is statistically significant in nine of the twelve cases (5 p-values below 0.01, 4 more below 0.05, two-tailed heteroscedastic t-tests). When β_p = 1 maximum Type 3 stability does not significantly exceed maximum Type 4 stability (p-values above 0.5 in five of six cases, same t-tests). When β_p = 1 and $\beta_h > 1$, Type 3 stability for Costs A virtually disappears.

Type 3 stability occurs when *AVR*-parasites alone coexist with *RES*- and *res*-hosts. If both species have low dispersal *AVR*-parasites will be locally adapted to hosts, *i.e.* found in regions of RES-hosts. If parasites have high dispersal some AVR-parasite spores will

encounter suitable hosts regardless of host dispersal. However if host dispersal is significantly greater than parasite dispersal, *AVR*-parasites will be unable to adapt to rapidly changing host population structures and *RES*-hosts will be able to take invade regions of *AVR*-parasites. Low basic parasite reproductive rate *y* means that *AVR*-parasites are vulnerable to being wiped out by such *RES*-host incursions.

3.5 - DISCUSSION

3.5.1 – Limited dispersal protects against stochastic rather than deterministic genotype loss

This chapter shows that limited dispersal of seeds and spores changes average genotype frequencies and oscillation dynamics, making polymorphism more likely to persist in a stochastic model population. I suggest these processes in real-world pathosystems could make long-term persistence of polymorphism more likely.

The stabilising effects of limited dispersal work in a fundamentally different way to ndFDS. ndFDS can lead to shrinking rather than growing oscillations and thus opposes **deterministic genotype loss**. Conversely the mechanisms discussed in this chapter, of moving equilibrium frequencies of the rarer genotypes further from zero and of altering oscillations in genotype frequency to have lower frequency and magnitude, work to oppose **stochastic genotype loss**. Although the latter cannot by itself lead to stable polymorphism, it could certainly protect the former against random fixation and make polymorphism last far longer in real pathosystems. Thus, low dispersal and the resulting spatial population structure alone cannot lead to long-term stable polymorphism. At best they can only delay the end of a transient polymorphism (Brown & Tellier 2011). However, when occurring in conjunction with ndFDS, low dispersal can make stochastic loss of genotypes less likely over any given period of time.

This protection against stochastic genotype loss does not depend on the source of ndFDS and only requires limited dispersal to operate. Thus such protection could occur in any system with any source(s) of ndFDS and some degree of genotype aggregation due to limited dispersal.

Model 3 is far from the first spatial model of host-parasite coevolution. Many spatial models in this area have been metapopulations rather than individual-based (Daamgard 1999, Thrall & Burdon 2002, Sasaki 2002). These three respectively included group-selection, mutation and mutation as potential sources of ndFDS, although in the latter two cases it is not clear whether mutation was important or whether the models merely experienced a delay in genotype fixation due to self-slowing dynamics caused by the population structure (Brown & Tellier 2011). A key difference between these studies and Model 3 is that in Model 3 I was able to distinguish between factors opposing deterministic and stochastic genotype loss and explore each separately. This separation of the two classes of factor and analysis of how they interact is the most important general contribution Model 3 makes to coevolutionary theory.

The multi-island metapopulation pathosystem studied by Laine (2005) is a real-world case where the interaction of factors opposing deterministic and stochastic genotype loss might apply. There may be various sources of ndFDS or analogous stabilising factors, including DLR of parasites (Model 1) and multiple parasite generations per host generation leading to auto-infection of hosts (Tellier & Brown 2007A). Laine reports spatial aggregation and local adaptation of parasites at an intermediate scale between demes and the whole metapopulation. This is consistent with parasite dispersal being limited compared to the environment size, allowing a fitness advantage due to increased local adaptation to hosts. This, in conjunction with sources of ndFDS, could help maintain polymorphism. Burdon & Thrall (1999) also found evidence that parasites were adapted at an intermediate scale in a wild pathosystem of wild flax and rust.

Gandon et al. (1996) described a matching-allele metapopulation model in which they varied host and parasite dispersal. They reported that the species with higher dispersal was always locally adapted, while I found that *AVR*-parasites were locally adapted if anything was. This apparent discrepancy is because Gandon et al. studied a wider range of disease costs in both species. Indeed, the subset of their results dealing with the equivalent of $c = 1$ found the same results I reported for simulation set 3.2.

3.5.2 – Other sources of protection from stochastic genotype loss?

The question arises, are there other non-ndFDS generating factors that limit stochastic genotype loss? From this chapter I suggest two candidates – large populations and inherently less dynamic oscillations. In large populations it is less likely that an allele at any given frequency will be lost by random genetic drift. If oscillations are less dynamic, allele frequencies will spend less time near zero and have a lower chance over any given time period of stochastic fixation.

Low dispersal is a specific source of less dynamic oscillations, although as discussed in Section 3.3.2 it is not clear whether this promotes stability in Model 3. Less dynamic oscillations also occur, for example, in Type 3 as opposed to Type 4 dynamics and in Type 4 dynamics with lower as opposed to higher parasite basic reproductive rate. This is shown in Figure 3.12, below. This occurs because lower *y* results in stronger DLR and more rapid convergence to the IEP. Lower parasite reproductive rate was also cited as promoting stability in Gandon et al. (1996) and May & Anderson (1983).

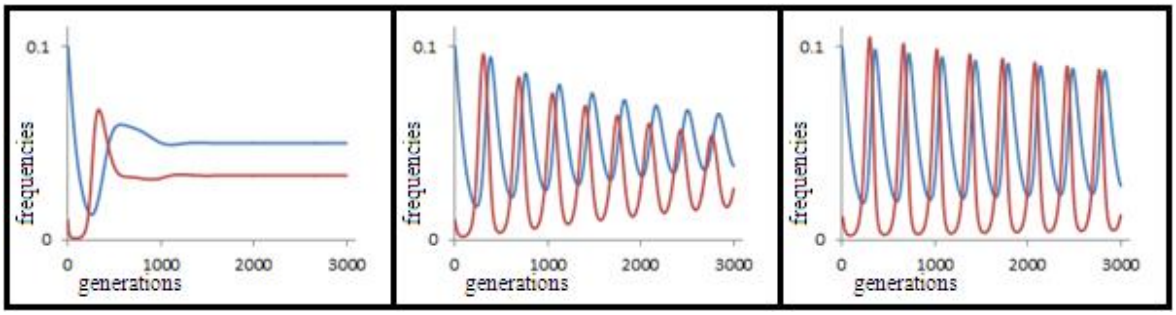

Figure 3.12 – oscillations in *R* (blue) and *A* (red) for different values of *y* from Model 1 (Chapter 2). x-axes are generations, y-axes are frequencies *R* and *A*. Costs are $u = 0.01$, $s = 0.3$, $b = 0.05$ and $c = 1$. Panel A is a Type 3 system ($y = 1.1$), panel B is a Type 4 system with lower *y* (= 1.2) and panel C is a Type 4 system with higher *y* (= 1.3). Oscillations are damped more rapidly in the lower *y* as opposed to the higher *y* type 4 system, which would make stochastic genotype loss less likely in a real-world population. Oscillations are entirely absent in the type 3 system, which would make stochastic genotype loss even less likely in a real-world population. A similar difference in oscillation magnitude can be seen between low-dispersal and high-dispersal systems (Figure 3.5).

The relative strength of ndFDS, and analogous processes such as DLR, is important for both deterministic and stochastic gene-loss. Deterministically a system is either stable or unstable. Stochastically stronger ndFDS leads to faster oscillation damping so genotype loss is less likely. This is why the stable zones in figures 3.10-3.11 have sloping

boundaries and low stability near the stable/unstable boundary values; as oscillations move from converging towards diverging they converge more slowly and stochastic genotype loss becomes more likely.

3.5.3 - Complexity revisited - multiple factors preserving polymorphism

It has been suggested that complex interacting factors are necessary to promote stable polymorphisms of immunity-antigenicity genes (Thrall & Burdon 2002; Brown 2003B). More recently Tellier & Brown (2007A) demonstrated simpler theoretical requirements and consequently simpler biological mechanisms for stable polymorphism. However it may be the case that, while not actually required for stable polymorphism, multiple factors frequently *do* occur together and their interaction strengthens stability.

Interacting factors preserving polymorphism could include multiple non-spatial sources of ndFDS (Tellier & Brown 2007A, Model 1 Chapter 2), oscillation damping analogous to ndFDS due to spatial variation in selection pressures (Tellier & Brown 2011, Model 2 Chapter 2) and, as shown in this chapter, non-ndFDS factors that oppose stochastic rather than deterministic genotype loss. Such multi-factor interactions have important implications for the stability and commonness of polymorphism in host-parasite coevolution because in real, finite populations, deterministically stable cycles that spend long periods of time with genotype frequencies close to zero are likely to experience stochastic genotype fixation.

3.5.4 - Limited dispersal helps preserve diversity in coevolutionary dynamics other than GFG

Brockhurst et al. (2006) reported that in bacteria-phage systems non-homogenised environments can lead to spatial refugia that preserve polymorphism in host resistance. Koskella & Lively (2007) reported that zones of sexual and asexual phenotypes are found in freshwater snails. These zones depend on limited dispersal of both the snails and the trematode parasites that select for sexual reproduction. While limited dispersal may lead to genotype aggregation in any species, host or parasite, the aggregation of one species can cause the over-dispersal of another. Packer & Clay (2000) reported that black cherry seedlings suffer increased mortality close to their parent trees due to concentrated *Pythium* oomycete parasites. This study is interesting as an example of long-term aggregation of

parasites leading to over-dispersal of hosts; the parasite thrives on the adult roots without killing the tree but this leads to an increase in parasite density that wipes out locally landing seedlings and drives increased diversity in the tree community. Augspurger (1983) found the same relation between distance from parent and seedling mortality in the tree *Platypodium elegans* due to damping-off. If limited dispersal in one species causes over-dispersal in another it presumably promotes greater community diversity at the level of that second species, in these cases at the level of the tree community in woodlands. Increased species diversity, meaning reduced density of any specific host, would itself lower parasite effective reproductive rate and strengthen DLR.

3.5.5 - Hypothesis – limited dispersal, or by analogy extended spatial scale, will alter genotype frequencies and thus make polymorphism more likely to persist in isolated populations of the same or closely related pathosystems

I make the following predictions about real-world pathosystems. When a host and parasite have low and (to a lesser extent) similar scales of dispersal, genotypes will be more clustered. Clustering will lead to increased average local adaptation, and thus to increased fitness and global equilibrium frequency, for *AVR*-parasites as the genotype most affected by genotypes in the other species. This will lead to higher equilibrium frequencies of *RES*-hosts and, because polymorphism is normally lost when *AVR*-parasites or *RES*-hosts go extinct, a higher likelihood of sustained polymorphism. The latter can be measured by the frequency of populations in which polymorphism is observed.

Testing these predictions would require comparing related pathosystems that differ primarily in dispersal distances. Alternatively, since appropriately similar pathosystems are rare, fully isolated populations of one pathosystem of different spatial sizes could be compared. Larger spatial size for a given population of a pathosystem is analogous to reduced dispersal. I would therefore expect more spatially extended populations to have increased average *A* and *R* and increased stability, matching simulation results of reduced dispersal (figures 3.2-3.3). Again, relative longevity of polymorphism in different-sized populations could be inferred from measuring the frequency of such populations in which polymorphism is currently extant. More spatially extended populations may well be larger populations, which itself favours stability, so population range and number would have to be separated in any such studies.

CHAPTER FOUR – VARIABLE HOST INCIDENCE AND A SECOND PARASITE IN SPATIAL INDIVIDUAL-BASED MODELS

OVERVIEW

Coevolution of a single host-parasite pairing does not occur in isolation. Numerous real-world factors interfere with pathosystems and can alter coevolutionary outcomes. In previous chapters I showed that density-dependent regulation of parasite incidence is a stabilising factor, in both simple deterministic and spatially extended stochastic models. Here I add further realistic factors to my spatial models and analyse their effects on coevolution. The factors addressed in this chapter are variable host incidence and a spatially heterogeneous second parasite. Variable host incidence both increases the strength of density-dependent regulation on avirulent parasites, favouring stability, and reduces host and parasite population sizes, favouring stochastic extinction. Overall this results in slightly increased stability, measured as average time to fixation in small arenas, when host dispersal is low. A spatially heterogeneous second parasite divides the arena into two different environments characterised by different fitnesses of both host and gene-for-gene parasite. If these environments occur in patches large enough relative to dispersal distances to have their own internal dynamics, transfer between environments occurs and leads to intermediate dynamics. In my models both of these factors only slightly alter stability compared to equivalent results from Model 3. However this may be due to the small size of the populations I modelled. I argue that both factors will promote stability more strongly in many pathosystems in the real world.

4.1 - INTRODUCTION

4.1.1 – Model 4 rationale - variable host density could affect parasite effective basic reproductive rate enough to change coevolutionary outcomes

Populations of hosts and prey are regulated in a density-dependent fashion by their exploiters. The higher the population density, the easier it is for predator populations or disease levels to increase and inflict greater costs on the host or prey population. A classic predator-prey study described the linked oscillations between snowshoe-hare and lynx populations (reviewed in Krebs 2001), while disease studies have shown links between host density and disease incidence (Holt 1985, Kohler 2001). In previous chapters I showed that variable parasite incidence can lead to stable polymorphism in gene-for-gene (GFG) coevolution by providing a source of density-limited reproduction (DLR). DLR has a stabilising effect on both parasite genotypes similar to negative direct frequency-dependent selection (ndFDS), discussed in Chapter 1 (Section 1.1.6). In this situation, the effect of variable host density can be critical because parasite effective basic reproductive rate and thus parasite incidence depend on host density. If host density decreases as parasite incidence increases, density-dependent regulation of parasite incidence is strengthened. In Model 1 (Chapter 2) I showed that relatively small changes in parasite incidence can alter the outcome of coevolution. Both modelling (Holt 1985) and observational (Kohler 2001) studies show that disease-regulated host density can cause much larger changes in parasite incidence.

In spite of its ubiquity and importance, variable host density is neglected in most modelling studies including Leonard (1977), Sasaki (2000), Salathe (2005), Segarra (2005) and Tellier & Brown (2007A, 2011). Exceptions are the metapopulation models in Thrall & Burdon (2002) and Damgaard (1999). Thrall allowed fully variable, disease-regulated host-density in each deme. Damgaard modelled local extinction and establishment of hosts in each deme, representing a change in global host density although not host density in a specific deme. Both models reported stable coevolution. However both models are problematic. Thrall's model is likely to be a case of prolonged transient polymorphism rather than true stability (Brown & Tellier 2011) and includes mutation between genotypes, which is problematic as discussed in Chapter 1 (Section 1.2.3). Damgaard's model has as its source of ndFDS the extinction of low-fitness host demes, theoretically

due to secondary disease ravaging the population. While plausible, in this model such extinction applied at an arbitrary cut-off fitness value imposed by the modeller.

In this chapter I introduce Model 4. This adds to the spatially explicit variable host-density models in two ways, by using an individual-based instead of a metapopulation model and by modelling a source of DLR (variable disease incidence) that is conceptually simple, biologically realistic, mathematically tractable in deterministic models and preserves long-term polymorphism rather than introducing new alleles via mutation.

4.1.2 – Model 5 rationale - a second, non-GFG parasite can change fitness for both hosts and GFG-parasites in patches where it is present

An important feature of real-world biology which has previously been almost completely neglected in theoretical models (local extinction caused by secondary parasites in Damgaard 1999 aside) is that multiple parasite species typically infect one host population. Gathering data and making general predictions about this phenomenon is difficult owing to the complexity of fitness outcomes in disease complexes, but it is likely to be important. Additional diseases can affect both host fitness and fitness of the GFG-parasite. By altering the magnitude of fitness costs in a dynamic, spatially and temporally variable way, additional parasite species can affect the outcome of coevolutionary dynamics between the host and a parasite which has a GFG interaction with that host. In this chapter I introduce Model 5. This is an individual-based spatially explicit model of a gfg-pathosystem where a second parasite, present in some but not all patches, alters the fitnesses of both the host and gfg-parasite.

