The effects of contact patterns and genetic specificity on host and parasite evolution

Ben Ashby
Linacre College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy, D.Phil
Trinity 2014

Acknowledgements

I would like to express my gratitude to my two supervisors, Sunetra Gupta and Angus Buckling, for their guidance and support, especially through the many ups and downs of peer review. I am also grateful to a large number of people who have provided me with extremely helpful discussions and comments on the work presented herein, especially Mike Boots, Mike Brockhurst, Damien Farine, Andy Gardner, Matt Keeling, Kayla King, Britt Koskella, Bridget Penman, Stuart West and Hywel Williams. On a more personal note, I thank my friends and family for their love and support over the course of my D.Phil. Lastly, I wish to thank my long-suffering partner Becky for her enduring patience, encouragement and unique scientific insight.

Abstract

The effects of contact patterns and genetic specificity on host and parasite evolution

Ben Ashby

D.Phil

Linacre College, University of Oxford, Trinity 2014

Many bacteria, viruses and other parasites cause severe morbidity or mortality in their host populations, creating strong selection for physiological or behavioural mechanisms to avoid disease. Likewise, changes in host susceptibility and contact patterns can dramatically influence the spread of infectious diseases, and hence selection for traits such as virulence and infectivity range in parasites. Understanding how ecological and evolutionary changes in one population affect selection in another represents a key challenge for theoreticians and empiricists alike, and is essential for gaining further insights into host-parasite relationships.

This thesis contains theoretical models that explore how genetic and environmental factors shape the evolutionary and coevolutionary dynamics of hosts and parasites. In particular, the roles of genetic specificity (i.e. genotype-by-genotype interactions) and population mixing patterns are investigated, using both mathematical models and computer simulations. A broad range of scenarios are covered, including the coevolution of broad resistance and infectivity ranges (generalism), the persistence of coevolutionary cycling and the maintenance of sex, the effects of mating behaviour on disease prevalence and evolution, and the evolution of sexual and social behaviour. The models presented herein develop our understanding of host-parasite relationships and highlight the importance of genetic interactions and ecological feedbacks.

Contents

List of Figures

So, naturalists observe, a flea
Has smaller fleas that on him prey;
And these have smaller still to bite 'em.
And so proceed *ad infinitum.*
- Jonathan Swift, *On Poetry: A Rhapsody*

Chapter 1

Introduction

1.1 Host- and parasite-mediated selection

Host-parasite relationships appear to be near-ubiquitous in nature, with many parasites[1] known to cause severe morbidity and mortality in their host populations. For example, estimates suggest that approximately 20% of the marine biomass is killed by viruses on a daily basis and that 10^{23} viral infections occur in the ocean every second (Suttle, 2007). In total, over 1,400 species of parasites have been identified that can infect humans (Woolhouse and Gowtage-Sequeria, 2005) and infectious diseases are currently responsible for approximately one in seven deaths, with tuberculosis alone claiming 1.4 million lives in 2011 (Zumla et al., 2013). While not all parasites kill their hosts directly, many still cause debilitating illnesses that reduce competitive ability or inhibit reproduction, which can lead to sudden population crashes (Weigler et al., 1988; Hudson et al., 1998).

Parasites that cause morbidity or mortality are expected to exert a strong force of selection on their hosts and have therefore been implicated in the evolution of a number of biological phenomena. The complex immune systems of plants and animals are obvious examples of evolutionary responses to parasitism, but several other behavioural and physiological mechanisms that enable hosts to avoid, resist or tolerate infectious agents have also evolved to limit the negative impacts of disease. In particular, patterns of social and sexual contact are expected to be influenced by

[1]Throughout, 'parasite' should be interpreted broadly as any infectious agent that causes disease, including viruses, bacteria, protozoa, fungi and helminths.

the prevalence and severity of infectious diseases, and so many evolutionary biologists have speculated that parasitism plays an important role in shaping sexual selection, mate choice and host sociality (Freeland, 1976; Hamilton and Zuk, 1982; Clayton, 1991; Sheldon, 1993; Able, 1996; Loehle, 1997; Knell, 1999; Thrall et al., 2000; Boots and Knell, 2002; Kokko et al., 2002; Ashby and Gupta, 2013). For example, Hamilton and Zuk (1982) suggested that secondary sex traits such as bright plumage or other male ornaments have evolved to signal the presence of 'good' (resistance) genes to prospective mates, although others have noted that these traits may simply signal current infection status (Loehle, 1997). It is also thought that the sickle cell trait is highly prevalent in sub-Saharan Africa because humans with a single copy of the recessive mutation experience high levels of protection from severe malaria, which is extremely common in the region (Allison, 1954). Individuals who are homozygous for the sickle cell trait suffer from chronic anaemia and a high probability of childhood mortality, reinforcing the notion that parasites can exert a strong force of selection on their hosts and lead to the evolution of costly defence mechanisms.

Similarly, host defence mechanisms along with changes in population size and structure can exert strong selection on parasites. Given that parasites typically have much larger population sizes and shorter generation times than their hosts, they are expected to rapidly evolve novel traits to overcome defence mechanisms and to adapt to constantly changing host populations. The influenza viruses, for example, elicit long-lasting immune responses in humans and consequently experience strong selection for antigenic variation to facilitate immune evasion (Smith et al., 2004). Many predictions for parasite evolution have focused on the question of virulence, with the scale of dispersal, within-host competition and trade-offs between transmissibility and the infectious period all known to influence selection for this trait (Levin and Pimentel, 1981; Anderson and May, 1982; Bremermann and Pickering, 1983; May and Anderson, 1983; Nowak and May, 1994; Boots and Sasaki, 1999; O'Keefe and

Antonovics, 2002). In particular, greater dispersal tends to select for increased virulence (Boots and Sasaki, 1999), which has important implications for public health in an increasingly globalised society. While evolution is often seen as a gradual process, dramatic changes in virulence can occur over relatively short time scales (Boots et al., 2004). Notable examples include the causative agents of myxomatosis (myxoma virus) and syphilis (*Treponema pallidum*), which both exhibited rapid reductions in virulence following their introductions to rabbit and human populations, respectively (May and Anderson, 1983; Knell, 2004). The myxoma-rabbit system provides strong evidence for a trade-off between transmissibility and infectious period (this explains why intermediate levels of virulence have evolved; May and Anderson, 1983). In contrast, syphilis is thought to have become milder due to disease-avoidance behaviour by European humans in the 15th century (Knell, 2004), which indicates that mate choice can have a profound impact on parasite evolution (this theory is explored in more detail in chapter 7).