4.1.3 – Biology behind multi-parasite interactions in Model 5 - Host defensive pathways responding to multiple parasites can up-regulate or down-regulate one another, leading to synergy or antagonism between immune responses

Different aggressors affect plants in different ways. The key signalling molecules are jasmonic acid (JA, often acting in concert with ethylene) and salicylic acid (SA) (Raymond & Farmer 1998, Thaler 2012). SA induces resistance against biotrophic parasites and some phloem-feeding insects and JA induces resistance against necrotrophic parasites, other phloem-feeding insects and chewing herbivores. SA and JA activate (mostly different) parasite-related genes, both in response to parasite elicitors and when exogenous

SA/JA is added experimentally. Different parasites induce either a JA-mediated or an SA-mediated response, rarely both.

Generally, pathways induced by SA decrease the effectiveness of JA-induced defence and *vice-versa* (Glazebrook 2005). SA can negatively regulate JA levels (Thaler 2012) and reduce the efficiency of JA-induced defence. Thaler et al (2002) found an SA analogue increased growth-rates of one caterpillar species on tomato plants but had no effect on other herbivore species. The same paper reported the SA analogue had opposite effects on *Pseudomonas syringea* lesion sizes in wild and domestic tomatoes, suggesting that crosstalk between SA and JA is more complex than pure antagonism. Still, it may be that early infection with an SA-inducing parasite could impair a JA-dependent response against a later parasite and *vice versa*.

Potentially, two simultaneous or at least overlapping infections which induce the two pathways could both benefit from this antagonistic cross-talk, to the detriment of the host. Conversely if two infections induce the same defence pathway the level of the response could be increased, to the detriment of both infections and the benefit of the host. Another way parasites could affect each other is direct competition for resources – this would require the parasites to colonise the same tissue and both be relatively aggressive, but this can happen with many foliar parasites. However host-mediated parasite interactions are more common. Thus between-parasite interactions could be synergistic, increasing the fitness of one or both parasites and the total cost of infection to the host, antagonistic, decreasing the fitness of one or both parasites and the total cost of infection to the host, or something in between, for example neutral so having no effect on parasite fitness or total disease cost.

4.1.4 – Models 4 and 5 add realistic detail to Model 3, respectively introducing variable host incidence and a second non-GFG parasite

Model 3 (Chapter 3) is a simplified version of spatial reality. Unrealistic assumptions include constant, total host presence and a single parasite species. Here I present more complex models that begin to address these issues. Models 4 and 5 are both extensions of the individual-based and spatially explicit Model 3. Model 4 allows variable host numbers, so a given patch can be occupied by a *RES*- or *res*-host or by no host. Model 5 introduces a second parasite species, a generalist with no gene-for-gene interaction, that by

its presence can alter the fitness of both focal species. Both factors can alter the coevolutionary outcomes of models.

Models 4 and 5 can affect the stability of polymorphism in two ways. Model dynamics will either change the strength of the stabilising factor (in these models DLR) and thus affect deterministic genotype loss, or they will alter the probability of chance genotype fixation and so affect stochastic genotype loss. Altering effective parasite basic reproductive rate to alter the strength of DLR is an example of the former (Model 1, Chapter 2). Varying dispersal distances to alter equilibrium genotype frequencies and oscillation dynamics is an example of the latter (Model 3, Chapter 3).

It is possible for one additional factor to affect both these processes. An example is the variable host incidence in Model 4. Variable host incidence means increased parasite incidence will reduce host incidence, thus effectively reducing parasite basic reproductive rate as parasite incidence increases. This will make DLR on parasites stronger and so increase the range of conditions under which deterministic stability will occur. Simultaneously variable host incidence will mean smaller populations of both host and parasite, as variable host incidence is replacing fixed and total host incidence. This will make stochastic genotype loss more likely in both species.

4.2 - MODELS AND METHODS

4.2.1 – Overview – Models 4 and 5 are both expansions of Model 3 and both model individual hosts and parasites in a spatially extended system

Models 4 and 5 are based on Model 3 (Chapter 3). Thus, both use most of the parameters defined for Model 3 (Chapter 3, Section 3.2.3, Table 3.1) and both calculate fitnesses for individual hosts and parasites rather than having an average fitness for each genotype. The constants and variables used in these models are defined in Tables 4.1 and 4.2, respectively, on the next two pages. Unless otherwise noted a constant or variable applies to both models. Fitnesses for individuals of each genotype of each species, in all possible interactions, are given for Models 4 and 5 in Tables 4.3 and 4.4 respectively. A detailed description of how Model 4 differs from Model 3 is given in Section 4.2.2. Similarly a detailed description of how Model 5 differs from Model 3 is given in Sections 4.2.3-4.2.5.

Name	Description	Typical values
u	Fitness cost to plant of having a resistance gene	0.01
MODEL 4 *s*	Fitness cost to plant of being diseased by the gfg-parasite, paid by *res*-plants with any gfg-parasite and *RES*-plants with *avr*-parasites	0.3
MODEL 5 s_1	Cost to host of infection by gene-for-gene parasite only -paid by *RES*-hosts with *avr* gfg-parasite only and *res*-hosts with any gfg-parasite only	0.3
MODEL 5 s_2	Cost to host of infection by generalist parasite only -paid by all hosts with gen-parasite only and *RES*-hosts with *AVR* gfg-parasite and gen-parasite	0.3
MODEL 5 s_3	Cost to host of infection by both parasites -paid by *RES*-hosts with *avr* gfg-parasite and gen-parasite and *res*-hosts with any gfg-parasite and gen-parasite	0.3-0.8 0.51 if no parasite interaction
b	Fitness cost to gfg-parasite of being *avr*	0.05
c	Fitness cost to gfg-parasite of being detected, paid by *AVR*-parasites on *RES*-plants	1
MODEL 4 y_h, y_p MODEL 5 y_p	Basic reproductive numbers for host and gfg-parasite (Model 4) and gfg-parasite only (Model 5) respectively	1.25-1.85
β_h, β_p	Dispersal parameters for host and gfg-parasite respectively	0.5-25
θ_{gen}	Effects of coinfection with the gen-parasite on gfg-parasite fitness	0.7-1.3
MODEL 5 m_{gen}	Species incidence for the gen-parasite, the fraction of patches with gen-parasites -all incidences are calculated relative to the number of patches in the arena, not the number of hosts	0.31-0.57
-	Size of NetLogo arena	63-by-63 or 255-by-255

Table 4.1 – Constants in Models 4 and 5. Constants are used in both models unless specifically noted as Model 4 or Model 5.

Name	Description	Typical values
	Individual variables	
$\boldsymbol{W_h}$, $\boldsymbol{W_p}$	Fitness of individual plant and gfg-parasite, expressions given in Tables 4.3 and 4.4	≤ 1
MODEL 4 $\boldsymbol{\sigma_h}$, $\boldsymbol{\sigma_p}$ MODEL 5 $\boldsymbol{\sigma_p}$	Seed- or spore-numbers of hosts and gfg-parasites (Model 4) and gfg-parasites only (Model 5) respectively -seed or spore number is calculated by multiplying fitness by basic reproductive number and rounding up or down probabilistically	$y_i W_i$ rounded
	Global variables	
R	Fraction of hosts *RES* at current time-step	
r	Fraction of hosts *res* at current time-step; $R + r = 1$	
A	Fraction of gfg-parasites *AVR* at current time-step	
a	Fraction of gfg-parasites *avr* at current time-step; $A + a = 1$	
MODEL 4 M_h, M_p MODEL 5 M_p	Species incidences, the fraction of patches with an individual of a given species, for hosts and gfg-parasites (Model 4) and gfg-parasites (Model 5) respectively at current time-step -all incidences are calculated relative to the number of patches in the arena, not the number of hosts	
M_pA, M_pa MODEL 4 M_hR, M_hr	Genotype incidences, the fraction of patches with an individual of a given species and a given genotype, for *AVR*- and *avr*-parasites (both models) and *RES*- and *res*-hosts (Model 4) respectively at current time-step -all incidences are calculated relative to the number of patches in the arena, not the number of hosts	

Table 4.2 – Variables in Models 4 and 5. Individual variables apply to each patch, while global variables apply to the whole system. Variables are used in both models unless specifically noted as Model 4 or Model 5.

host **parasite**	***RES*** ***AVR***	***RES*** ***avr***	***RES*** **none**	***res*** ***AVR***	***res*** ***avr***	***res*** **none**	**none** ***AVR***	**none** ***Avr***	**none** **none**
host fitness $\boldsymbol{W_h}$	$1-u$	$(1-u)$ $(1-s)$	$1-u$	$1-s$	$1-s$	1	–	–	–
parasite fitness $\boldsymbol{W_p}$	$1-c$	$1-b$	–	1	$1-b$	–	0	0	–

Table 4.3 – Fitnesses in Model 4.

host **gfg-parasite**	***RES*** ***AVR***	***RES*** ***avr***	***RES*** **None**	***res*** ***AVR***	***res*** ***avr***	***Res*** **None**
without gen-parasite						
host fitness $\boldsymbol{W_h}$	$1-u$	$(1-u)$ $(1-s_1)$	$1-u$	$1-s_1$	$1-s_1$	1
gfg-parasite fitness $\boldsymbol{W_p}$	$1-c$	$1-b$	–	1	$1-b$	–
with gen-parasite						
host fitness $\boldsymbol{W_h}$	$(1-u)$ $(1-s_2)$	$(1-u)$ $(1-s_3)$	$(1-u)$ $(1-s_2)$	$1-s_3$	$1-s_3$	$1-s_2$
gfg-parasite fitness $\boldsymbol{W_p}$	$(1-c)$ θ_{gen}	$(1-b)$ θ_{gen}	–	θ_{gen}	$(1-b)$ θ_{gen}	–

Table 4.4 – Fitnesses in Model 5.

For Models 4 and 5, as for Model 3, fitness is calculated for each individual host and parasite. It is assumed that fully fit hosts and parasites have relative fitnesses of 1. There are intrinsic fitness costs *u* to being a *RES*-host and *b* to being an *avr* gfg-parasite. There is a fitness cost *c* to the gfg-parasite of being unable to infect, i.e. of being *AVR* on a *RES*-host. Finally there are different fitness costs to the host of being diseased with one or both parasites and interaction parameters θ_{gen} and θ_p between the parasites, as discussed above. Fitness affects the number of seeds or spores individuals produce, defined in Table 4.2.

4.2.2 - Model 4

Model 4 differs from Model 3 in having variable host incidence - patches can have *RES*, *res* or no hosts. To allow this variable host incidence the model handles host reproductive rate, seed-number and seed dispersal in the same way as parasite reproductive rate, spore-

number and spore dispersal. Hosts and parasite thus have basic reproductive numbers y_h and y_p respectively. For each individual, the basic reproductive number is multiplied by fitness and rounded probabilistically to give seed- and spore-numbers σ_h and σ_p respectively. Each individual in turn, in random order, disperses all their seeds or spores a random direction and random-exponential distance with mean of β_h and β_p for hosts and parasites respectively. Each patch holds at most one seed and one spore, which exclude other seeds and spores respectively. In the absence of hosts, parasites have fitness and thus spore-number of 0.

As host density can change there are incidence variables for both species, M_h and M_p. M_h is the fraction of patches with hosts in the current generation. R and r are the fractions of hosts with of *RES*- and *res*-genotype respectively, making $M_h R$ and $M_h r$ the fractions of patches with hosts of these genotypes. Similarly M_p is the fraction of patches with parasites in the current generation and A and a are the fractions of parasites with *AVR*- and *avr*-genotypes respectively, making $M_p A$ and $M_p a$ the fractions of patches with parasites with these genotypes. As in Model 3, $R + r = 1$ and $A + a = 1$. It is possible for M_p to be higher than M_h, both for one generation and on average, if parasite basic reproductive rate q_p is high enough. However this does not occur in the simulations discussed here.

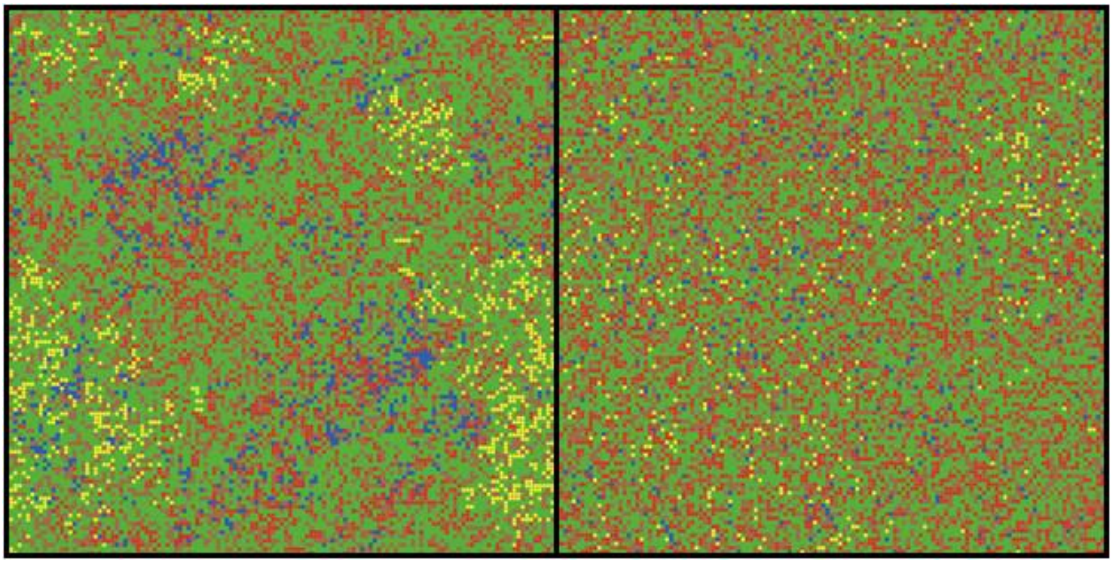

Figure 4.1 – Model 4 arenas. Blue, green and brown patches contain *RES*-hosts, *res*-hosts and no hosts respectively. Red and yellow circles are *AVR*-parasites and *avr*-parasites respectively. Parameters and costs are $y_h = y_p = 1.85$, $u = 0.01$, $s = 0.3$, $b = 0.05$ and $c = 1$). Dispersal parameters β_h and β_p are 1 and 4 in the first and second panels respectively. As in Model 3, limited dispersal leads to aggregation of genotypes (first panel) and higher dispersal breaks up these aggregations (second panel).

4.2.3- Model 5

Model 5 differs from Model 3 in having a generalist (gen-) parasite as well as a gene-for-gene (gfg-) parasite. The gen-parasite is modelled as a permanent environmental factor, neither reproducing nor dying – it is a property of patches that is either present or absent and remains so throughout model a given simulation. The gen-parasite is modelled this way because I am interested in how different spatial arrangements of the gen-parasite affect the coevolution of the host and gfg-parasite. Generalist parasites might disperse in a very different way to hosts and gfg-parasites for multiple reasons, e.g. if the gen-parasite is a soil-based fungus that spreads primarily by hyphal growth (such as the take-all fungus *Gaeumannomyces graminicola)* while the host and gfg-parasite are wind-dispersed or if the gen-parasite has a more limited environmental range than the host and gfg-parasite. In such cases the gfg-pathosystem will be divided into areas where the gen-parasite is present and absent. I predict that the dynamics of coevolution will be affected by the size of these areas and how strongly they are connected by seed and gfg-spore dispersal. While modelling the gen-parasite in this way means feedback between all three species cannot occur, it does allow precise control over the incidence and distribution of the gen-parasite.

The effects of different spatial aggregations of the permanent gen-parasite are studied. Gen-parasites can be scattered randomly, in 2x2, 4x4, 8x8 and 16x16 blocks, in two lines or in one patch. Random, 8x8, two-lines and one-patch arrangements of the generalist parasite are shown in Figure 4.2, below.

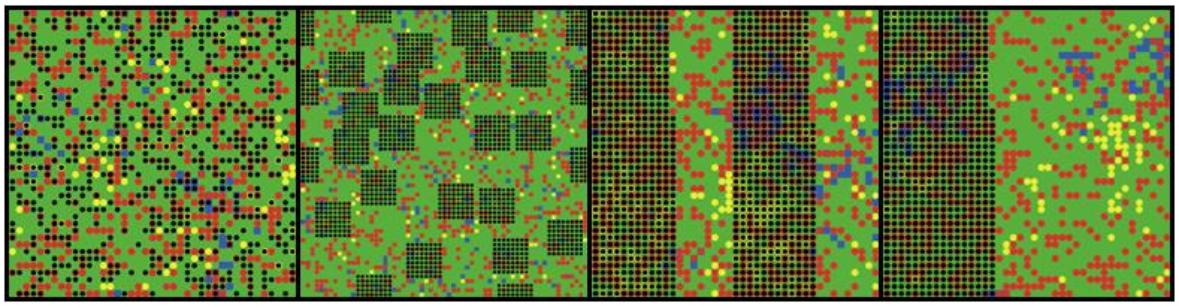

Figure 4.2 – Model 5 arenas showing the gen-parasites fixed at varying levels of aggregation. Blue and green patches contain *RES*-hosts and *res*-hosts respectively. Red, yellow and black circles are *AVR* and *avr* gfg-parasites and gen-parasites respectively. From left to right the panels show gen-parasites randomly aggregated, in 8-by-eight blocks, aggregated in two patches and aggregated in one patch. Costs are $u = 0.01$, $b = 0.05$ and $c = 1$. Other costs and parameters vary.

The presence of the gen-parasite can change the fitness of both the host and the gfg-parasite. The fitness cost to the host will depend on the negative effects of both parasites, so s is replaced with s_1, s_2 and s_3. s_1 is the fitness cost to the plant of infection by the gfg-parasite alone, s_2 is the cost of infection by the gen-parasite alone and s_3 is the cost of infection by both parasites. θ_{gen} is the effect the gen-parasite has on the fitness of the gfg-parasite when both infect the same host. It is modelled as a multiplier of the gfg-parasite's basic reproductive rate. Thus values of θ_{gen} above 1 increase gfg-fitness, representing synergy between parasites or at least an advantage to the gfg-parasite. Similarly values of θ_{gen} below 1 decrease gfg-fitness, representing antagonism between parasites or at least a cost to the gfg-parasite. Biologically θ_{gen} could represent direct interactions between the parasites but, perhaps more likely, could also represent parasite fitness being altered by cross-talk between host defence pathways. This is discussed in Section 4.1.3.

Like Model 3 and unlike Model 4, Model 5 has total host incidence so each patch contains one host every generation. Thus in Model 5, as in Model 3, each patch selects a potential donor patch in a random direction and at a random-exponential distance with mean β_h. The fitness of the host in the potential donor patch is then compared to a random number to see if it donates a seed. Thus hosts do not have explicit seed-numbers, rather relative seed-number depends on fitness. Full details are in Chapter 3 (Section 3.2.3).

In the case of *RES*-hosts with *AVR* gfg-parasites and gen-parasites, the hosts experience the cost of infection by the generalist parasite only (s_2) and the gfg-parasite experiences altered fitness due to the gen-parasite's presence (θ_{gen} is applied). The biological justification for this is that the *RES*-host stops the *AVR* gfg-parasite from infecting, so the host experiences the costs of infection by only the generalist parasite, while the host response has been triggered by the gen-parasite and affects the fitness of the gfg-parasite.

4.2.4 – Model 5 - modelling the effects of parasite-parasite interaction on gfg-parasite and host fitnesses

The total fitness cost of both diseases to the plant is related to the total reproductive success of the parasites. This suggests a relationship between the parasite interaction parameter θ_{gen} and s_3, or more precisely s_3 relative to s_1 and s_2. If there is no interaction between the diseases and no unexpected change in the total cost they inflict on the host, $\theta_{gen} = 1$ and $s_3 = 1 - (1 - s_1)(1 - s_2)$. Increased parasite reproductive success implies

higher costs to the plant and *vice-versa*. Thus values of θ_{gen} above or below 1 imply that s_3 should be higher or lower, respectively, than $s_3 = 1 - (1 - s_1)(1 - s_2)$.

4.3 - SIMULATIONS

4.3.1 - Model 4 Simulations

To allow comparison with Model 3 results, costs were set to [u=0.01 s=0.3 b=0.05 c=1]. These costs are biologically plausible and provide interesting coevolutionary outcomes, as discussed in Chapter 3 (Section 3.2.3). Arenas of 63-by-63 patches with 30 replicates were used to measure average time to fixation (TTF) as a measure of stability. As discussed in Chapter 3 (Section 3.3.2), arenas of this size are small enough for stochastic fixation to occur normally within a few hundred generations but large enough for their internal dynamics to affect the probability of stochastic fixation and lead to significant differences between average TTF. Average genotype frequencies and average frequencies of all nine different patch types (*RES*/*res*/no host with *avr/AVR*/no parasites) were measured from the same simulations.

Initial experiments with the above costs showed that reproductive parameters of $y_h = 1.85$ and $y_p = 1.85$ gave high host incidences, about 60%, and intermediate parasite incidences, about 20%, leading to polymorphism with all four genotypes. Thus these basic reproductive rates were used throughout. Dispersal parameters were set to all combinations of $\beta_h = [1\ 3\ 5]$ and $\beta_p = [1\ 3\ 5]$. Results were compared with each other and with the equivalently dispersing Model 3 results.

4.3.2 – Model 5 Simulations

To allow comparison with Model 3 results, costs were set at [u=0.01 s_1=0.3 s_2=0.3 b=0.5 c=1] and gfg-parasite reproduction was set at y_p=1.25. To keep the parameter space manageable and to maximise the spatial impact of gen-parasite aggregation, dispersal parameters were set to [$\beta_h = \beta_p = 1$]. Arenas of 63-by-63 patches with 30 replicates were used to measure average time to fixation (TTF) as a measure of stability.

The two main factors investigated were the aggregation of gen-parasites and the relationship between parasites. Aggregation levels of the generalist parasite were set at six levels. These were random scattering of individuals, random scattering in square blocks of 4, 16 and 64 individuals, aggregation into two stripes and aggregation into one patch. Figure 4.2 shows examples of individual scattering, squares of 64 scattering, two stripes and one patch. The relationship between gfg- and gen-parasites was synergistic, antagonistic or neutral and different combinations of parameters were used to represent varying degrees of synergy and antagonism between parasites (Table 4.5, overleaf). In antagonistic cases both the reproductive success of the gfg-parasite and the fitness cost experienced by the host were decreased, while in synergistic cases both were increased.

The incidence of the generalist parasite m_{gen} was set to a value calculated to be that of the internal equilibrium point in a version of the deterministic, non-spatial Model 1 (Chapter 2, Section 2.2.2) altered to include a variable-incidence second parasite and the relevant parameters. This model assumed identical costs, that gen-parasite basic reproductive rate was equal to gfg-parasite basic reproductive rate y_p and that the effect of gfg-parasite coninfection on gen-parasite fitness was equal to the effect of gen-parasite co-infection on gfg-parasite fitness θ_{gen}. Thus, m_{gen} was set at different values for each interaction. These are also shown in Table 4.5, below.

parasite interaction	s_3	θ_{gen}	m_{gen}
antagonistic	0.3	0.7	0.31387
antagonistic	0.3	0.8	0.321
antagonistic	0.3	0.9	0.3424
neutral	0.51	1	0.3803
synergistic	0.8	1.1	0.43098
synergistic	0.8	1.2	0.495
synergistic	0.8	1.3	0.56673

Table 4.5 – Parameters used in Model 5 experiments relating to between-parasite interaction. Antagonistic interactions between parasites imply reduced gfg-parasite fitness and reproductive success (θ_{gen}), reduced gen-parasite fitness and reproductive success (m_{gen}) and reduced fitness costs to the host (s_3), while synergistic parasite interactions imply the opposite.

4.4 - RESULTS

4.4.1 – Model 4 results - increasing β_h and the difference between β_h and β_p decreases average values for equilibrium *A*, equilibrium *R* and stability

Model 4 stability, measured as average TTF, declines from around 1200 generations at $\beta_h = \beta_p = 1$ to less than 500 generations at $\beta_h = 3$ and less than 400 generations at $\beta_h = 5$. This is shown in Figure 4.3, below (blue line). Stability declines with both increasing β_h and an increasing difference between β_h and β_p. Average genotype frequencies *A* and *R* also decrease with both increasing β_h and increasing difference between β_h and β_p. This is shown in Figure 4.4, overleaf (second panel). This relationship between dispersal, genotype frequencies and stability matches Model 3, discussed in Chapter 3 (Sections 3.4.1-3.4.4). As in Model 3, increasing dispersal alters average genotype frequencies and makes stochastic extinction more likely over any given period of time. Model 4 has slightly higher stability than Model 3 at $\beta_h = 1$ and $\beta_h = 3$.

Average host incidence M_h decreases with increasing β_p and is unaffected by increasing β_h, while average parasite incidence M_p increases with increasing β_p and decreases slightly with increasing β_h (Figure 4.4, first panel). Thus, parasite incidence increases when parasites can disperse faster than their hosts.

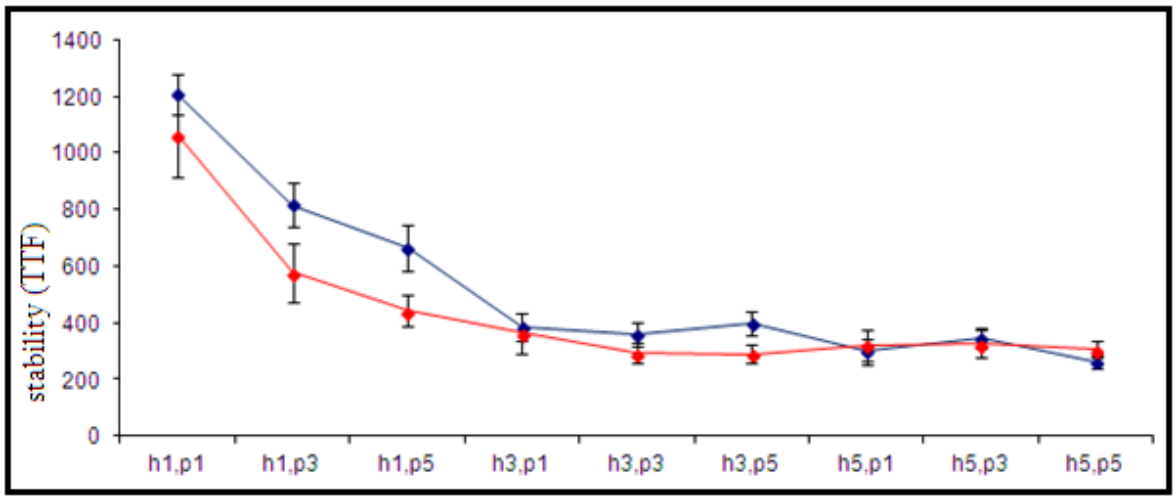

Figure 4.3 – Model 4 (blue) and Model 3 (red) stabilities against dispersal functions. y-axis is stability (TTF) and x-axis is dispersal (β_h is h, β_p is p). Error bars are standard errors. Stability in Model 4 follows the main trend of stability in Model 3, decreasing with increasing β_h and to a lesser extent with increasing β_p. However Model 4 stability is slightly higher than Model 3 stability at $\beta_h = 1$ and $\beta_h = 3$.

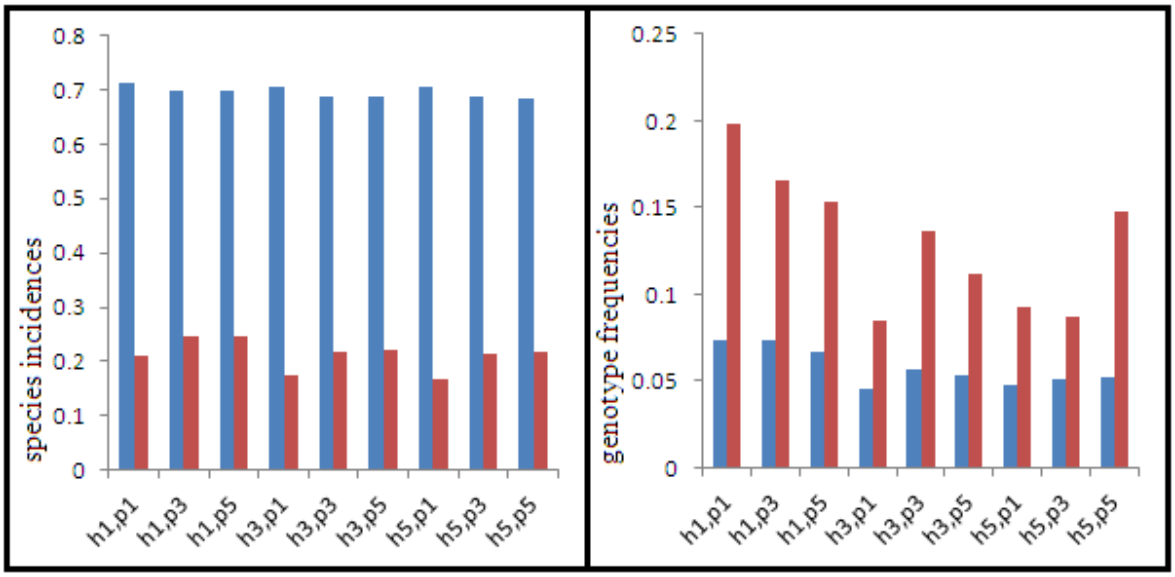

Figure 4.4 – Model 4, average species incidences and genotype frequencies against dispersal. The first panel shows average host incidence M_h (blue) and parasite incidence M_p (red), incidences on *y*-axis. The second panel shows average genotype frequencies *R* (blue) and *A* (red), frequencies on y-axis. (y-axis). x-axes are dispersal parameters β_h and β_p (labelled h and p as in Figure 4.3).

4.4.2 – Model 4 results explained – how variable host incidence affects stability depends on dispersal – at low dispersal it increases stability, while at high dispersal it has no effect

As discussed in the introduction, variable host incidence is predicted to both reduce deterministic genotype loss and increase stochastic genotype loss. The former is because at high parasite incidence host incidence is reduced, lowering parasite effective basic reproductive rate and strengthening DLR on parasites. The latter is because variable host incidence reduces population size of both species and makes random fixation more likely. The overall effect of variable host incidence depends on the balance between these two effects and is altered by host dispersal relative to arena size. In these results, Model 4 simulations have slightly higher stability than Model 3 simulations with equivalent parameters when $\beta_h = 1$ and $\beta_h = 3$ but stability does not differ between Models 3 and 4 when $\beta_h = 5$.

It follows that at low dispersal the extra DLR stabilising the system has a stronger impact than the reduced population size destabilising the system, while at higher dispersal the impacts are equivalent. This may be because at higher host dispersal oscillations in genotype frequency are shorter and have greater amplitude (Chapter 3, Section 3.3.4). This makes stochastic genotype loss over any period of time more likely, as do the reduced

population sizes in both species caused by variable host incidence. In conjunction, shorter and more pronounced oscillations and smaller population sizes may make stochastic genotype loss even more likely. Another possible explanation would be that higher host dispersal leads to faster replacement of dead hosts, minimising variation in host density and thus the occurrence of extra stabilising DLR, but this is not supported by the data on average host incidence (Figure 4.4, first panel).

Overall, allowing variable host incidence made little difference to stability. Relatively high basic reproductive rates ensured that most patches contained plants at any time-step, minimising the effects of variable host incidence. While this could have been altered, in this particular modelling system doing so would have made stochastic loss of parasite genotypes overwhelmingly likely and made it difficult to collect any data. In the real world, where disease regulation can reduce host densities by a larger amount than that shown here, the impact of variable host incidence on stability of polymorphism may be considerably greater. One possibility is that in the real world low host density leads to frequent stochastic extinction and recolonisation of parasite demes. This is discussed further in Section 4.5.2.

4.4.3 - Model 5 results – increasing spatial aggregation makes disease incidence and stability move towards an intermediate value regardless of synergy or antagonism

Simulations were designed to test the effects on the focal gfg pathosystem of both between-parasite synergy or antagonism, modelled as linked variation of the effect of the gen-parasite on the gfg-parasite's fitness θ_{gen} and the cost to the host of infection by both parasites s_3, and aggregation of the generalist parasite, modelled as that parasite occurring randomly or in blocks of various sizes (Section 4.3.2). Figure 4.5, overleaf, shows how stability and disease incidence are affected by these two factors. The x-axes show different levels of aggregation of the generalist parasite and the data series are different degrees of antagonism or synergy. The y-axes are average stability (TTF) and average incidence of the gfg-parasite in Panels A and B respectively.