Along with virulence, conditions that allow parasites to adapt to infect novel genotypes through host switches or range expansion are of great interest to evolutionary biologists and epidemiologists alike (Woolhouse and Gowtage-Sequeria, 2005; Benmayor et al., 2009; Hall et al., 2011a; Herfst et al., 2012; Poullain and Nuismer, 2012; Russell et al., 2012; chapter 2 herein). While recent 'gain of function' experiments with avian influenza viruses, which aim to understand the mutations that allow shifts to mammalian hosts, have been the centre of considerable debate regarding the relative risks and rewards of this research (Herfst et al., 2012; Russell et al., 2012; Rey et al., 2013), the majority of systems that are studied do not pose a significant threat to public health. Much of the research on the experimental evolution of parasites has involved bacteriophages (phages), which typically exhibit rapid adaptation to changes in bacterial populations (Buckling et al., 2009; Koskella and Brockhurst, 2014). Crucially, several studies involving bacteria and phages have demonstrated

the importance of reciprocal adaptations (i.e. coevolution) for the emergence of novel traits such as broader resistance and infectivity ranges (Poullain et al., 2008; Paterson et al., 2010; Hall et al., 2011a). For example, Paterson et al. (2010) showed that the rate of molecular change in phages approximately doubles when bacteria are allowed to coevolve, while Poullain et al. (2008) and Hall et al. (2011a) showed that broader infectivity ranges emerge if hosts are not held in evolutionary stasis. These studies (along with many others, e.g. Anderson and May, 1982; May and Anderson, 1983; Thompson, 1994; Dybdahl and Lively, 1998; Gandon et al., 2002; Bonds et al., 2005; Buckling et al., 2006; Koskella and Lively, 2007; Best et al., 2009; Prado et al., 2009; King et al., 2009; Koskella and Brockhurst, 2014) illustrate the need to understand host-parasite interactions in a coevolutionary context due to the presence of genotype-by-genotype interactions and ecological feedbacks, which both play a fundamental role in shaping selection for novel traits in hosts and parasites.

1.2 Antagonistic coevolution and the importance of ecological feedbacks

Much of today's interest in host-parasite coevolution can trace its roots back to the 1970s and the realisation that a constantly changing biotic environment could resolve one of the most prominent questions in evolutionary biology: why do organisms reproduce sexually? Although coevolution had been discussed prior to this period by many renowned scientists, including Darwin, limited progress had been made regarding specific predictions for coevolving populations (Thompson, 1994). Reciprocal adaptations were widely accepted as the cause of specialisation among flowers and their insect pollinators or among predators and their prey (Thompson, 1994), and a genetic basis for infectivity had been identified among plant-pathogens (Flor, 1956). Yet, few had fully grasped the wider importance of antagonistic coevolution until sev-

eral individuals (Hamilton, 1975; Levin, 1975; Clarke, 1976; Jaenike, 1978) noted that negative frequency-dependent selection among genes governing host-parasite specificity could produce conditions that would allow sexually-reproducing individuals to offset the "twofold cost of males" (Maynard Smith, 1978). Bell (1982), referred to this as the 'Red Queen Hypothesis' (RQH), due to the earlier observation by Van Valen (1973) that interactions between species are analogous to the Red Queen's description of Looking-Glass Land in Lewis Carroll's *Through the Looking Glass*: "it takes all the running you can do to keep in the same place". In other words, sexual reproduction enables hosts to produce genetically diverse offspring that are more likely to avoid contemporaneous parasites compared to clonal populations.

Hamilton (1980) provided the first explicit demonstration of this principle using a theoretical model, which has led to an ongoing debate as to the generality of the Red Queen Hypothesis (RQH) for the maintenance of sex (Lively, 2010a). Much of this debate has focused on whether theoretical models exploring the RQH make realistic assumptions about population genetics (Otto and Nuismer, 2004) and dynamics (May and Anderson, 1983). For example, many models of the RQH have been implemented using traditional population genetics approaches that omit population dynamics and hence do not incorporate ecological feedbacks (Hamilton, 1980; Hamilton et al., 1990; Otto and Nuismer, 2004; Agrawal, 2009; but see Lively, 2010b). May and Anderson (1983) demonstrated that the inclusion of realistic epidemiological dynamics greatly reduces the conditions under which sex is likely to be favoured. Other issues have also been raised with existing theoretical approaches, such as the preference for discrete-time models that are inherently prone to oscillatory dynamics (Kouyos et al., 2007) and the lack of diploid models that account for the effects of segregation (Agrawal, 2009). Addressing these issues will be a major aspect of future theoretical work on the RQH.

While the maintenance of sex as envisaged by the RQH was responsible for much

of the initial growth in the literature on host-parasite coevolution, many soon realised that these antagonistic relationships were of much broader interest and could be used to explain a variety of other biological phenomena, including local adaptation (Gandon and Michalakis, 2002) and the maintenance of diversity (Frank, 1993b; Penn and Potts, 1999; Penman et al., 2013). Considerable attention has also been paid towards the general coevolutionary dynamics that emerge from host-parasite interactions (Sasaki, 2000; Agrawal and Lively, 2002, 2003; Fenton and Brockhurst, 2007; Fenton et al., 2009, 2012). As with the majority of theory relating to the RQH, ecological feedbacks are often omitted from these more general models of coevolution. However, the importance of ecological feedbacks is well established in epidemiological scenarios (Anderson and May, 1991), so a number of authors have begun to include more realistic population dynamics in models of host-parasite coevolution (Best et al., 2009; Gandon and Day, 2009; Boots et al., 2014).

A fundamental component in many of the aforementioned studies on the RQH and more general coevolutionary dynamics is the precise nature of genotype-by-genotype interactions between hosts and parasites, otherwise known as 'genetic specificity'. Thus, it is important to understand the genetic basis of infectivity in real host-parasite systems and how different forms of specificity shape coevolutionary dynamics.