In these simulations increasing synergy increases gfg-parasite incidence and reduces stability, while increasing antagonism reduces gfg-parasite incidence and a middling level of antagonism leads to maximum stability. Aggregation of the gen-parasite affects both incidence and stability. Higher aggregation of the gen-parasite increases gfg-parasite

incidence in low-incidence systems with antagonistic parasite interactions. Similarly higher aggregation reduces the stability of the more stable, antagonistic systems and increases the stability of the less stable, synergistic systems. Aggregation has little effect on systems with intermediate incidence and stability (θ_{gen} = 0.9 and θ_{gen} = 1.0). Thus, high aggregation reduces the impacts of strong synergy or antagonism on both incidence and stability.

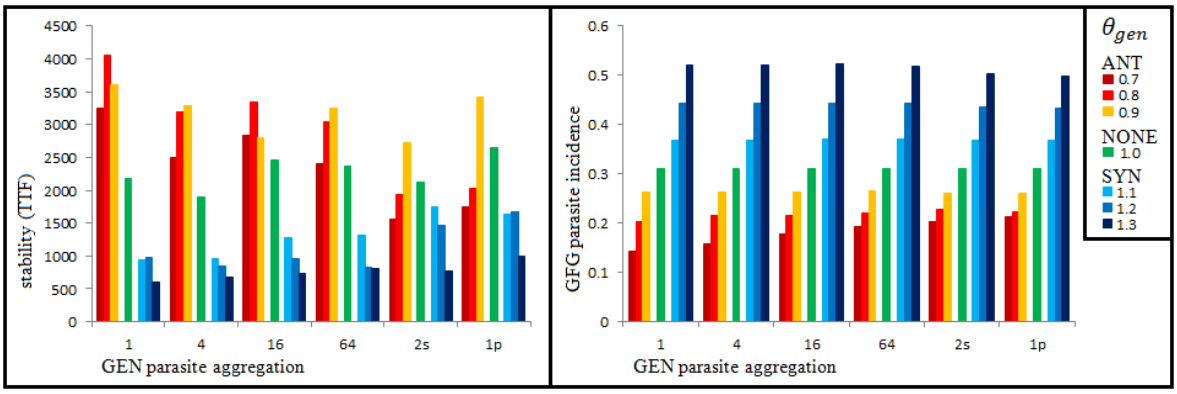

Figure 4.5 – Model 5, average stability (first panel) and average gfg-parasite incidence (second panel) against aggregation of gen-parasite and disease interaction. Aggregation is the x-axes and gen-parasites are aggregated as individuals, in blocks of 4, 16 or 64, in two stripes or in one patch (1, 4, 16, 64, 2s and 1p respectively on x-axes, described in Section 4.3.2). Parasite interactions are defined in the legend and are antagonistic (ANT), neutral (NONE) or synergistic (SYN). The numbers are values of θ_{gen} for each series. Increasing the degree of synergy (θ_{gen}) strongly increases gfg disease incidence. Increasing aggregation of the gen-parasite makes both stability and, much more weakly, gfg disease incidence for different degrees of synergy more similar, tending towards levels associated with intermediate stability of about 2000 generations and intermediate disease incidence of about 0.35.

4.4.4 – Model 5 results explained – synergy and antagonism affect gfg-parasite basic reproductive rate and thus stability

The degree of synergy or antagonism alters the gfg-parasite's effective basic reproductive rate. Synergy increases gfg-parasite fitness in patches where the gen-parasite is present and thus increases the gfg-parasite's global average reproductive success, while antagonism decreases these traits. Stability is affected by the basic reproductive rate and is highest when the rate is low to intermediate (Model 1, Chapter 2). Thus, given all simulations had gfg-parasite basic reproductive rate $y_p = 1.25$ which is above the optimum rate for stability, the effect of synergy and antagonism on stability is in retrospect predictable. The specific outcome depends on the value of y_p. If the gfg-parasite had a

sufficiently low basic rate of reproduction, synergy would be necessary to make its incidence high enough to allow stable polymorphism.

4.4.5 – Model 5 results explained – aggregation makes overall system more like system without gen-parasite

Overall, higher aggregation serves to make all systems more similar in both stability and gfg-disease incidence to the system with no interaction between parasites. Stability is correlated with effective basic reproductive rate and thus incidence (Section 4.4.4) so altered stability may be due to altered incidence. The question then becomes why is incidence affected by aggregation?

Higher degrees of aggregation make the system behave less like one system with average dynamics and more like several linked populations, one type with the generalist parasite and one without, each with their own internal dynamics and an overall dynamic depending on the strength of their coupling. This is similar to my Model 2 (Chapter 2) and Tellier & Brown's multi-patch models (2011). Incidence and therefore stability are thus driven by two population dynamics, one without the generalist parasite and one with, each driving the whole system towards its own equilibrium. The overall effect is a slight movement towards the no-aggregation levels of incidence. This becomes more obvious as the magnitude of antagonism or synergy between the parasites and thus the difference between dynamics in the two environments increases. This explains the trend in incidences and probably the trend in stability.

Linkage between patches is theoretically a source of oscillation damping leading towards stability (Bar-Eli 1985, Aronson 1990). However this would imply universally increasing stability with greater aggregation, rather than stability tending towards an intermediate value as occurred in these simulations. A possible explanation is that the rate of transfer between the patches is too high to cause oscillation damping that stabilises polymorphism. This is consistent with excessive transfer leading to instability in Tellier & Brown (2011) and Model 2 (Chapter 2). In this case transfer would synchronise patch dynamics, rather than damping oscillations to cause stability.

4.5 - DISCUSSION

4.5.1 – Model 4 - variable host density can increase or decrease stability – quantifying this in nature would require comparing pathosystems with different oscillation dynamics

Model 4 results show that variable host incidence, reinforcing DLR on parasites, can increase stability of polymorphism. However, in 63-by-63 arenas, this only happens when host dispersal is low. This is because high host dispersal increases the frequency and amplitude of oscillations in genotype frequency. This increases the chance of stochastic genotype loss, which is also increased by the smaller populations caused by variable host incidence. Model 4 results thus suggest that if a pathosystem has low populations, has highly dispersing hosts or is otherwise more prone to stochastic extinction, the overall effect of variable host density would be to decrease stability. Conversely, if a pathosystems is not particularly prone to stochastic genotype loss, the effect of variable host density in strengthening DLR and opposing deterministic genotype loss will increase stability.

Testing these contrasting hypotheses would involve comparing stability of polymorphism between multiple pathosystems, or isolated cases of the same pathosystem, that differ in size, dispersal characteristics or other factors affecting period and amplitude of genotype frequency oscillations. Surrogate measures for stability could include average age of known polymorphisms, measured from molecular clock data as in Bakker et al. (2006), or average number of polymorphisms in multi-locus GFG systems.

4.5.2 – Model 4 – host density in natural pathosystems often varies more than in Model 4 and its impact on stability may thus be much stronger

My results only cover cases where host density is slightly reduced. Large reductions in host density in small arenas would have made stochastic genotype loss overwhelmingly likely. The much stronger regulation of host density observed in wild pathosystems suggests that the impact on stability, whether increasing or decreasing it, may be correspondingly greater than in my Model 4 results.

Populations of most species vary over time (Futuyma 1996, Hedrick 2009), Variation in both host and parasite density is thus a common feature of natural pathosystems. Well-studied examples include wild flax *Linum marginale* with rust fungus *Melampsora lini* (Thrall & Burdon 2000) and *Plantago lanceolata* and *Podosphaera plantaginis* (Laine 2004). Both pathosystems exist as metapopulations characterised by repeated local extinction and reestablishment of parasites. Hosts have variable local density but local populations generally persist between years. Host density is negatively regulated by parasites in both systems. Thus, strong reduction of host density leading to local parasite extinction clearly does not render stable polymorphism impossible. It may in fact promote stable polymorphism when pathosystems exist as transient metapopulations. This is likely to be a common phenomenon in natural pathosystems.

4.5.3 – DLR on both parasites and hosts is well-known in nature, leads to linked density cycles and appears to stabilise the coexistence of victims and exploiters

Parasites grow most rapidly when host incidence is highest and fall when host incidence lowers. Conversely host incidence increases when parasites are scarce and falls when parasites are common. This could be termed indirect density-limited reproduction, or iDLR. Linked cycles of host and disease incidence, similar to Lotka-Volterra predator-prey cycles, result. Such host-parasite incidence cycles are well established in the literature (Holt & Pickering, 1985). Example pathosystems include caddis-fly and microsporidia (Kohler & Hoiland 2001) and numerous forest-dwelling insects and their parasites (Anderson & May 1981). A modelling study by Packer et al. (2003) suggested that deliberate removal of predators as a conservation measure, effectively increasing host basic reproductive rate and thus host density, can trigger unstable, expanding oscillations in host-parasite cycles. Indirect DLR on one or more species in a pathosystem thus stabilises the existence of the pathosystem as a whole, as well as the existence of polymorphism.

4.5.4 – Model 4 – variable or low density of specific host species may manifest as high species diversity at the community level

Many plant communities have high species diversity and thus a low density of most individual species. Low density often lowers disease transmission risk (Keesing et al 2006). Mitchell et al (2002) found that increased diversity in grass communities leads to a

two-thirds reduction in foliar fungal incidence and determined that this correlated with the reduced density of specific host species. A similar effect of cultivar diversity occurs in agriculture (Mundt 2002, Wolfe 1980, Chin & Wolfe 1984A, 1984B). Augspurger (1983) and Packer (2000) both reported that parasites cause heavier mortality on seedlings near parent trees, probably because parasites are transmitted with higher frequency and therefore inflict more damage on young plants when host density (roots of the parent tree) is high.

From the above, it seems likely that over millions of years low density of specific hosts and thus high species diversity has been driven by parasites. This in turn could strengthen DLR caused by variable disease incidence and may explain the known great age of some gene-for-gene polymorphisms (Tian et al. 2003, Bakker et al. 2006). However, the reduced and possibly more variable species population sizes caused by such community-level diversifying selection may also make stochastic loss of genotypes (or species) more likely.

In my own simulations, only one host and one parasite are modelled. This simplifies analysis and allows modelling with arenas of manageable sizes. However, larger NetLogo arenas lead to much longer periods of coevolution (Chapter 3) and would potentially allow for modelling such a multi-species framework. Strong regulation of host density, by both host-specific parasites and competing hosts, could not be modelled in small NetLogo arenas (Section 4.5.2) but is common in natural communities. Model 4 results already show a stabilising effect of DLR on the host as well as the parasite at low dispersal (Figure 4.5). I predict that strong DLR regulation on hosts would lead to stronger stabilising effects on stability of polymorphism, assuming populations are sufficiently large that stochastic extinction is unlikely.

4.5.5 – Model 4 – *avr* and *AVR* parasites may have different optimal dispersal strategies

Model 4 simulations show that total gfg-incidence always increases with parasite dispersal, while the frequency of *AVR*-parasites increases when parasite dispersal is similar to host dispersal. This is because *AVR*-parasites benefit from local adaptation due to avoiding *RES*-hosts and so do best when their dispersal matches their host, while *avr*-parasites can colonise any host and so do best when they have high dispersal and avoid competition with

sibling spores. The two genotypes of parasite thus have different optimal dispersal studies in Model 4, although in the real world high dispersal of *avr*-parasites carries the risk of missing populations of hosts altogether. Given the age of some *avr*/*AVR* polymorphisms (Tian et al. 2003) genotype-specific variation in dispersal is at least a possibility. Trade-offs between increased host range and efficiency of host exploitation are common and have been studied extensively (Leggett et al. 2013). However to my knowledge no authors have reported variation in dispersal correlated with host-range. This may be because genetic recombination makes it hard for dispersal traits to segregate with pathotype.

4.5.6 – Model 5 - multi-parasite dynamics are a specific case of environmental heterogeneity and the relevant signalling biology is well-understood, so further modelling has a sound theoretical basis

In the Model 5 results, the effects of the generalist parasite on the focal gfg-pathosystem are due to the generalist changing aspects of that focal interaction. These aspects include average costs, spatial uniformity of parasite incidence and spatial uniformity of costs. Respectively these are analogous to altered costs and parameters as in Model 1 (Chapter 2), linked populations varying in incidence as in Model 2 (Chapter 3) and linked populations of various costs as in Tellier & Brown (2011). As with any environmental factor, the challenge then lies in identifying exactly how the generalist parasite alters the basic interaction and where and when this effect occurs.

While environmental heterogeneity is not a new subject in host-parasite modelling, to my knowledge this is the first time parameters have been varied specifically to simulate the effects of host-mediated parasite synergy and antagonism. Multiple diseases are common in both natural and agricultural pathosystems, making this specific case of environmental heterogeneity of wide import. Variable environmental conditions have been studied in metapopulations (Tellier & Brown, 2011) and there is a broad understanding of how plant responses to multiple parasites interact (Glazebrook et al. 2005, Thaler et al. 2012, Pieterse et al. 2012). Thus, there exists a sound theoretical framework to support further investigation of this specific environmental variable.

4.5.7 – Model 5 - linkage between patches can lead to oscillation damping and increased stability or, if linkage is stronger, to averaged dynamics and intermediate stability

In Model 5, high aggregation that splits the population into different environments does not lead to increased stability through oscillation damping as in Model 2 (Chapter 2) and Tellier & Brown (2011). In those models such oscillation damping is the only source of ndFDS and breaks down with stronger linkage between demes, leading to instability. However Model 5 has DLR of gfg-parasites as a stabilising factor and, even without oscillation damping, linked environments display stability. The two linked environments develop one set of dynamics that is intermediate in both the average genotype frequencies and the average time to fixation (Figure 4.5).

The incidence of secondary parasites can vary at highly local spatial scales, as can other environmental factors, so this dynamic of averaging behaviour without oscillation damping could apply in the real world. While the breakdown of stability caused by oscillation damping when linkage between demes is sufficiently strong has been observed before, (Tellier & Brown 2011, Model 2), the individual-based nature and intrinsic stabilising factor of Model 5 makes it possible to study such breakdown in more detail.

Weakly linked, damping oscillations and strongly linked, averaging oscillations are simply two regions on a continuum that ranges from unlinked demes to two demes behaving as one. Making qualitative predictions about differences and especially about boundaries between the two dynamics is therefore difficult, although if the systems are approximated as deterministic equations linear algebra can be used to find boundaries. With this in mind, I make the following three hypotheses about systems best described as strongly linked with averaged oscillations compared to those that are weakly linked with dampened oscillations:

Average genotype frequencies and species incidences will be more similar in strongly linked than in weakly linked environments. This can be measured, although long-term and extensive sampling would be required.

In systems with two strongly connected environments, the stability of long-term polymorphsim will be intermediate between the stabilities of the two environments in isolation. This is because conditions will be averaged across the system as a whole.

Conversely, in systems with two weakly linked environments stability will be higher than in either environment in isolation. This is because weak linkage allows oscillation damping as in Tellier & Brown (2011) and Chapter 3. The stability of long-term polymorphism in a system can be quantified as average survival-time of polymorphism in populations of a given size or as average age of known polymorphism.

Strongly linked environments will be more likely than weakly linked environments to be associated with a continuous range of the focal pathosystem and a zone of gradually transitioning average frequencies and incidence between environments, as opposed to spatial separation and no zones with intermediate dynamics.

The differences between strongly-linked averaged and weakly linked damping systems could be studied in metapopulation models as well as individual-based models. The advantages of metapopulation models would be reduced stochasticity, reduced computational difficulty and potentially easier mathematical analysis. The main disadvantage of metapopulation models would be that the rate of exchange between demes would be set by the modeller, rather than emerging from dispersal parameters as in individual models. Individual-based and metapopulation spatial models are discussed in more detail in Chapter 5 (Section 5.2.3).

4.5.8 – Model 5 – three-species dynamics from allowing generalist parasites to reproduce, disperse and die rather than being fixed might lead to more complex feedback, potentially affecting the stability of long-term polymorphism

It is possible to construct a version of Model 5 where the generalist parasite, rather than being a fixed environmental condition, can reproduce and die like the other species in the simulation. Such a model would include parameters for gen-parasite basic reproductive rate, dispersal and the effect of co-infection with the gfg-parasite on the generalist parasite. I did not use such a model in the experiments reported in this chapter because I chose to focus on spatial aggregation of the generalist and wanted the ability to control this precisely. For investigations where the focal host-parasite pairing is of primary interest, this approach may often be all that is required. Conversely the three-way feedback the full model allows may lead to interesting dynamics, although further speculation would be unsubstantiated at this point.