1.3 Genetic specificity of host-parasite interactions

Considerable evidence indicates that there exists a strong genetic basis for resistance and infectivity in many host-parasite systems, including plants and plant-pathogens (Flor, 1956; Thompson and Burdon, 1992), bacteria and phages (Flores et al., 2011; Scanlan et al., 2011), molluscs and trematodes (Dybdahl and Lively, 1998; King et al., 2011), crustaceans and bacteria (Carius et al., 2001; Luijckx et al., 2013) and even among humans (Allison, 1954; Hill et al., 1991). As such, considerable effort has been

made by empiricists to reveal the underlying patterns of genetic specificity in real populations and by theoreticians to understand how different forms of specificity shape coevolutionary dynamics. Recent advances in sequencing have allowed empiricists to identify mutations that emerge in coevolving communities, revealing complex links between genotype and phenotype (Brockhurst et al., 2010; Paterson et al., 2010; Scanlan et al., 2011), which presents problems when trying to infer patterns of genetic specificity from coevolutionary dynamics alone (Frank, 1993c, 1996; Heath and Nuismer, 2014). These relationships are also likely to depend on environmental factors such as temperature (Zhang and Buckling, 2011) and resource availability (Lopez-Pascua and Buckling, 2008; Gomez and Buckling, 2011; Lopez-Pascua et al., 2012; Harrison et al., 2013), but our understanding of genotype-by-genotype-by-environment interactions is currently poor. In addition, studies of both natural and laboratory populations show strong evidence of local adaptation, which indicates that multiple routes to resistance are likely to exist and that antagonistic coevolution can allow segregated populations to diverge (Buckling and Rainey, 2002b; Thrall and Burdon, 2002; King et al., 2009; Vos et al., 2009; Koskella et al., 2011). However, Antonovics et al. (2013) noted that our understanding of how patterns of resistance and infectivity emerge is likely to be hindered by the fact that hosts may gain non-evolved resistance if parasites specialise on other hosts. This appears to be the case among water snails (*Potamopyrgus antipodarum*) and parasitic trematodes (*Microphallus* sp.), as parasites only coevolve with their hosts in shallow-water, yet deep-water snails are less susceptible to infection (King et al., 2009).

Theoreticians have demonstrated that the nature of genetic specificity is crucial for shaping coevolutionary dynamics, and plays a fundamental role in the maintenance of sex, local adaptation and sympatric diversity (Parker, 1994; Damgaard, 1999; Agrawal and Lively, 2002; Gandon and Michalakis, 2002; Otto and Nuismer, 2004). The majority of models of host-parasite specificity are based on either the

'matching alleles' (MA) or 'gene-for-gene' (GFG) frameworks. The MA framework is based on self/non-self recognition mechanisms among animals and assumes that infectivity depends on how well parasites match their hosts (Hamilton, 1980; Frank, 1993c; Penn and Potts, 1999; Grosberg and Hart, 2000). This means that different parasite genotypes specialise on dissimilar hosts, which can lead to negative frequency-dependent (fluctuating) selection. In other words, parasites track the most common host genotypes, which selects for rare alleles in the host. Examples of MA-like specificity are found among many invertebrate host-parasite systems (Dybdahl and Lively, 1998; Luijckx et al., 2013) and in the major histocompatibility complex (MHC), which plays a vital role in the vertebrate immune system (Hill et al., 1991; Penman et al., 2013).

The GFG framework differs from the MA model by assuming that hosts and parasites can exhibit various degrees of generalism, so that certain genotypes are able to resist or infect a broader range of genotypes than others. This can lead to a series of reciprocal adaptations as part of a coevolutionary arms race, but broader resistance and infectivity ranges are often associated with fitness costs (Chao et al., 1977; Webster and Woolhouse, 1999; Thrall and Burdon, 2003; Lopez-Pascua and Buckling, 2008; Poullain et al., 2008), which can ultimately curtail directional selection (Hall et al., 2011b). The GFG theory is based upon observations of plant immune systems, which show evidence of arms race dynamics with certain infectious agents (Flor, 1956; Thompson and Burdon, 1992; Thrall and Burdon, 2002; Jones and Dangl, 2006). Some plants rely on pattern recognition receptors to recognise the molecular patterns (e.g. flagellin) of certain parasites, but these can be overcome during infection by the delivery of effector molecules that interfere with host immunity. Successful parasites may then have to breach a second line of defence if hosts are able to respond to these effectors (Jones and Dangl, 2006).

Similarly, different bacterium and phage genotypes vary in their resistance and

infectivity ranges in a GFG-like manner (Flores et al., 2011; Scanlan et al., 2011). Bacteria may prevent phage adsorption through mutations that modify key surface receptors, but phages may subsequently adapt to utilise novel entry points (Meyer et al., 2012). Bacteria can also produce extracellular proteins that block entry to the cell, but these can be degraded by phages that produce suitable enzymes. Alternatively, restriction-modification mechanisms may be used to break down phage DNA once it is inside the cell or the bacterial CRISPR-Cas system (clustered regularly interspaced palindromic repeats (CRISPRs) and associated *cas* genes) may be used to incorporate phage DNA into the host's genome as a form of adaptive immunity; again, certain phages are able to overcome these forms of resistance (see Labrie et al., 2010 for a comprehensive review of defence and counter-defence mechanisms among bacteria and phages). These reciprocal adaptations are indicative of a coevolutionary arms race between bacteria and phages (Chao et al., 1977; Buckling and Rainey, 2002a; Mizoguchi et al., 2003; Forde et al., 2008; Hall et al., 2011b; Scanlan et al., 2011).

The MA and GFG frameworks are well-studied and much attention has been paid to conditions that promote coevolutionary cycling, given that these dynamics can produce patterns of local adaptation (Gandon and Michalakis, 2002) and are generally more favourable to the RQH for the maintenance of sex (Hamilton, 1980; Parker, 1994; Otto and Nuismer, 2004; Agrawal, 2009; Lively, 2010b; Gokhale et al., 2013). Both frameworks are able to produce coevolutionary cycling (Sasaki, 2000; Agrawal and Lively, 2002), but MA models typically produce oscillatory dynamics with shorter periods and higher amplitudes, and are more likely to select for sexual reproduction (Otto and Nuismer, 2004). This has important consequences for the generality of the RQH given that plant-pathogen infection genetics appear to be governed by GFG-like interactions (Flor, 1956; Thompson and Burdon, 1992; Thrall and Burdon, 2002; Jones and Dangl, 2006).