CHAPTER 5 – GENERAL DISCUSSION

INTRODUCTION

In this chapter, I evaluate how this thesis expands theoretical knowledge of host-parasite coevolution. I also comment on other modelling studies and on hypotheses from Models 1-5 and set out ideas for future research in modelling and real-world studies. Key topics include:

The expansions made to Tellier & Brown's one- and two-patch models (2007A, 2011) and the concepts of non-ndFDS mechanisms including limited dispersal opposing stochastic genotype loss and of possible multiple causes of stability in the real world (Section 5.1).

The advantages and disadvantages of individual-based models compared to population and metapopulation models and the possibility of modelling more realistic multi-gene GFG interactions inlcuding multiple alleles, multiple loci or both (Section 5.2)

Biological evidence for and against hypotheses made in earlier chapters from both wild pathosystems and controlled laboratory experiments on coevolution (section 5.3).

Possible relevance to other fields, including subsistence agriculture and human epidemiology (Section 5.4).

5.1 - ADVANCES IN KNOWLEDGE FROM THIS THESIS

This section contains a discussion of my key results and how they advance coevolutionary theory. Key results are:

Variable disease incidence in time modelled as logistic parasite growth, regulated by density-dependent growth of total parasite incidence, leads to regulation of incidence in both parasite genotypes. This is density-dependent selection or density-limited reproduction (DLR). DLR, like negative direct frequency-dependent selection (ndFDS) in

Tellier & Brown (2007A, 2009, recapped in Section 5.1.1), can potentially stabilising long-term polymorphism (Section 5.1.2).

Variable disease incidence in space can cause oscillation damping between demes as in Tellier & Brown (2011), potentially stabilising long-term polymorphism (Section 5.1.3).

Limited dispersal of hosts, parasite or both alters average genotype frequencies and oscillation dynamics, making stochastic loss of genotypes in finite populations less likely over any given period of time (Sections 5.1.4-5.1.6).

Multiple sources of ndFDS can oppose deterministic genotype loss, while multiple factors can oppose stochastic genotype loss. While complexity is not *per se* required for stable polymorphism, this interaction of multiple factors may make loss of genotypes less likely and may occur in many real pathosystems. This is not an experimental result, but is a useful insight for considering long-term polymorphism in the real world (Section 5.1.6).

Avirulent parasites were the species and genotype most likely to exhibit local adaptation when dispersal was limited. This seems in contrast to earlier results (Gandon et al. 1996, Gandon 2002), where the species with higher dispersal was locally adapted. However, different values of costs of maladaptation in mine and Gandon's models explain this apparent discrepancy (Section 5.1.7)

5.1.1 – Recap of framework developed by Tellier & Brown (2007A, 2011)

Models 1 and 2 (Chapter 2) fit into and expand the framework developed by Tellier & Brown (2007A) and (2011) respectively. Tellier & Brown (2007A) showed that negative frequency dependent selection, ndFDS, is required for stable polymorphism. This ndFDS can be generated by multiple ecological or epidemiological factors including auto- versus allo-infection (Tellier & Brown 2007A) and seed-banks (Tellier & Brown 2009). Tellier & Brown (2011) showed that variation in costs between linked demes can stabilise polymorphism in metapopulations. This occurs because demes with different conditions have different periods of oscillation and linking such demes can cause oscillation damping, a source of ndFDS (Bar-Eli 1985, Aronson et al. 1990).

5.1.2 - Variable disease incidence in time is a non-spatial stabilising factor analogous to ndFDS

Model 1 establishes that temporally variable parasite incidence, regulated as logistic growth (Murray 1989), causes negative density-dependent regulation on parasites. I refer to this as DLR, density-limited reproduction. Like ndFDS, DLR is capable of stabilising polymorphism. DLR is not actually ndFDS, as it regulates the total density of a species rather than the frequency of genotypes within a species.

DLR and ndFDS from auto-infection (Tellier & Brown 2007A) and seed-banks (Tellier & Brown 2009) are similar in that they do not require environmental heterogeneity or spatially structured populations. I will therefore refer to these factors as non-spatial stabilising factors.

A plethora of non-spatial stabilising factors are now known to exist. Their numbers imply that many or most real-world pathosystems have multiple sources of stability. Many deterministically stable systems spend a great deal of time with gene-frequencies very near zero (Jayakar 1977, Leonard 1993, Tellier & Brown 2007A, Models 1 & 2) and for real-world pathosystems with finite numbers and stochastic fluctuations this threatens stability. The existence of many factors, each individually capable of stabilising polymorphism, suggests multiple factors could interact to increase stability. These factors include both non-spatial and spatial (Section 5.1.3) sources of ndFDS, as well as similar sources of deterministic stability such as DLR. They also include factors opposing stochastic genotype loss (Section 5.1.4). In combination these factors could dampen oscillations faster, thus keeping genotype frequencies further from fixation and making genotype loss in real, finite pathosystems less likely.

5.1.3 – Variable disease incidence in space is an example of spatial heterogeneity leading to oscillation damping, a source of ndFDS.

Model 2 (Chapter 3) shows that spatial variation of disease incidence can lead to oscillation damping and stable polymorphism. Varying disease incidence between demes is analogous to varying the cost of infection between demes in Tellier & Brown (2011). Model 2 is thus both interesting in its own right and as an indicator that many more factors could tie into the framework of oscillation damping as a spatial source of ndFDS. For

example, one could assume the intensity of inter-host competition varies between environments. In terms of Tellier & Brown's model (2011) this would be analogous to varying costs of resistance, infection or both.

Numerous other factors could vary between environments and in many real-world pathosystems one or more such factors may vary. This supports the idea that real-world pathosystems can contain multiple stabilising factors.

5.1.4 – Limited dispersal is an example of self-slowing in a spatial system and, in conjunction with a source of ndFDS, can make stochastic genotype loss less likely. Self-slowing factors opposing stochastic genotype loss may be common in nature

The importance of spatial structure in stabilising polymorphism has long been debated (Chapter 1, Section 1.3.1). Brown & Tellier (2011) divided spatial factors into truly stabilising factors and factors that prolong transient polymorphism. The former are spatial sources of ndFDS such as oscillation damping between linked demes, discussed in Section 5.1.3. The latter are factors that slow down coevolution, such as metapopulation structure, and are referred to as self-slowing. They are not sources of ndFDS and cannot stabilise an unstable system, but can greatly increase the time it takes such a system to go to fixation.

In a finite population model subject to stochastic genotype loss, and thus presumably in natural populations, self-slowing factors make genotype loss less likely over any given time period. Model 3 has both stabilising and self-slowing factors, respectively DLR from temporally variable disease incidence and limited dispersal. Reduced dispersal increases the average time to fixation. This is due to altered genotype frequencies and possibly altered oscillation dynamics (Chapter 3, Section 3.3.2), both of which are self-slowing mechanisms.

In real pathosystems, which are finite and subject to stochastic as well as deterministic genotype loss, these self-slowing mechanisms act in concert with spatial and non-spatial stabilising factors to make prolonged polymorphism more likely. This assessment of self-slowing mechanisms as mechanisms opposing stochastic genotype and acting in concert with, but not replacing, sources of ndFDS is the most important general result of Model 3.

5.1.5 – A combination of ndFDS and self-slowing factors may explain stability in earlier metapopulation models

Sasaki et al. (2002) and Thrall & Burdon (2002) both described models with mutation and metapopulation structure. Both cited spatial structure as the stabilising factor. I propose these models are examples of ndFDS and self-slowing mechanisms acting in concert. The models have mutation as a source of ndFDS opposing deterministic genotype loss. However, as low-frequency mutation is a weak source of ndFDS, stability is only observed when low dispersal reduces stochastic genotype loss. In nature, of course, mutation is unlikely to be a significant source ndFDS (Chapter 1, Section 1.2.3).

5.1.6 – Long-term polymorphism often occurs in spatially structured pathosystems, suggesting self-slowing factors such as spatial structure are important in stability in nature

Both Thrall & Burdon (2000) and Laine (2005) described natural pathosystems where stable polymorphism is observed and where the systems are large relative to host and parasite dispersal. Limited dispersal as an oscillation-damping mechanism and thus as a source of protection from stochastic genotype loss could be important in real pathosystems.

I suggest multiple factors, in particular the combination of spatial and non-spatial stabilising factors opposing deterministic genotype loss and self-slowing or equilibrium-altering factors opposing stochastic genotype loss, is not a mathematical necessity for long-term stable polymorphism but is probably common in the real world. Further possible self-slowing mechanisms are discussed in Chapter 3 (Section 3.5.2).

5.1.7 – Model 3 shows that limited dispersal allows local adaptation and suggests that the species locally adapted is the species with the higher cost of being maladapted

A more specific result from Model 3 is that limited dispersal allowed local adaptation. In Model 3 parasites were locally adapted if anything was. More precisely, *AVR*-parasites were locally adapted in that they tended to aggregate in areas where only *res*-hosts were present. When host dispersal exceeded parasite dispersal, this *AVR*-parasite local adaptation was reduced but *AVR*-parasites were still the only genotype showing local adaptation. By contrast Gandon (2002) and Gandon & Michalakis (2002) reported that in

a matching-allele spatial model pathosystem the species with higher dispersal tended to be locally adapted.

Model 3 differs from Gandon's model in that the cost of local maladaptation is always far greater for the parasite than for the host. The cost to *AVR*-parasites of being unable to infect was modelled as 1, higher than the cost to hosts of being resistant, infected or both. In contrast, in Gandon's model both parameters ranged from 0 to 1. In fact Gandon's simulations also indicated that when the cost of maladaption is total for the parasite but not for the host the parasite is typically locally adapted.

Greischar & Koskella (2007) conducted a meta-analysis of 54 host-parasite systems and found local adaptation of parasites in 18 and of hosts in 6. They found that parasite local adaptation was correlated with parasite dispersal exceeding host dispersal, but that host local adaptation did not increase when host dispersal exceeded parasite dispersal. 25 of their studies were based on plants and fungal parasites, including 13 where parasite dispersal was lower than host dispersal. These 25 cases included nine locally adapted parasites and three locally adapted hosts, including two of each when parasite dispersal was lower than host-dispersal. Thus, fungal parasites were much more likely than host plants to be locally adapted when they dispersed more and still equally likely to be locally adapted when they dispersed less.

Greischar and Koskella's results are intermediate between mine and Gandon's, in that they support parasite local adaptation when parasites disperse more but not host local adaptation when hosts disperse more. This relative lack of host local adaptation may be because selective pressure on fungal parasites is generally greater than on hosts. From my work, Gandon & Michalakis (2002) and Greischar & Koskella (2007) I would suggest that local adaptation is influenced primarily by the relative strength of selection pressure and secondarily by relative dispersal.

5.1.8 – Models 4 and 5 add further detail to the spatial coevolutionary dynamics simulated in Model 3

Model 4 results suggest variable host density can increase stability, measured as average time to fixation. This is probably because variable host density, regulated by disease incidence, increases stabilising DLR on parasites and thus opposes deterministic genotype

loss (Chapter 4). Conversely, I predict that reduced host density, resulting in smaller populations of both host and parasite, could make stochastic genotype loss more likely. This suggests the net effect of variable host density on real-world coevolutionary outcomes would depend on the relative strength of these two factors. In Model 4 such reduced stability with variable host density was not seen, a result I attribute to the relatively minor reductions in host density I was able to model.

Model 5 results showed that, in a spatial system with a non-spatial source of stability, strong linkage between demes with different environmental conditions leads to averaged dynamics (Chapter 4). This is in contrast to the situation where weak linkage between such demes leads to oscillation-damping and thus stability (Tellier & Brown 2011, Model 2 in Chapter 2). Environments can clearly vary at scales small enough for demes not to be isolated, suggesting this dynamic could be common in real-world pathosystems.

5.2 – MODELLING GFG-COEVOLUTION – SPATIAL, NON-SPATIAL AND MULTI-LOCUS MODELLING - IDEAS FOR FUTURE DIRECTIONS

This section contains:

A discussion of the limitations of simple, non-spatial deterministic models (Section 5.2.1).

A discussion of the advantages and limitations of metapopulation and individual spatial models (Section 5.2.2).

A proposed framework for future individual spatial coevolutionary models (Section 5.2.3).

A discussion of how to model more realistic GFG genetics (Section 5.2.4).

5.2.1 - Limitations of simple models include no stochasticity and no spatial dimension

The simplest models of GFG coevolution are non-spatial deterministic models such as Jayakar (1970), Leonard (1993), Tellier & Brown (2007A) and Model 1. These models are easy to describe, comprehend, create and alter and are mathematically tractable, *e.g.* by linear algebra (Roughgarden 1979). As the models are simple, one can pinpoint the biological and mathematical causes of stability. However, for all their advantages, such

models do have limitations. Specifically they assume infinite and spatially unstructured populations, two key differences from reality.

An infinite population has no random fluctuations in gene frequencies. A considerable subset of stable parameter space in deterministic models involves systems spending a lot of time with frequencies *R* and *A* close to zero, meaning this lack of fluctuation conceals what could be a very high likelihood of genotype loss in finite populations. This is stochastic genotype loss as opposed to deterministic genotype loss, as discussed in Section 5.1.4.

Determining the likelihood of stochastic genotype loss is difficult. The probability of stochastic genotype loss at a single point in time relates to the allele frequency and the population size (Hedrick 2009). However to use this theory to calculate the probability of allele fixation over a period of time, one must know or have reasonable estimates for the age of the coevolutionary relationship, the historical fluctuations in population numbers and the historical periods and amplitudes of oscillations in genotype frequencies. Ultimately, what this amounts to is that a hard-to-define subset of deterministically stable cases would not be stable in anything approaching real conditions. This represents a slight and difficult to quantify shrinking of stable phase-space as seen in figure 5.1, below. However, except in the specific case of small populations, this finite/infinite issue is ultimately a less serious limitation of simple models than their very limited realism with regard to factors other than population size. These factors include spatial structure, ecological complexity and genetic realism.

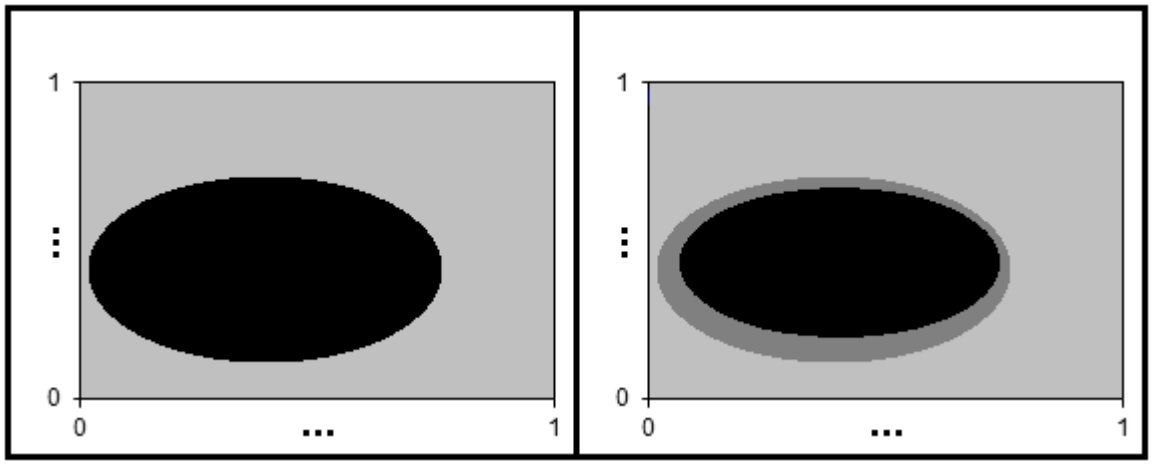

Figure 5.1 – General phase space of stable polymorphism in non-spatial models. Axis could represent two costs (as in Tellier & Brown's models) or parasite basic reproductive rate against cost *s* (as in my model). The first panel shows the clear-cut boundary between stable (black) and unstable (grey) polymorphism given by infinite populations, while the second panel shows the stochastic transition zone (dark grey) given by finite populations.

5.2.2 - Spatial modelling - metapopulations and massive individual-based models

A common assumption in metapopulation models is that, even if between-deme interactions are stochastic, within-deme interactions are deterministic (*e.g.* Thrall & Burdon 2002, Sasaki et al. 2002). This leads to the problem of deterministic as opposed to stochastic stability discussed above. Additionally, dividing organisms into demes with no spatial structure within and group-based spatial structure between is a limited way of modelling populations. While such a hierarchical organisation may be appropriate for some fragmented populations or cases where suitable habitat occurs in discrete patches such as islands or fields (Laine 2005, Laine & Hanski 2006), it is not generally descriptive of species that can disperse anywhere but not any distance within a habitat (Thrall 2000). Additionally, within-deme and between-deme dynamics emerge naturally from the same individual-level processes in individual-based models but are entirely distinct in metapopulation models. This is why I choose to use individual-based models, but these have their own problems. One issue is modelling arenas large enough and sparse enough to represent realistic pathosystems, although computing power is now seldom an issue in ecological modelling. Another issue is that deterministic metapopulation models are more amenable to analysis than the necessarily stochastic individual-based models.

5.2.3 – A proposed framework for large-scale, sparse, individual-based spatial models of host-parasite coevolution

My models and several other studies suggest limited dispersal relative to arena-size tends to stabilise polymorphism (Thrall & Burdon 2002, Sasaki 2002, Gandon 2002, Nuismer 2006, Gandon & Nuismer 2009). More realistic dispersal functions than I use in Models 3-5 would thus appear to be a logical area for modelling research. Such realistic dispersal would require much larger arenas than I have used in this thesis. Such arenas and the massive populations they could support would also allow diseases to persist with lower contact rates due to more variable and potentially much lower host incidence, more closely resembling real-world pathosystems (Laine & Hanski 2006, Thrall & Burdon 2003).

The advantages and disadvantages of individual-based and metapopulation models are discussed above (Section 5.2.2). For studying emergent processes based on dispersal I

would favour individual-based models, as demes and linkages between them would result from individual dynamics rather than being imposed by the modeller. Wingen et al. (2007, 2013) are examples of very large individually-based models generating new insights about spatial dynamics. Wingen et al. (2007) discussed the impact of different dispersal function on spatial structures of population. A key result was that power-law as opposed to exponential, or fat-tailed as opposed to thin-tailed, distributions of dispersal distances lead to more widely distributed populations even when mean dispersal was the same.

Another issue relating to large arenas is variation in costs and parameters. These change abruptly between demes in Tellier & Brown's multi-deme models (2011) and in my Model 2. While many environmental variations are quite abrupt (soil-type, shift between wild, farmed and industrial lands), others are gradual (temperature, day-length, humidity). A very large individual-based arena would allow either kind of transition. Biologically this would enable modelling of a wider range of environmental factors. In particular, gradual as opposed to abrupt parameter shifts could affect the range of linkage strengths leading to oscillation damping as opposed to system averaging (Chapter 4, Section 4.5.7). I predict gradual transitions would increase isolation between environments, favouring oscillation-damping dynamics over averaging dynamics.

5.2.4 – Increasing genetic realism by modelling multiple genes, alleles and loci in GFG coevolution

All the models in this thesis, together with many previous modelling studies, have assumed one host locus with *RES* and *res* alleles and one parasite locus with *avr* and *AVR* alleles. Multiple sets of *RES*/*res* and *avr*/*AVR* genes have been modelled in numerous studies (Sasaki 2000, Salthe et al. 2005, Segarra 2000, Tellier & Brown 2007B), but these studies have implicitly assumed separate loci for each *RES*/*res* and *avr*/*AVR* gene. This is not always true, as discussed below.

In plants, multiple *RES*-genes quite often share a single locus. Examples include the *Mla* locus in barley (Seeholzer et al. 2010, Wei et al. 2002), the *Pm3* locus in wheat (Huang & Roder 2001, Srichumpa et al. 2005, Yahiaoui et al. 2006) and loci in *Arabidopsis thaliana* (Holub 2001). Conversely while parasite *AVR*-genes can be associated with genetic elements that favour gene duplication such as LINE-1 retrotransposons (Sacristan et al.

2009), which allow new specificities to evolve, locus-sharing by multiple *AVR*-genes has not been reported.

In biological terms, locus-sharing as a mechanism of avoiding costs for *RES*-alleles but not for *avr*-alleles makes sense. *RES*-genes might have constitutive costs (Bergelson & Purrington 1996, Tian et al. 2003, Laine & Barres 2013). Locus-sharing thus means paying one constitutive cost rather than several such costs. In contrast the cost of *avr*-genes is not a cost of having that *avr*-allele, but of lacking a specific *AVR*-gene with beneficial effector activity (Dodds & Rathjen, 2010). Locus-sharing would thus incur several costs of not having a number of effectors, rather than minimising costs of *avr*-genes. Thus, locus-sharing to minimise total costs is more likely for *RES*-genes than for *AVR*-genes.

A logical place to start investigating multi-gene interactions would be expanding a simple model of one *RES*/*res* gene in hosts and one *avr*/*AVR* gene in parasites to include two different gene-pairs where *RES1* detects *AVR1* and *RES2* detects *AVR2*. The question then arises do these alleles share a locus or have different loci. Ignoring sex and diploidy, there are four possible combinations of genetic architecture in a two-species pathosystem. The first two combinations both exist in nature, while the last two include locus-sharing of *AVR*-genes and are not known or predicted to exist (above) but are included for completeness. In all cases, each locus will have a single *res*- or *avr*-allele. The four combinations are:

Two loci in each species: *RES1* and *RES2* in the host interacting with *AVR1* and *AVR2* respectively in the parasite.

One locus in host, two in parasite: *RES1-1* and *RES1-2* in the host interacting with *AVR1* and *AVR2* respectively in the parasite. Example include the *Mla* locus in barley (Seeholzer et al. 2010, Wei et al. 2002), the *Pm3* locus in wheat (Huang & Roder 2001, Srichumpa et al. 2005, Yahiaoui et al. 2006) and loci in *Arabidopsis thaliana* (Holub 2001).

Two loci in host, one in parasite: *RES1* and *RES2* genes in the host interacting with *AVR1-1* and *AVR1-2* respectively in the parasite. I am not aware of any biological evidence for this model.

One locus in both species: *RES1-1* and *RES1-2* in the host interacting with *AVR1-1* and *AVR1-2* respectively in the parasite. I am not aware of any biological evidence for this variation of the GFG-model. However, except for the *res*- and *avr*-alleles, this system does resemble the allele-for-allele model suggested by some authors for animal immunes systems (e.g. Koskella & Lively 2007).

By themselves, these four different architectures do not generate stabilising ndFDS (work not shown, same result in Tellier & Brown 2007B). However the question remains whether these architectures would expand the range of stable polymorphism in a system with a source of ndFDS. Tellier & Brown (2007B) reported that synergy of costs of virulence and complementation of costs of resistance would both increase range of stable conditions in systems with sources of ndFDS. Such cost-structures seem biologically plausible and would promote the sharing of loci in hosts but not parasites, as discussed above.

To reiterate a common theme in this thesis, the effects of complex genetic architecture on coevolution may alter dynamics in a way that makes stochastic genotype loss more or less likely and in finite real-world pathosystems this could be important. Possible support for this theory is provided by the fact that a locus-sharing genetic architecture for resistance genes is common in hosts. This could also reflect intrinsic costs of having many loci, however the prevalence of gene-duplication events in the course of evolution (Seeholzer et al. 2010) suggest such costs are not large. From a modelling perspective this issue is amenable to simple, non-spatial analysis.

5.3 – MODEL HYPOTHESES AND EXISTING BIOLOGICAL DATA

This section contains:

A recap of testable hypotheses from Models 1-5 (Section 5.3.1).

A discussion of how wild and agricultural pathosystems, mostly plants and parasites with GFG relationships, support or refute these hypotheses (Section 5.3.2).

A discussion of how coevolutionary microcosms, without GFG genetics but in some cases with comparable relationships, support or refute these hypotheses (Section 5.3.3).

A discussion of potential future coevolutionary microcosm experiments relevant to GFG coevolutionary theory (Section 5.34).

5.3.1 – Recap of hypotheses from this thesis

Model 1: Variable parasite incidence within a population generates stabilising DLR. There is a link between parasite effective basic rate of reproduction R_0, functionally measurable as parasite incidence, and the stability and nature of polymorphism. Stable polymorphism for all four genotypes (Type 4 stability) occurs at low to intermediate parasite R_0. Stable polymorphism for *RES*/*res* genes with fixed avirulence (Type 3 stability) occurs at even lower R_0. Below this range of R_0 fixation for avirulent parasites and susceptible hosts occurs, while above this range oscillations spiral outwards towards fixation.

Model 2: Linking two (or more) patches with different levels of disease incidence and therefore different periods of oscillations in gene-frequencies can lead to oscillation-damping and stabilise polymorphism in both patches.

Model 3: Limited dispersal of hosts and parasites leads, via spatial clustering of genotypes and local adaptation of avirulent parasites, to increased equilibrium genotype frequencies *A* and *R* and genotype frequency oscillations of longer period and lower magnitude. Both factors oppose stochastic genotype loss but do not necessary affect deterministic allele loss.

Model 4: Variable host incidence regulated by the parasite both strengthens density-dependent regulation on the parasite by expanding the range of R_0 for which deterministically stable polymorphism occurs and lowers effective population size, making stochastic genotype loss more likely. The overall effect on coevolutionary outcomes depends on the relative strength of the two factors.

Model 5: Spatial systems including environmental heterogeneity in which environments are too strongly linked for oscillation damping to lead to stability will exhibit averaging of

properties between environments. Averaged properties include longevity of polymorphism. Such systems can include stabilising factors other than oscillation damping, such as density-limited reproduction.

5.3.2 - Wild and agricultural pathosystems and hypothesis testing

From Models 1 and 4 I predict stable polymorphism will be most common when parasite incidence is low to intermediate and that low to intermediate host density, as is often the case in wild but not agricultural pathosystems, will make this more likely. Agricultural and wild pathosystems also differ because hosts are subject to selection by farmers, rather than fitness-based selection.

Studies of wild pathosystems support the Model 4 result that disease incidence decreases with reduced host density. Disease incidence also decreases with increased host diversity, which amounts to reduced density of any specific host species (Keesing et al. 2006, Mitchel et al. 2002). I suggest observed high diversity of host species in wild communities is due in large part to historical parasite regulation of host densities. If parasite regulation of host densities is often historical, as well as augmented by hosts competing with other plant species, it may be difficult to observe in current populations. Still, such regulation may have led to low host density and thus to lower parasite incidence.

Hypothetical regulation of population sizes in this way in the past could help explain why wild pathosystems feature low to intermediate densities of both hosts within the community (Augspurger 1983, Packer & Clay 2000) and parasites on hosts (Laine & Hanski 2006, Thrall & Burdon 2003). Wild pathosystems also feature ancient polymorphisms in *RES-AVR* gene pairs (Tian et al. 2003) and this link between disease incidence and age of polymorphism matches the Model 1 prediction of type 4 stability. A specific prediction from Model 1 is that higher disease density, possibly caused by higher host density (from Model 4), would lead to a break-down of stable polymorphism. Model 1 also suggests even lower disease incidence will cause Type 3 stability.

Conserved genetic features of *RES*-genes and modern sequencing technologies make potential *RES*-genes easy to find. However moving from potential to functional *RES*-genes, identifying corresponding parasite *AVR*-genes and quantifying genotype frequencies *R* and *A* in wild populations remains a difficult and unappealing prospect for the kinds of

low-incidence, and thus likely wild and commercially unimportant, pathosystems predicted to display Type 3 stability. Fortunately, genetic sequencing is not necessarily required. Using simple inoculation experiments one could demonstrate that a plant population has qualitative, binary variation in resistance. This would imply the existence of resistance genes with non-functional, presumably susceptible alleles. One could then investigate whether this resistance is effective against all collected parasite strains, implying no virulent strains and a Type 3 system, or whether some parasite strains are unaffected by resistance, implying virulent strains and a Type 4 system. Such a study, while time-consuming, would be easier and cheaper than combined genetic and population analysis. Quantifying the presence/absence of resistance by inoculation experiments rather than by genetic analysis is relatively common. Published examples include Bevan et al. (1993) and Thrall & Burdon (1999).

However, while a low-incidence pathosystem with *RES* but no *avr* genes could be a case of Model 1 Type 3 stability, it could also be that *avr*-alleles have simply not had time to evolve. If any pathosystem shows no virulence where disease incidence is low (for example because hosts are scarce or environmental conditions do not favour parasites) but virulence where disease incidence is higher, this would be a direct validation for the Type 3 as opposed to Type 4 stability prediction.

Model 1 hypotheses about distinctions between high incidence with fixation, intermediate incidence with Type 4 stability and low incidences with Type 3 stability require multiple, isolated examples of the same pathosystem to test. Isolation requires the distance between populations to be large relative to parasite dispersal. Wind-dispersed plants on multiple isolated islands could be a system to investigate this. Laine (2005) studied a plant-fungus pathosystem on a group of islands that were both isolated from other populations and sufficiently isolated from each other for local contact-rates and density to differentiate local coevolutionary outcomes. More generally, rare, long-distance dispersal events can establish pathosystems that then develop in isolation (Brown & Hovmoller 2002, Hovmoller et al. 2008) and island populations would be likely to differ in disease incidence through biotic or abiotic factors.

A hypothesis from Model 3 is that limited dispersal will both increase local adaptation of *AVR*-parasites, ultimately increasing average *A* and *R*, and decrease frequency and amplitude of oscillations in gene frequency. The former factor is well-established –

limited dispersal is a logical necessity for local adaptation and such adaptation is regularly observed. Both factors are predicted to increase the stability of polymorphism.

Significant host local adaptation (parasite local maladaptation) is rarer than parasite local adaptation, because parasites pay higher costs for being maladapted, but has been reported. For example, populations of the generalist parasite *Pseudomonas syringae* can utilise multiple hosts and has been reported to have local maladaptation to specific host-species (Kniskern et al 2011), which would be predicted when the cost of maladaptation on a specific host is diluted by a wide range of hosts. Such local maladaptation of the parasite would be sustained because the availability of many hosts means the evolutionary pressure on a specific host species to be locally adapted to the parasite would be greater than the average evolutionary pressure on the parasite to be locally adapted to that specific host. This is a reverse of the typical situation, where species-specific parasites are under stronger evolutionary pressure than hosts as the cost to the *AVR*-parasite of being detected exceeds the cost to the host of being infected.

5.3.3 - Experimental coevolutionary microcosms and hypothesis testing – advantages of microcosms include short generations and controllable variables

Multi-generation oscillations in gene frequencies are difficult to observe in wild populations. They require long periods of observation, frequent sampling and accounting for many biotic and abiotic complicating factors. These issues can be alleviated somewhat by using laboratory-grown populations. Organisms with short generation times can be selected, sampling is inevitably quicker and easier and environments can be controlled and simplified to focus on one host species and one parasite species in constant or near-constant conditions.

Pathosystems that have been grown for multiple generations in controlled laboratory conditions include bacteria and phage (Lenski & Levin 1985, Bohannon et al. 1999, 2000, Elena & Lenski 2003, Brockhurst et al. 2006, Poullain et al. 2008), water-fleas and their parasites (Ebert et al. 2008) and moth larva and viral parasites (Boots & Mealor 2007, Boots et al. 2009). Questions addressed have include the effects of competition between host-strains with costly resistance (Lenski & Levin 1985) and the importance of spatial structure on both genotype co-existence (Brockhurst et al. 2006) and frequency and amplitude of oscillations in populations (Boots et al. 2009).

The moth-virus system had no reported genetic variation in host-parasite interaction. However a result from this system is that limited dispersal dampens genotype oscillation and makes stochastic loss of the virus less likely, supporting Model 3 hypotheses. The bacteria-phage systems include genetics somewhat similar to GFG interactions (discussed in Section 5.3.4). Brockhurst et al. (2006) reported that phage-resistant and susceptible bacteria could coexist in unmixed but not in mixed environments and suggested coexistence required transient phage-free refugia. This is equivalent to low dispersal leading to aggregation of genotypes, a stabilising factor in Model 3. Lenski & Levin (1985) reported that co-existence of resistant and susceptible bacteria requires resistance to have a cost. Costs of bacterial resistance are common in bacteria-phage systems (Elena & Lenski 2003).