Several authors have proposed modified or extended versions of the MA and GFG frameworks, which may better represent the genetic specificity of real populations (Frank, 1994; Parker, 1994; Frank, 2000; Agrawal and Lively, 2002, 2003; Otto and Nuismer, 2004; Forde et al., 2008; Fenton et al., 2009, 2012). These include a modification to the GFG framework so that infectivity is always highest on ancestral genotypes (Forde et al., 2008) and inverse forms of both frameworks that shift the burden of adaptation from hosts to parasites (inverse-GFG, IGFG; Fenton et al., 2009) or vice versa (inverse-MA, IMA; Frank, 1994). Others have suggested that genetic specificity may be more complex and that a combination of these frameworks may be more realistic. For example, the MA and GFG models can be considered as two extremes of a continuum, which allows for intermediate forms of specificity (Parker, 1994; Agrawal and Lively, 2002). Alternatively, each locus may conform to fundamentally different forms of infection genetics (Frank, 2000; Agrawal and Lively, 2003; Fenton et al., 2012), which mirrors the multiple lines of defence and counter-defence that have been observed in real systems (Jones and Dangl, 2006; Labrie et al., 2010). Accordingly, models of specificity have been developed that incorporate both MA and GFG loci (Agrawal and Lively, 2003) or both GFG and IGFG loci (Fenton et al., 2012). These 'two-step' models of specificity are able to produce more complex patterns of coevolutionary dynamics than are possible with a single framework.

Clearly, the genetic specificity of host-parasite interactions plays a vital role in shaping evolutionary and coevolutionary dynamics. Hence, understanding how different forms of specificity influence patterns of selection in these systems is a fundamental aspect of current research. However, genetic specificity is not the only factor that has a strong influence on host and parasite evolution. In particular, heterogeneous contact patterns that cause non-random mixing within a population are known to greatly affect the spread of infectious diseases (Rhodes and Anderson, 1996; Keeling et al., 2001; Eames and Keeling, 2002), which will inevitably influence selection

in both hosts and parasites.

1.4 The role of contact patterns in epidemiological and evolutionary dynamics

Accurately modelling the spread of infectious diseases can be challenging, as real populations consist of many autonomous individuals that interact in complex and unpredictable ways. The traditional method for overcoming these issues is to divide populations into different classes according to attributes such as infection status or genotype and to assume that interactions between individuals can be approximated by average interactions between classes. For example, if S individuals in a population are susceptible and I individuals are infectious, then in many cases we can assume that the rate at which disease spreads is roughly proportional to $S \times I$. This 'compartmental' approach based on mean field theory is the foundation for the majority of epidemiological models as it reduces a complex and stochastic many-body problem to a deterministic one-body problem, which is usually much easier to analyse. Behavioural heterogeneities such as assortative mixing or variation in the frequency of sexual contact can be incorporated into compartmental models (Hethcote and Yorke, 1984; Anderson et al., 1986; Gupta et al., 1989), but interactions between individuals are still based on the average effects of their respective classes, which assumes a certain degree of random mixing.

However, in many situations the interactions between some individuals (e.g. relatives) may be much stronger than between others (e.g. strangers), which can have a considerable impact on epidemiological dynamics (Keeling and Eames, 2005). This is especially true when populations are spatially structured, as interactions between individuals may be restricted to local neighbourhoods. Similarly, most individuals tend to have a small number of relatively stable sexual contacts, although popula-

tions can be highly heterogeneous (Liljeros et al., 2001). Consequently, a rich body of theory on population structure and its effects on epidemiology has been developed, much of which has been driven by the need to understand the spread of sexually transmitted infections (STIs) in humans (Gupta et al., 1989; Garnett and Anderson, 1996; Morris and Kretzschmar, 1997; Eames and Keeling, 2002; Potterat et al., 2002). Social or sexual mixing patterns are typically captured through the use of a contact network, which represents a shift away from the population-level approach of compartmental models and allows heterogeneity to be incorporated at the level of individuals. Simulations of disease spread through networks have demonstrated that clustering and heterogeneity in contact number (degree) are often decisive factors for determining the outcome of an epidemic (Keeling and Eames, 2005; Keeling, 2005; Lloyd-Smith et al., 2005). For example, Keeling (2005) showed that random-mixing models overestimate the initial growth rate of an epidemic if real contact patterns are highly clustered. Heterogeneity in degree has been shown to increase the likelihood that parasites will go extinct before causing an epidemic, but if an epidemic does occur then it will tend to spread through the population very quickly (Lloyd-Smith et al., 2005).

While an individual-based approach can produce more realistic dynamics than simple compartmental models, this method is not without its drawbacks, as it can be computationally intensive and lack analytic tractability. In practice, this limits the extent to which different contact patterns and parameter ranges can be explored, so there has been a recent push towards models that approximate spatial and network structure in mathematically rigorous ways. In particular, 'pair approximation', 'neighbour exchange' and 'spatial correlation' models have been developed that offer much of the simplicity and analytic tractability of compartmental models, while still capturing many of the effects of population structure on epidemiology (Keeling, 1999; Eames and Keeling, 2002; Volz and Meyers, 2007; Lion and Boots,

2010). Pair approximations, for example, still compartmentalise the population according to infection status or other attributes, but focus on the rates at which pairs of classes change. Hence, rather than focusing solely on the number of susceptible and infectious individuals, pair approximations also track the number of susceptible-susceptible, susceptible-infectious and infectious-infectious pairs, which matches the epidemiological dynamics of network-based models much more closely than traditional compartmental approaches (Eames and Keeling, 2002).

Creating a realistic portrayal of contact patterns is clearly vital for understanding and predicting epidemiological dynamics, but it is equally important for gaining insights into host and parasite evolution. This has been highlighted in a number of theoretical and experimental studies that have focused on the role of population structure in shaping evolutionary and coevolutionary dynamics. On the theoretical side, spatial structure has been shown to constrain the evolution of transmission rates (Read and Keeling, 2003) and virulence (Boots and Sasaki, 1999; Haraguchi and Sasaki, 2000; Heilmann et al., 2010; Lion and Boots, 2010; Best et al., 2011), but may select for higher resistance in the host (Best et al., 2011) or altruistic suicide (Débarre et al., 2012). In addition, heterogeneous mixing patterns and assortativity have been shown to allow parasites with different life-history strategies to coexist by specialising on distinct subsets of the population (Eames and Keeling, 2006). Methods that approximate population structure are especially useful when modelling the evolutionary or coevolutionary dynamics of hosts and parasites, which may occur over much longer timescales than ecological dynamics. Best et al. (2011) used this approach to show that selection for resistance and virulence depends on the ratio of local to global dispersal of both populations. Meanwhile, studies from the experimental community have demonstrated that spatial structure slows down coevolutionary dynamics (Brockhurst et al., 2003), promotes host-parasite coexistence (Brockhurst et al., 2006) and can lead to local adaptation (Lively and Dybdahl, 2000; Thrall

and Burdon, 2002) even at very fine scales (Vos et al., 2009; Koskella et al., 2011). Environmental heterogeneity and different rates of dispersal have also been shown to influence coevolutionary dynamics and local adaptation (Forde et al., 2004, 2007; Morgan et al., 2005, 2007; Vogwill et al., 2008, 2009; King et al., 2009; Lopez-Pascua et al., 2012), in line with the geographic mosaic theory of coevolution (Gomulkiewicz et al., 2000). Together, these studies emphasise the fundamental role that contact patterns play in mediating the ecological and evolutionary dynamics of host-parasite interactions and highlight the need to consider population structure when constructing theoretical models.

1.5 Thesis outline

The roles of parasite-mediated selection in host evolution and vice versa have been discussed at length in the preceding sections, with a particular focus on the importance of coevolutionary dynamics, ecological feedbacks, genetic specificity and contact patterns. Together, the studies that constitute this thesis cover these four themes in a variety of host-parasite contexts, including the coevolution of broad resistance and infectivity ranges (i.e. generalism; chapters 2-4), the persistence of coevolutionary cycling and the maintenance of sex (chapter 5), the effects of mating behaviour on disease prevalence and evolution (chapters 6 and 7), and the evolution of sexual and social behaviour (chapters 7 and 8). These topics are explored using a range of theoretical models that are analysed algebraically or through the use of computer simulations. The structure of this thesis is described below.

Chapter 2 considers how different forms of genetic specificity affect the evolution of broader infectivity ranges. This study was motivated by recent interest in the importance of reciprocal adaptations for the emergence of novel traits (Poullain et al., 2008; Paterson et al., 2010; Schulte et al., 2010; Hall et al., 2011a; Morran et al.,

2011). The model presented in this chapter demonstrates that the type and strength of epistasis between mutations is crucial for shaping coevolutionary dynamics and that gradual, rather than sudden changes in host phenotype are not always more likely to lead to successful range expansion among parasites. A published version of this chapter is included in the appendices (Ashby et al., 2014a).

Chapter 3 explores how population mixing influences coevolutionary dynamics as the strength of fitness costs associated with range expansion is varied. The model, which is loosely based on the infection genetics and ecological dynamics of bacteria and lytic phages, demonstrates that spatial structure can facilitate range expansion even when this trait is costly. A published version of this chapter is included in the appendices (Ashby et al., 2014b).

Chapters 2 and 3 are based on the gene-for-gene (GFG) framework of host-parasite specificity, but empirical observations suggest that the genetic basis for infectivity among bacteria and lytic phages is more complex. Chapter 4 introduces a new model of specificity, which builds on several previous frameworks and is able to produce coevolutionary dynamics that match empirical observations under a variety of environmental conditions.

Chapter 5 moves away from the realm of bacteria and phages and focuses instead on conditions that allow coevolutionary cycling (fluctuations in the frequencies of genes that govern specificity) to persist. Parasitic castration is found to be necessary for these dynamics to continue indefinitely, which has important implications for the maintenance of sex according to the Red Queen Hypothesis, as castrators are shown to disproportionately harm sexual populations by reducing the availability of fertile mates. A published version of this chapter is available online (Ashby and Gupta, 2014).

Chapter 6 begins by reviewing how heterogeneity in sexual contact patterns influences the epidemiological dynamics of sexually transmitted infections (STIs). The

consequences of different host sexual contact patterns for the evolution of STI virulence are then explored using a model of asymmetrical mating systems. A published version of this chapter is included in the appendices (Ashby and Gupta, 2013).

Chapter 7 continues on the theme of STIs and mating behaviour, but explores these in a coevolutionary context. This chapter focuses on the effects of disease-avoidance behaviour by hosts on the evolution of virulence among sterilising STIs and how ecological feedbacks affect selection for mate choice. The strength of the trade-off between transmissibility and virulence is shown to have a strong influence on coevolutionary dynamics, with fluctuating selection, stable equilibria and runaway virulence all possible outcomes.

The final study in this thesis (chapter 8) concerns the evolution of social mixing patterns when hosts are capable of transmitting both beneficial information (e.g. related to foraging sites or predation risk) and infectious disease. A stochastic individual-based model is presented, which captures fundamental properties of a well-studied empirical system of wild songbirds (*Parus major*), and is used to show that multiple switches in the optimal social strategy can occur as disease incidence increases. The appendices include extended methods and additional results that supplement the material in the main text.

1.6 Statement of author contributions

I designed and analysed the models presented in this thesis and was the primary author of the text. Chapters 2 and 3 were co-written with my supervisors, Sunetra Gupta[2] and Angus Buckling[3]; chapter 4 was co-written with Angus Buckling[3]; chapters 5 and 6 were co-written with Sunetra Gupta[2]; chapter 8 was co-written with Damien Farine[2].

[2]Department of Zoology, University of Oxford, Oxford, UK
[3]Biosciences, University of Exeter, Cornwall Campus, Penryn, UK

Chapter 2

Effects of epistasis on infectivity range during host-parasite coevolution

Ben Ashby[1], Sunetra Gupta[1] and Angus Buckling[2]

Abstract

Understanding how parasites adapt to changes in host resistance is crucial to evolutionary epidemiology. Experimental studies have demonstrated that parasites are more capable of adapting to gradual, rather than sudden changes in host phenotype, as the latter may require multiple mutations that are unlikely to arise simultaneously. A key, but as yet unexplored factor is precisely how interactions between mutations (epistasis) affect parasite evolution. Here, we investigate this phenomenon in the context of infectivity range, where parasites may experience selection to infect broader sets of genotypes. When epistasis is strongly positive, we find that parasites are unlikely to evolve broader infectivity ranges if hosts exhibit sudden, rather than gradual changes in phenotype, in close agreement with empirical observations. This is due to a low probability of fixing multiple mutations that individually confer no immediate advantage. When epistasis is weaker, parasites are more likely to evolve broader in-

[1]Department of Zoology, University of Oxford, Oxford, UK
[2]Biosciences, University of Exeter, Cornwall Campus, Penryn, UK

fectivity ranges if hosts make sudden changes in phenotype, which can be explained by a balance between mutation supply and selection. Thus, we demonstrate that both the rate of phenotypic change in hosts and the form of epistasis between mutations in parasites are crucial in shaping the evolution of infectivity range.