5.3.4 - Microcosms and future hypothesis testing – bacteria-phage systems have similarities to GFG-genetics and could be used to test Model 1, 3 and 4 hypotheses

Microcosm studies of coevolution have already produced results relevant to the hypotheses in this thesis (Section 5.3.3). Here I suggest further studies that could be done. The simplicity, controllability, short generation times and ease of sampling of such systems are all reasons to consider doing this. A key issue with these systems is that they do not exhibit GFG genetics. However the bacteria-phage system comes close, in that naïve bacteria are vulnerable to phage but can evolve resistance and phage can sometimes evolve to overcome this resistance. The difference is that bacteria often evolve a final immunity to infection that phage are unable to overcome (Hofnung et al 1976). Such mutation may be difficult to avoid and would be undesirable if the system is being used to model GFG dynamics where parasites can generally evolve virulence alleles to avoid detection by any particular *RES*-gene.

The Model 1 hypothesis is that low disease incidence leads to *RES*/*res* stable polymorphism with fixed *AVR*, intermediate incidence leads to *RES*/*res* and *avr*/*AVR* stable polymorphism and high incidence leads to unstable polymorphism. In bacteria-phage systems, host density can be manipulated by nutrient levels and rate of flow in continuous culture. The Model 1 hypotheses applied to bacteria-phage genetics would be that at low incidence resistant and susceptible hosts co-exist, at intermediate incidence resistant hosts are more common and phage capable of overcoming resistance appear and at high

incidence resistant hosts go to fixation. This should be examined in a mixed (non-spatial) population, to remove one complicating factor and avoid refugia effects as discussed below (Brockhurst et. Al 2006).

A practical issue is that phage have both very high basic reproductive rates and very high costs of infection, as a successfully infecting lytic phage results in thousands of new phage and destroys the bacterial host (Elena & Lenski 2003). Thus local disease incidences would typically be either zero or virtually total. Model 1 predictions apply to low to intermediate disease incidence. In bacteria-phage systems, the range of host-densities that allow such densities would be narrow and difficult to manipulate.

A key difference between bacterial resistance and plant GFG-resistance is that in most studied cases bacterial resistance has a clear and consistent fitness cost (Elena & Lenski 2003), as opposed to common but difficult-to-observe environmentally labile costs (Bergelson & Purrington 1996, Laine & Barres 2013). This makes the system attractive for validating theoretical work involving such costs.

The Model 3 hypothesis is that spatial structure changes equilibrium frequencies and dampens oscillations. Both of these have already been observed in experimental systems, the oscillation-damping in Boots et al. (2009) and the altered equilibrium in a bacteria-phage system where susceptible hosts survived in an unmixed system and did not survive in a mixed system (Brockhurst et al. 2006). In the latter case it is not clear that coexistence is a permanent, dynamic effect rather than a temporary spatial refuge effect. However, since spatial variation for phage incidence (normally either zero or very high at local levels) is ubiquitous in the wild, this refuge effect is presumably permanent and important even if any specific refuge is temporary. It would be interesting to try to create a long-term spatially structured population in which local extinctions and recolonisations can be observed directly, rather than inferred from overall gene levels, and demonstrate that the system is stable for either bacterial resistance polymorphism alone or both bacterial resistance and phage virulence polymorphism. The dynamics of connected, differing environments (differing perhaps in nutrient levels and thus host density) could be investigated as part of this. Such a system would have a mixture of factors opposing deterministic and stochastic genotype loss.

5.4 - VARIABLE INCIDENCE AND STABILITY IN OTHER ECOLOGICAL INTERACTIONS

5.4.1 – Coevolution in agriculture – pathosystems occupy a continuum between wild systems, subsistence agriculture and modern agricultural monocultures

Model 1 suggests lower parasite basic reproductive rates are more likely to preserve polymorphism in GFG systems. Lower host density reduces effective parasite basic reproductive rate, as propogules are less likely to encounter hosts. This may help explain the different coevolutionary outcomes often observed in wild and agricultural pathosystems (Chapter 1, Section 1.4.2).

Resistance genes routinely fail in large-scale monocultures and for practical reasons many commercial farmers are unlikely to abandon such systems. I would expect resistance genes to exhibit the greatest longevity in crops grown in conditions most like wild pathosystems, i.e. in crops which are not grown in large, high-density cultures but are instead grown in small crops or at low density across wide areas. Such crops include the foods and fodders of traditional, subsistence-level agricultural practice.

Crops grown using traditional agricultural practices are intermediate between wild plant populations and modern agricultural crops in terms of species diversity and density (Malezieux 2012). Differences in disease dynamics between wild, subsistence and modern agricultural pathosystems present a chance to test the Model 1 hypotheses relating to parasite incidence (correlated with host density) and, more broadly, to test if and when switches between stable and unstable coevolution can be detected. Such studies could potentially shed light on when subsistence crops might experience the same kind of high-incidence diseases as large-scale agricultural crops, for instance when their density increases above a certain threshold or when related monocultures are grown near them.

In many parts of the developing world greater need owing to population growth and greater ability owing to advances in technique such as water-harvesting have led to an increase in the amounts, hence both the density and range, of crops grown by subsistence farmers (Bouma 2011). This increase in host prevalence and density may lead to rapid, abrupt switches in coevolutionary dynamics.

5.4.2 – Coevolution in human epidemiology – dispersal and diversity in malaria

In general, both my models and other studies (Rothman 2012) predict that increased host density and increased movement between populations will lead to parasites evolving to overcome host resistance. This suggests that limiting disease dispersal could reduce the ability of diseases to overcome local resistance.

As an example, malaria parasites have highly variable surface proteins and are thought to be under diversifying selection driven by both innate and vaccine-boosted immune responses (Takala & Plowe 2009). Model 3 results suggest malaria parasites are more likely to retain polymorphism for vulnerability to specific vaccines (at a metapopulation level) and be universally vulnerable to vaccines (at a local level) if gene-flow between parasite populations is limited. This implies breaking up populations of carrier mosquitoes may have beneficial impacts. This may be a useful strategy in malaria control – vector mosquito species require stagnant water to breed, bodies of water differ in their accessibility and draining and eliminating some bodies of stagnant water would clearly be easier than eliminating all such bodies. Recent reports on current and potential future spatial dynamics of malaria are Bousema et al. (2012) and Tonnang et al. (2010), respectively.

5.5 - CONCLUSIONS

The hypotheses generated in this thesis are about the dynamics of polymorphism when parasites negatively regulate their own incidence and possibly their host's density. In the real world such regulation may be most obvious manifesting as transient extinction-recolonisation dynamics, or may be surmised to have occurred from current high diversities and low densities of hosts and parasites. Key hypotheses are that reduced incidence corresponds to stable polymorphism, that large and spatially structured populations are resistant to stochastic genotype loss and that multiple stabilising factors are likely to commonly co-occur and reinforce stable polymorphism.

Directions for future research include more detailed ecology and genetics in deterministic models, more accurate spatial dynamics and densities in individual-based models, further analysis of both wild and agricultural pathosystems and further coevolutionary microcosm studies.

BIBLIOGRAPHY

Afzal, A. J., Wood, A. J. & Lightfoot, D. A. 2008 Plant receptor-like serine threonine kinases: roles in signalling and plant defence. *Molecular Plant-Microbe Interactions* **21**, 507-517.

Anderson, R. M. & May, R. M. 1981 The population dynamics of microparasites and their invertebrate hosts. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **291**, 451-524.

Apanius, V., Penn, D., Sleve, P. R., Ruff, L. R. & Potts, W. K. 1997 The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology* **17**, 179-224.

Aronson, D.G, Ermentrout, G. B. & Kopell, N. 1990 Amplitude response of coupled oscillators. *Physica D: Nonlinear Phenomena* **41**, 403-449.

Augspurger, C. K. 1983 Seed dispersal of the tropical tree *Platypodium elegans*, and the escape of its seedlings from fungal pathogens. *Journal of Ecology* **71**, 759-771.

Axtell, M. J. & Staskawicz, B. J. Initiation of *RPS2*-specified disease resistance in *Arabidopsis* is coupled to the AvrRpt2-directed elimination of RIN4. *Cell* **112**, 369-377.

Bakker, E. G., Toomajian, C., Kreitman, M. & Bergelson, J. 2006 A geneome-wide survey of *R* gene polymorphisms in *Arabidopsis*. *The Plant Cell* **18**, 1803-1818.

Barr-Eli, K. 1985 On the stability of coupled chemical oscillators. *Physica D: Nonlinear Phenomena* **14**, 242-252.

Bergelson, J. & Purrington, C. B. 1996 Surveying patterns in the cost of resistance in plants. *The American Naturalist* **148**, 536-558.

Bergelson, J., Kreitman, M., Stahl, E. A. & Tian, D. 2001 Evolutionary dynamics of plant *R*-genes. *Science* **292**, 2281-2285.

Bevan, J. R., Crute, I. R. & Clarke, D. D. 1993 Variation for virulence in *Erysiphe fischeri* from *Senecio vulgaris. Plant Pathology **42**, 622-635.*

Blanquart, F., Gandon, S. & Nuismer, S. L. 2012 The effects of migration and drift on local adaptation to a heterogeneous environment. *Journal of Evolutionary Biology* **25**, 1351-1363.

Bohannan, B. J. M. & Lenski, R. E. 1999 Effect of prey heterogeneity on the response of a model food chain to resource enrichment. *The American Naturalist* **153**, 73-82.

Bohannan, B. J. M. & Lenski, R. E. 2000 Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecology Letters* **3**, 362-377.

Boller, T. & Feliz, G. 2009 A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual Review of Plant Biology* **60**, 379-406.

Boots, M. & Mealor, M. 2007 Local interactions select for lower pathogen infectivity. *Science* **315**, 1284-1286.

Boots, M., Childs, D., Reuman, D. C. & Mealor, M. 2009 Local interactions lead to pathogen-driven change to host dynamics. *Current Biology* **19**, 1660-1664.

Bousema, T., Griffen, J. T., Sauerwein, R. W., Smith, D. L., Churcher, T. S., Takken, W., Ghani, A., Drakeley, C. & Gosling, R. 2012 Hitting hotspots: spatial targeting of malaria for control and elimination. *Public Library of Science: Medicine* **9**, e1001165.

Brading, P. A., Hammond-Kosack, K. E., Parr, A. & ones, J. D. G. 2000 Salicylic acid is not required for *Cf-2*- and *Cf-9*-dependent resistance of tomato to *Cladosporium fulvum. The Plant Journal **23**, 305-318.*

Briggs, C. J. & Hoopes, M. F. 2004 Stabilizing effects in spatial parasitoid-host and predator-prey models: a review. *Theoretical Population Biology* **65**, 299-315.

Brockhurst, M. A., Buckling, A & Rainey, P. B. 2006 Spatial heterogeneity and the stability of host-parasite coexistence. *Journal of Evolutionary Biology* **19**, 374-379.

Brown, J. K. M. 2002 Yield penalties of disease resistance in crops. *Current Opinion In Plant Biology* **5**, 339-344.

Brown, J. K. M. 2003A Little Else But Parasites. *Science* **289**, 1680-1681.

Brown, J. K. M. 2003B A cost of disease resistance: paradigm or peculiarity? *Trends In Genetics* **19**, 667-671.

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

Burdon, J. J. & Thrall, P. H. 1999 Spatial and temporal patterns in coevolving plant and pathogen associations. *The American Naturalist* **153**, 15-33.

Caplan, J, Padmanabhan, M. & Denish-Kumar, S. P. 2008 Plant NB-LRR receptors: from recognition to transcriptional programming. *Cell Host & Microbe* **3**, 126-135.

Carius, H. J., Little, T. J. & Ebert, D. 2001 Genetic variation in a host-parasite association: potential for coevolution and frequency-dependent selection. *Evolution* **55**, 1136-1145.

Catanzariti, A., Dodds, P. N., Ve, T., Kobe, B., Ellis, J. G. & Staskawicz, B. J. The AvrM effector from flax rust has a structured C-terminal domain and interacts directly with the M resistance protein. *Molecular Plant-Microbe Interactions* **23**, 29-57.

Chaboudez, P. & Burdon, J. J. 1995 Frequency-dependent selection in a wild plant-pathogen population. *Oecologia* **102**, 490-493.

Chin, K. M. & Wolfe, M. S. 1984A The spread of *Erysiphe graminis* f. sp. *hordei* in mixtures of barley varieties. *Plant Pathology* **33**, 89-100.

Chin, K. M. & Wolfe, M. S. 1984B Selection on *Erysiphe graminis* in pure and mixed stands of barley. *Plant Pathology* **33**, 535-546.

Chisholm, S. T., Coaker, G., Day, B. & Staskawicz, B. J. 2006 Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* **124**, 803-814.

Conrath, U., Beckers, G., Flors, V. & 13 others. 2006 Priming: getting ready for battle. *Molecular Plant-Microbe Interactions* **19**, 1062-1071.

Damgaard, C. 1999 Coevolution of a plant host-pathogen gene-for-gene system in a metapopulation model without cost of resistance or cost of virulence. *Journal of Theoretical Biology* **201**, 1-12.

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

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

Dodds, P. N. & Rathjen, J. P. Plant immunity: towards an integrated view of plant-pathogen interactions. 2010 *Nature Reviews: Genetics* **11**, 539-548.

Elena, S. F. & Lenski, R. E. 2002 Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews: Genetics* **4**, 457-469.

Ebert, D. 2008 Host-parasite coevolution: insights from the *Daphnia*-parasite model system. *Current Opinion in Microbiology* **11**, 290-301.

Ellstrand, N. & Antonovics, J. 1984 Experimental studies of the evolutionary significance of sexual reproduction. I. A test of the frequency-dependent selection hypothesis. *Evolution* **38**, 103-115.

Feng, F. & Zhou, J. M. 2012 Plant-bacterial pathogen interactions mediated by type III effectors. *Current Opinions in Plant Biology*, **15**, 469-476.

Flor, H. 1955 Host-parasite interactions in flax-rust – its genetics and other consequences. *Phytopathology* **45**, 680-685.

Flor, H. 1971 Current status of the gene-for-gene concept. *Annual Review of Phytopathology* **9**, 275-296.

Frank, S. A. 1992 Models of plant-pathogen coevolution. *Trends in Genetics* **8**, 213-219.

Frank, S. A. 1993 Coevolutionary genetics of plants and pathogens. *Evolutionary Ecology* **7**, 45-75.

Futuyma, D. J. 1996 *Evolutionary Biology, Third Edition*. Massachusetts, USA: Sinauer Associates.

Gandon, S., Capowiez, Y, Dubois, Y, Michalakis, Y. & Olivieri, I. 1996 Local adaptation and gene-for-gene coevolution in a metapopulation model. *Proceedings of the Royal Society B* **263**, 1003-1009.

Gandon, S. 2002 Local adaptation and the geometry of host–parasite coevolution. *Ecology Letters* **5**, 246-256.

Gandon, S. & Michalakis, Y. 2002 Local adaptation, evolutionary potential and host–parasite coevolution: interactions between migration, mutation, population size and generation time. *Journal of Evolutionary Biology* **15**, 451-462.

Gandon, S. & Nuismer, S. L. 2009 Interactions between genetic drift, gene flow and selection mosaics drive parasite local adaptation. *The American Naturalist* **173**, 212-224.

Garcia-Arenal, F. & Fraile, A. 2013 Trade-offs in host range evolution of plant viruses. *Plant Pathology* doi: 10.1111/ppa.12104.

Glazebrook, J. 2005 Contrasting mechanisms of defence against biotrophic and necrotrophic pathogens. *Annual Review of Phytopathology* **43**, 205-227.

Greischar, M. A. & Koskella, B. 2007 A synthesis of experimental work on parasite local adaptation. *Ecology Letters* **10**, 418-434.

Groth, J. V. 1976 Multilines and "super races": a simple model. *Phytopathology* **66**, 937-939.

Hedrick, P. W. 2009 G*enetics of Populations, Fourth Edition*. Massachusetts, USA: Jones & Bartlett.