2.1 Introduction

Antagonistic coevolution between hosts and parasites can lead to directional selection for more effective defence and counter-defence mechanisms (Thrall and Burdon, 2003; Labrie et al., 2010; Schulte et al., 2010; Brown and Tellier, 2011). In many cases, these dynamics (often referred to as 'coevolutionary arms races') are characterised by reciprocal expansions in the range of genotypes that the host can resist and the parasite can infect, which means that populations tend to fare better than their ancestors when confronted with contemporary antagonists (Buckling and Rainey, 2002a; Mizoguchi et al., 2003; Thrall and Burdon, 2003; Brown and Tellier, 2011; Scanlan et al., 2011). Understanding precisely why some parasites develop broader infectivity ranges than others has important implications for our ability to predict how parasites will evolve in response to shifting patterns of host resistance or other environmental changes, with particular relevance for the use of biocontrol in industry and medicine (Tait et al., 2002; Levin and Bull, 2004). While variation in infectivity range is typically explained by selection (e.g. fitness costs; Fenton and Brockhurst, 2007; Ashby et al., 2014b) or fundamental genetic constraints (e.g. parasites may be forced to specialise on one group of hosts or another; Dybdahl and Lively, 1998; Decaestecker et al., 2007; Koskella and Lively, 2007), a lack of broad infectivity ranges may also result from the need to fix multiple, rather than single, mutations (Benmayor et al., 2009; Paterson et al., 2010; Hall et al., 2011a; Meyer et al., 2012). Here, we investigate how key parameters (epistasis and the rate of phenotypic change in the

host) affect the fixation of multiple mutations, and hence infectivity range, during coevolution.

Parasites frequently require multiple amino acid substitutions to infect a novel host, and the likelihood of several beneficial mutations occurring simultaneously or in quick succession is usually slim (Benmayor et al., 2009; Hall et al., 2011a; Scanlan et al., 2011; Gururani et al., 2012; Meyer et al., 2012; Russell et al., 2012). In some cases, however, subsets of mutations may confer an immediate fitness advantage on contemporaneous hosts, increasing the probability that a complete set will eventually become fixed (Meyer et al., 2012). Empirical observations using bacteria and viruses suggest that these conditions are most likely to be realised when parasites are exposed to genetically diverse host populations, such that they experience gradual, rather than sudden changes in phenotype during coevolution, as individual mutations may increase performance on subsets of the host population (Hall et al., 2011a; Meyer et al., 2012). Crucially, a wide range of genetic and ecological processes, such as recombination and gene flow, could alter the rate of phenotypic change in the host population (Sasaki, 2000; Gandon, 2002; Gandon and Nuismer, 2009), and hence the likelihood of range expansion.

Empirical studies that demonstrate the importance of coevolution for the emergence of broad infectivity ranges have used host-parasite interactions that are governed by strong positive epistasis between infectivity mutations (Paterson et al., 2010; Hall et al., 2011a; Meyer et al., 2012). This means that parasites with an incomplete set of mutations will fare no better (or even worse) on the novel host than parasites with none. However, both quantitatively and qualitatively different forms of epistasis governing infectivity have been identified, including weak positive, negative and no epistasis (Lenski, 1984b; Wilfert and Schmid-Hempel, 2008), and it is currently unclear how this will impact on infectivity range during coevolution. Here, we demonstrate theoretically that different forms of epistasis have contrasting effects on

the ability of parasites to expand their infectivity ranges when hosts exhibit gradual or sudden changes in phenotype during coevolution. Our results are in good agreement with empirical observations when epistasis is strongly positive (gradual changes in host phenotype promote broader infectivity ranges), but notably, we find that the opposite outcome is expected for weaker forms of epistasis (sudden changes in host phenotype promote broader infectivity ranges).

2.2 Methods

2.2.1 Model description

We compare two types of genetic specificity that govern host-parasite interactions, both of which allow the evolution of parasites with broader infectivity ranges. The first is similar to the multilocus gene-for-gene framework proposed by Sasaki (2000), with interactions occurring at n biallelic loci in both host and parasite. Increasing the number of infectivity alleles improves infectivity to a wider range of host genotypes, and increasing the number of resistance alleles improves resistance to a wider range of parasite genotypes. We refer to this as a symmetric (SYM) interaction, because there is a one-to-one correspondence between resistance and infectivity alleles. We compare this scenario to an asymmetric (ASYM) form of genetic specificity, where interactions occur between a single locus in the host and n loci in the parasite (one-to-many). In this case, there are only two possible host genotypes (susceptible and resistant), and increasing the number of infectivity alleles can improve performance on the resistant host. Genotypes are represented by binary strings (hosts: $h_1^i...h_n^i$ (SYM) or h_1^i (ASYM); parasites: $p_1^j...p_n^j$; superscripts identify each genotype), where each locus corresponds to the presence (1) or absence (0) of a resistance (host) or infectivity (parasite) allele. Infectivity alleles interact with each other and with resistance alleles to modulate the overall strength of infectivity, Q, on a given host, such that:

SYM

(a)	P_{000}	P_{001}	P_{010}	P_{100}	P_{011}	P_{101}	P_{110}	P_{111}
H_{000}	1	1	1	1	1	1	1	1
H_{001}	2/3	1	2/3	2/3	1	1	2/3	1
H_{010}	2/3	2/3	1	2/3	1	2/3	1	1
H_{100}	2/3	2/3	2/3	1	2/3	1	1	1
H_{011}	1/3	2/3	2/3	1/3	1	2/3	2/3	1
H_{101}	1/3	2/3	1/3	2/3	2/3	1	2/3	1
H_{110}	1/3	1/3	2/3	2/3	2/3	2/3	1	1
H_{111}	0	1/3	1/3	1/3	2/3	2/3	2/3	1

ASYM

(b)	P_{000}	P_{001}	P_{010}	P_{100}	P_{011}	P_{101}	P_{110}	P_{111}
H_0	1	1	1	1	1	1	1	1
H_1	0	1/3	1/3	1/3	2/3	2/3	2/3	1

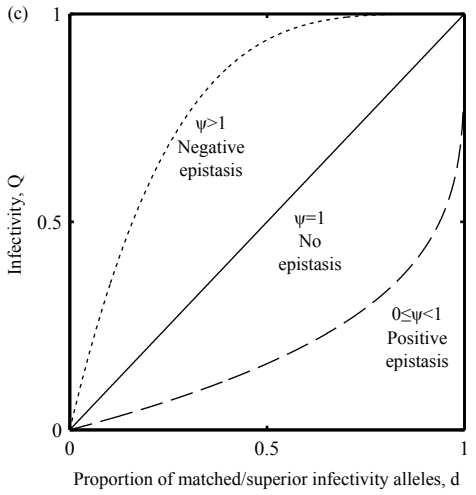

Figure 2.1: Genetic interactions between hosts and parasites. The tables show the proportion of parasite loci, d, that match or are superior to (a) corresponding loci (symmetric (SYM) scenario) or (b) the sole locus (asymmetric (ASYM) scenario) in the host. Interactions between hosts (H) and parasites (P) are shown for $n = 3$, where subscripts correspond to the presence (1) or absence (0) of a resistance or infectivity allele at a given locus. (c) Infectivity (Q; equation 2.1) as a function of the proportion of matched/superior infectivity alleles (d) in the parasite for different values of the epistasis parameter, ψ, which modifies the type and strength of epistasis between infectivity alleles: $0 \leq \psi < 1$, $\psi > 1$ and $\psi = 1$ give positive (dashed), negative (dotted) and no (solid) epistasis, respectively.

$$Q_{ij} = \begin{cases} 1 - (1 - d_{ij})^{\psi} & \text{if } d_{ij} < 1 \\ 1 & \text{otherwise} \end{cases} \tag{2.1}$$

where d_{ij} is the proportion of infectivity alleles that match or exceed either (i) the resistance allele at each corresponding locus (SYM: $d_{ij} = 1 - \frac{1}{n}\sum_{k=1}^{n} h_k^i(1 - p_k^j)$), or (ii) the sole resistance allele in the host (ASYM: $d_{ij} = 1 - \frac{h_1^i}{n}\sum_{k=1}^{n}(1 - p_k^j)$). The parameter ψ modulates the type and strength of epistasis between infectivity alleles, such that $0 \leq \psi < 1$, $\psi > 1$ and $\psi = 1$ give positive, negative and no epistasis, respectively (figure 2.1). Values of ψ further away from 1 give stronger forms of epistasis; in the special case of $\psi = 0$ infection is only possible if $p_k^j \geq h_k^i$ at all loci.

We base the epidemiological dynamics of our model on the SI framework, where hosts of genotype i are classed as either susceptible (S_i) or infected by parasite genotype j (I_{ij}). Hosts are haploid and reproduce asexually with a maximum per-capita birth rate of $\bar{r}$, and experience a density-dependent per-capita mortality rate of at least $\bar{\mu}N$, where $\bar{\mu}$ modifies the strength of competition for resources and N is the total population size. We set $\bar{\mu} = \bar{r}/K$, so that the host population tends towards a carrying capacity of K in the absence of disease. Initial populations are composed of K susceptible and $K/20$ infected hosts, with no resistance or infectivity alleles present. The host population mixes randomly and exhibits either frequency- (FD) or density-dependent (DD) contact patterns, so that a susceptible host of genotype i will be infected with parasite j at a rate of $\lambda_{ij} = \beta_j Q_{ij} \sum_k I_{kj}/N$ (FD) or $\lambda_{ij} = \beta_j Q_{ij} \sum_k I_{kj}$ (DD) per unit time, where β_j is the transmission coefficient of the parasite (base transmission rate: $\bar{\beta}$). Infected hosts are unable to recover and suffer an increased mortality rate, given by the parameter α_j (base disease-associated mortality rate: $\bar{\alpha}$); coinfection does not occur. New generations are subject to mutation rates of ϵ_H and ϵ_P at each locus for hosts and parasites, respectively, with the restriction that multiple mutations cannot arise simultaneously (i.e. the genotypes of parent and progeny never differ at more than one locus).

Broader resistance and infectivity ranges are often associated with a fitness cost (Chao et al., 1977; Webster and Woolhouse, 1999; Bohannan et al., 2002; Poullain et al., 2008), which we incorporate into either the per-capita birth (r_i) or the coefficient of density-dependent mortality (μ_i), and either the disease-associated mortality rate (α_j) or transmission coefficient (β_j) rate for parasites. We limit simulations to one type of fitness cost per population, giving a total of four combinations. For example, if hosts experience a fitness cost in the form of a reduced birth rate, then the coefficient of density-dependent mortality remains constant for all genotypes ($\mu_i = \bar{\mu}$) and if parasites with broader infectivity ranges have a lower transmission coefficient, then

the disease-associated mortality rate does not vary ($\alpha_j = \bar{\alpha}$). When fitness costs do affect a particular life-history trait, they do so based on the following equations:

$$\left.\begin{aligned} r_i &= \bar{r}\left(1 + \left(\frac{1}{c_H} - 1\right) q_i^{\phi_H}\right) \\ \mu_i &= \bar{\mu}\left(1 + (c_H - 1)\, q_i^{\phi_H}\right) \\ \alpha_j &= \bar{\alpha}\left(1 + (c_P - 1)\, q_j^{\phi_P}\right) \\ \beta_j &= \bar{\beta}\left(1 + \left(\frac{1}{c_P} - 1\right) q_j^{\phi_P}\right) \end{aligned}\right\} \tag{2.2}$$

where for hosts (parasites), q_i (q_j) gives the proportion of loci that contain a resistance (infectivity) allele, $c_H \geq 1$ ($c_P \geq 1$) is the maximum strength of the fitness cost and ϕ_H (ϕ_P) controls whether costs are accelerating ($\phi_H, \phi_P > 1$), decelerating ($0 < \phi_H, \phi_P < 1$) or linear ($\phi_H, \phi_P = 1$). Note that for $q_i = 1$, $r_i = \bar{r}/c_H$ or $\mu = \bar{\mu} c_H$, which means that the birth rate or coefficient of density-dependent mortality is c_H times lower/higher for individuals with a full complement of resistance alleles than the base rates (similarly for $q_j = 1$: $\beta_j = \bar{\beta}/c_P$ and $\alpha_j = \bar{\alpha} c_P$). Although resistance and infectivity alleles can behave epistatically for both specificity (Q) and fitness costs (see e.g. Fenton and Brockhurst, 2007), we shall only refer to epistasis in the context of the parasite's ability to infect a given host (i.e. in terms of the parameter ψ) to avoid confusion. Variations in ψ will be referred to as positive, negative or no epistasis, whereas variations in ϕ_H and ϕ_P will be referred to accelerating, decelerating or linear fitness costs.