Hill, A. V. S. 1997 The genomics and genetics of human infectious disease susceptibility. *Annual Review of Genomics and Human Genetics* **2**, 373-400.

Holt, R. D. & Pickering, J. 1985 Infectious disease and species coexistence: a model of Lotka-Volterra form. *The American Naturalist* **126**, 196-211.

Holub, A. B. 2001 The arms race is ancient history in *Arabidopsis*, the wild flower. *Nature Review Genetics* **2**, 516-527.

Hovmoller, M. S., Yahyaoui, A. H., Milius, E. A. & Justesen, A. F. 2008 Rapid global spread of two aggressive strains of a wheat rust fungus. *Molecular Ecology* **17**, 3818-3826.

Huang, R., Kranz, J. & Welz, H. G. 1994 Selection of pathotypes of *Erysiphe graminis* f.sp. *hordei* in pure and mixed stands of spring barley. *Plant Pathology* **43**, 458-470.

Huang, X. Q. & Roder, M. S. 2004 Molecular mapping of powdery mildew resiatnce genes in wheat: a review. *Euphytica* **137**, 203-223.

Huang, Y. J., Balesdent, M. H., Li, Z. Q., Evans, N., Rouxel, T. & Fitt, B.D.L. 2010 Fitness cost of virulence differs between the *AvrLm1* and *AvrLm4* loci in *Leptosphaeria maculans* (phoma stem canker of oilseed rape). *European Journal of Plant Pathology* **126**, 279-291.

Huckelhoven, R., Eichmann, R., Weis, C., Hoefle, C. & Proels, R. K. 2013 Genetic loss of susceptibility: a costly route to disease resistance? *Plant Pathology* doi: 10.1111/ppa.12103.

Jayakar, S. D. 1970 A mathematical model for interaction of gene frequencies in a parasite and its host. *Theoretical Population Biology* **1**, 140-164.

Jiang, R. H. Y., Tyler, B. M. & Govers, F. 2006 Comparative analysis of *Phytophthora* genes encoding secreted proteins reveals conserved synteny and lineage-specific gene duplications and deletions. *Molecular Plant-Microbe Interactions* **19**, 1311-1321.

Jones, J. D. G. & Dangl, J. L. 2006 The plant immune system. *Nature* **444**, 323-329.

Juroszek, P. & von Tiedemann, A. 2013 Plant pathogens, insect pests and weeds in a changing global climate: a review of approaches, challenges, research gaps, key studies and concepts. *The Journal of Agricultural Science* **151**, 163-188.

Kelly, J. D. & Vallejo, V. A. 2004 A comprehensive review of the major genes conditioning resistance to anthracnose in common bean. *HortScience* **39**, 1196-1207.

Keesing, F., Holt, R. D. & Ostfeld, R. S. 2006 Effects of species diversity on disease risk. *Ecology Letters* **9**, 485-498.

Kirby, G. C. & Burdon, J. J. 1997 Effects of mutation and random drift on Leonard's gene-for-gene coevolution model. *Phytopathology* **87**, 488-493.

Kjaer, B., Jensen, H. P., Jensen, J. & Jorgenson, H. 1990 Associations between three *ml-o* powdery mildew resistance genes and agronomic traits in barley. *Euphytica* **46**, 185-193.

Kniskern, J. M., Barrett, L. G. & Bergelson, J. 2011 Maladaptation in wild populations of the generalist plant pathogen *Pseudomonas syringae*. *Evolution* **65**, 818-830.

Kohler, S. L. & Hoiland, W. K. 2001 Population regulation in an aquatic insect: the roles of disease. *Ecology* **82**, 2294-2305.

Koskella, B. & Lively, C. M. 2007 Advice of the rose: experimental coevolution of a trematode parasite and its snail host. *Evolution* **61**, 152-159.

Krebs, C. J., Boonstra, R., Boutin, S. & Sinclair, A. R. E. 2001 What drives the 10-year cycle of snowshoe hares? *BioScience* **51**, 25-35.

Laine, A. L. 2004 Resistance variation within and among host populations in a plant–pathogen metapopulation: implications for regional pathogen dynamics. *Journal of Ecology* **92**, 9990-1000.

Laine, A. L. 2005 Spatial scale of local adaptation in a plant-pathogen metapopulation. *Journal of Evolutionary Biology* **18**, 930-938.

Laine, A. L. & Barres, B. 2013 Epidemiological and evolutionary consequences of life-history trade-offs in pathogens. *Plant Pathology* doi: 10.1111/ppa.12129.

Laine, A. L. & Hanski, I. 2006 Large-scale spatial dynamics of a specialist plant pathogen in a fragmented landscape. *Journal of Ecology* **94**, 217-226.

Leggett, H. C., Buckling, A., Long, G. H. & Boots, M. 2013 Generalism and the evolution of parasite virulence. *Trends in Ecology and Evolution* in press.

Lenski, R. E. & Levin, B. R. 1985 Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. *The American Naturalist* **125**, 585-602.

Leonard, K. J. 1977 Selection pressures and plant pathogens. *Annals of the New York Academy of Sciences* **287**, 207-222.

Leonard, K. J. 1994 Stability of equilibria in a gene-for-gene coevolution model of host-parasite interactions. *Phytopathology* **84**, 70-77.

Lewontin, R. C. 1958 A general method for investigating the equilibrium of gene frequency in a population. *Genetics* **43**, 419-434.

Mackey, D., Belkhadir, Y., Alonso, J. M., Ecker, J. R. & Dangl, J. L. 2003 *Arabdiopsis* RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. *Cell* **112**, 379-389.

Malezieux, Z. 2012 Designing cropping systems from nature. *Agronomy for sustainable development* **32**, 15-29.

May, R. M. & Anderson, R. M. 1983 Epidemiology and genetics in the coevolution of parasites and hosts. *Proceedings of the Royal Society B* **219**, 281-313.

Mitchell, C. E., Tilman, D. & Groth, J. V. 2002 Effects of grassland plant species diversity, abundance and composition on foliar fungal disease. *Ecology* **83**, 1713-1726.

Monaghan, J. & Zipfel, C. 2012 Plant pattern recognition receptor complexes at the plasma membrane. *Current Opinion in Plant Biology* **15**, 349-357.

Montarry, J., Hamelin, F. M., Glais, I., Corbiere, R. & Andrivon, D. 2010 Fitness costs associated with unnecessary virulence factors and life history traits: evolution insights from the potato late blight pathogen *Phytophthora infestans*. *BMC Evolutionary Biology* **10**, e283.

Morgan, A. D., Gandon, S. & Buckling, A. 2005 The effect of migration on local adaptation in a coevolving host-parasite system. *Nature* **437**, 253-256.

Mundt, C. C. 2002 Use of multiline cultivars and cultivar mixtures for disease management. *Annual Review of Phytopathology* **40**, 381-410.

Murray, J. D. 1989 *Mathematical Biology I: An Introduction, Third Edition*. Berlin, Germany: Springer.

Nuismer, S. L. & Otto, S. P. 2005 Host-parasite interactions and the evolution of gene expression. *Public Library of Science Biology* **3**, e203.

Nuismer, S. L. 2006 Parasite local adaptation in a geographic mosaic. *Evolution* **60** 24-30.

Packer, A. & Clay, K. 2000 Soil pathogens and spatial patterns of seedling mortality in a temperate tree. *Nature* **404**, 278-281.

Packer, C., Holt, R. D., Hudson, P. J., Lafferty, K. D. & Dobson, A. P. 2003 Keeping the herds healthy and alert: implications of predator control for infectious disease. *Ecology Letters* **6**, 797-802.

Pieterse, C. M. J., Van der Does, D., Zamioudis, C., Leon-Reyes, A. & Van Wees, S. C. M. 2012 Hormonal modulation of plant immunity. *Annual Review of Cell and Developmental Biology* **28**, 489-521.

Poullain, V., Gandon, S., Brockhurst, M., Buckling, A. & Hochberg, M. E. 2007 The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage. *Evolution* **2008**, 1-11.

Rayapuram, C., Jensen, M. K., Maiser, F., Shanir, J. V. & Hornshoj, H. 2012 Regulation of basal resistance by a powdery mildew-induced cysteine-rich receptor-like protein kinase in barley. *Molecular Plant Pathology* **13**, 135-147.

Reymond, P. & Farmer, E. E. 1998 Jasmonate and salicylate as global signals for defense gene expression. *Current Opinion in Plant Biology* **1**, 404-411.

Roughgarden, J. 1996 *Theory of Population Genetics and Evolutionary Ecology: An Introduction*. New York, USA: Prentice Hall.

Salathe M., Scherer A. & Bonhoeffer S. 2005 Neutral drift and polymorphism in gene-for-gene systems. *Ecology letters* **8**, 925-932.

Sarcristan, S. & Garcia-Arenal, F. 2008 the evolution of virulence and pathogenicity in plant-pathogen populations. *Molecular Plant Pathology* **9**, 369-384.

Sarcristan, S., Vigouroux, M., Pedersen, C., Skamnioti, P., Thordal-Christensen, H., Micali, C., Brown, J. K. M. & Ridout, C. J. 2009 Coevolution between a family of parasite virulence effectors and a class of LINE-1 retrotransposons. *Public Library of Science One* **4**, e7463.

Sasaki, A. 2000 Host-parasite coevolution in a multi-locus gene-for-gene system. *Proceedings of the Royal Society B* **267**, 2183-2188.

Sasaki, A., Hamilton, W. D. & Ubeda, F. 2002 Clone mixtures and a pacemaker: new facets of Red-Queen theory and ecology. *Proceedings of the Royal Society B* **269**, 761-772.

Srichumpa, P., Brunner, S., Keller, B. & Yahiaoui, N. 2005 Allelic series of four powdery mildew resistance genes at the *Pm3* locus in hexaploid bread wheat. *Plant Physiology* **139**, 885-895.

Seeholzer, S., Tsuchimatsu, T., Jordan, T., Bieri, S., Pajonk, S., Yang, W., Jahoor, A., Shimizu, K. K., Keller, B. & Schulze-Lefert, P. 2010 Diversity at the *Mla* powdery mildew resistance locus from cultivated barley reveals sites of positive selection. *Molecular Plant-Microbe Interactions* **23**, 497-509.

Segarra, J. 2005 Stable polymorphisms in a two-locus gene-for-gene system. *Phytopathology* **95**, 728-736.

*Skamnioti, P. & Ridout, C. J. 2005 Microbial avirulence determinants: guided missiles or antigenic flak? Molecular Plant Pathology **6**, 551-559.*

Smedegaard-Petersen, V. & Tolstrup, K. 1985 The limiting effects of disease resistance on yield. *Annual Review of Phytopathology* **23**, 475-490.

Spiess, E. P. 1968 Low frequency advantage in mating of *Drosophila seudoobscura* karyotypes. *The American Naturalist* **102**, 363-379.

St Clair, D. 2010 Quantatative disease resistance and quantitative resistance loci in breeding. *Annual Review of Phytopathology* **48**, 247-268.

Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M. & Bergelson, J. 1999 Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* **400**, 667-671.

Stukenbrock, E. H. & McDonald, B. A. 2008 The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology* **46**, 75-100.

Tack, A. G. M., Thrall, P. H., Barrett, L. G, Burdon, J. J. & Laine, A. L. 2012 Variation in infectivity and aggressiveness in space and time in wild host-pathogen systems: causes and consequences. *Journal of Evolutionary Biology* **25**, 1918-1936.

Tellier, A. & Brown, J. K. M. 2007A Stability of genetic polymorphism in host-parasite interactions. *Proceedings of the Royal Society B* **274**, 809-817.

Tellier, A. & Brown, J. K. M. 2007B Polymorphism in multilocus host-parasite coevolutionary interactions. *Genetics* **177**, 1777-1790.

Tellier, A. & Brown, J. K. M. 2009 The influence of perenniality and seed banks on polymorphism in plant-parasite interactions. *The American Naturalist* **174**, 769-779.

Tellier, A. & Brown, J. K. M. 2011 Spatial heterogeneity, frequency-dependent selection and polymorphism in host-parasite interactions. *BMC Evolutionary Biology* **11**, e319.

Thaler, J. S., Karban, R., Ullman, D. E., Boege, K. & Bostock, R. M. 2002 Cross-talk between jasmonate and salicylate plant defence pathways: effects on several plant parasites. *Oecologia* **131**, 227-235.

Thaler, J. S., Humphrey, P. T. & Whiteman, N. K. 2012 Evolution of jasmonate and salicylate crosstalk. *Trends in Plant Science* **17**, 260-270.

Thompson, J. N. & Burdon, J. J. 1992 Gene-for-gene coevolution between plants and parasites. *Nature* **360**, 121-125.

Thrall, P. H. & Burdon, J. J. 2000 Effect of resistance variation in natural plant host-pathogen metapopulation on disease dynamics. *Plant pathology* **49**, 767-773.

Thrall, P. H. & Burdon, J. J. 2002 Local adaptation in the *Linum marginale-Melampsora lini* host-pathogen interaction. *Evolution* **56**, 1340-1351.

Thrall, P. H. & Burdon, J. J. 2002 Evolution of gene-for-gene systems in metapopulations: the effect of spatial scale of host and pathogen dispersal. *Plant Pathology* **51**, 169-184.

Thrall, P H. & Burdon, J. J. 2003 Evolution of virulence in a plant host-pathogen metapopulation. *Science* **299**, 1735-1737.

Thrall, P. H., Laine, A. L., Ravensdale, M., Nemri, A., Dodds, P. L., Barrett, L. G. & Burdon, J. J. 2012 Rapid genetic change underpins antagonistic coevolution in a natural host-pathogen metapopulation. *Ecology letters* **15**, 425-435.

Tian, D., Traw, M. B. Chen, J. Q., Krietmen, M. & Bergelson, J. 2003 Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**, 74-77.

Tinbergen, N. 1951 *The Study of Instinct*. Oxford, UK: Oxford University Press.

Tonnang, H. E. Z., Kangalawe, R. Y. M. & Yanda, P. Z. 2010 Predicting and mapping malaria under climate change scenarios: the potential redistribution of malaria vectors in Africa. *Malaria Journal* **9**, e111.

Vogel, J. P., Raab, T. K., Somerville, C. R. & Somerville, S. C. 2004 Mutations in *PMR5* result in powdery mildew resistance and altered cell wall composition. *The Plant Journal* **40**, 968–978.

Vos, I. A., Pieterse, C. M. J. & van Wees, S. C. M. 2013 Costs and benefits of hormone-regulated plant defence. *Plant Pathology* doi: 10.1111/ppa.12105.

Wei, F., Wing, R. A. & Wise, R. P. 2002 Genome dynamics and the evolution of the *Mla* (powdery mildew) resistance locus in barley. *The Plant Cell* **14**, 1903-1917.

Wichman, G. & Bergelson, J. 2004 Effector Genes of *Xanthamonas axonopodis* pv. *Vesicatoria* promote transmission and enhance other fitness traits in the field. *Genetics* **166**, 693-706.

Wilensky, U. 1999 NetLogo. http://ccl.northwestern.edu.netlogo/ Center for Connected Learning and Computer-based Modeling, Northwestern University. IL, USA.

Wingen, L. U., Brown, J. K. M. & Shaw, M. 2007 The population genetic structure of clonal organisms generated by exponentially bounded and fat-tailed dispersal. *Genetics* **177**, 435-448.

Wingen, L. U., Shaw, M. & Brown, J. K. M. 2013 Long-distance dispersal and its influence on adaptation to host resistance in a heterogeneous landscape. *Plant Pathology* **62**, 9-20.

Wolfe, M. S. & Barrett, J. A. 1980 Can we lead the pathogen astray? *Plant Disease* **64**, 148-151, 153-155.

Workneh, F., Tylka, G. L., Yang, X. B, Faghihi, J. & Ferris, J. M. 1999 Regional assessment of soybean brown stem rot, *Phytophthora sojae*, and *Heterodera glycines* using area-frame sampling: prevalence and effects of tillage. *Phytopathology* **89**, 204-211.

Yahiaoui, N., Brunner, S. & Keller, B. 2006 Rapid generation of new powdery mildew resistance genes after wheat domestication. *The Plant Journal* **47**, 85-98.

Ye, T. Z., Yang, R. C. & Yeh, F. C. 2003 Coevolution in natural pathosystems: effects of dominance on host-pathogen interactions. *Phytopathology* **93**, 633-639.