The dynamics of our model (excluding mutations) are captured by the following set of coupled ordinary differential equations:

$$\begin{aligned} \frac{dS_i}{dt} &= S_i\left(r_i - \mu_i N - \sum_j \lambda_{ij}\right) \\ \frac{dI_{ij}}{dt} &= \lambda_{ij} S_i - (\mu_i N + \alpha_j) I_{ij} \end{aligned} \tag{2.3}$$

We translate this deterministic framework into a stochastic model by using the τ-leap

method proposed by Gillespie (2001), which uses a fixed step size, τ, and assumes the number of events occurring within a time step is Poisson distributed. The optimal genotype will always emerge in a deterministic framework with no extinction threshold, but we should expect parasites to struggle to accumulate infectivity alleles when demographic stochasticity is included, especially if resistance spreads rapidly. Thus, by comparing the deterministic and stochastic models, we are able to establish if broader infectivity ranges do not evolve due to selection (i.e. broader infectivity ranges are not beneficial), or if mutations are struggling to reach fixation due to stochasticity.

2.2.2 Analysis

We analyse the deterministic and stochastic versions of our model to evaluate how the previously described forms of genetic specificity (SYM and ASYM) and different types of epistasis (ψ) influence the evolution of broader infectivity ranges. In other words, we establish how these genetic factors affect the ability of parasites to accumulate infectivity alleles.

At each time point, we measure the average proportion of parasite loci that contain an infectivity allele and define 'peak infectivity range', E, to be the maximum of this value over the course of a simulation ($20,000$ time units). Thus, if $x_k(t)$ is the proportion of parasites that have a total of k infectivity alleles at time t, then:

$$E = \max_t \left(\frac{1}{n} \sum_{k=1}^{n} k x_k(t) \right) \tag{2.4}$$

We measure the maximum value over the duration of each simulation as GFG frameworks can produce fluctuations in range (e.g. Sasaki, 2000), but the focus of the present study is whether genetic factors affect the initial emergence of broader infectivity ranges and not whether they are evolutionarily stable. We wish to determine

	description	range
$\bar{\alpha}$	base disease-associated mortality rate	$(0.01, 0.1)$
$\bar{\beta}$	base transmission coefficient	FD: $(0.1, 1)$, DD: $(10^{-9}, 10^{-5})^*$
ϵ_H	host mutation rate	$1/K^{**}$
ϵ_P	parasite mutation rate	$1/K^{**}$
$\bar{\mu}$	base coefficient of density-dependent mortality	$\bar{r}/K^{**}$
c_H	maximum strength of host fitness costs	$(1.05, 1.2)$
c_P	maximum strength of parasite fitness costs	$(1.05, 1.5)$
K	carrying capacity	$(10^5, 10^9)$
$\bar{r}$	base per-capita birth rate	$(0.01, 0.1)$

Table 2.1: Parameter distributions used for the Latin Hypercube Sample. * FD=frequency-dependent transmission; DD=density-dependent transmission. ** Values are fixed by the base per-capita birth rate and/or carrying capacity.

the general behaviour of our model, but since we are not modelling a particular host-parasite system, the parameter space is somewhat arbitrary. To overcome this issue, we fix $\tau = 0.1$ and vary ψ, ϕ_H and ϕ_P incrementally to cover all qualitatively different forms of epistasis (positive, negative or none) and fitness costs (accelerating, decelerating or linear) and use a Latin hypercube sample (LHS) to draw the remaining parameters from the distributions in table 2.1, the majority of which are varied over at least an order of magnitude, covering biologically plausible areas of parameter space (e.g. population sizes of $10^5 - 10^9$ are appropriate for microbial communities). Note that the mutation and natural mortality rates are not used in the construction of the LHS, but are instead fixed by the base per-capita birth rate and/or carrying capacity, which are part of the LHS design. The parameters c_H and c_P vary over relatively narrow ranges compared to the other parameters as previous studies have demonstrated that high fitness costs greatly limit range expansion in well-mixed populations (Sasaki, 2000; Ashby et al., 2014b). The LHS contains 1,000 parameter

combinations, each of which is tested in both SYM and ASYM scenarios with all four possible combinations of fitness costs (equation 2.2) for different values of ψ, ϕ_H and ϕ_P. This method allows us to determine if one form of genetic specificity (SYM or ASYM) consistently allows broader infectivity ranges to evolve than the other, and if this relationship holds for all forms of epistasis (ψ) and fitness costs (ϕ_H and ϕ_P). We discard simulations where the parasite dies out in either the SYM or ASYM scenarios for a given set of parameters.

2.3 Results

For the sake of brevity, here we only present results for parasite populations with three loci ($n = 3$), contact patterns based on frequency-dependence (FD) and decelerating fitness costs ($\phi_H = \phi_P = 0.5$) affecting only the per-capita birth rate (r_i) and transmission coefficient (β_j). However, the results are qualitatively similar for other numbers of loci (figure A.1), density-dependent contact patterns (figure A.2), linear and accelerating fitness costs (figure A.3) and different combinations of cost functions (figure A.4).

2.3.1 Algebraic analysis

Using equation 2.3, we can derive the basic reproductive ratio, R_0^j, for a parasite in a naive, fully-susceptible host population when transmission is frequency-dependent:

$$R_0^j = \frac{\beta_j}{\bar{\mu}N + \bar{\alpha}} \tag{2.5}$$

In the absence of resistant hosts, this quantity is maximised when the parasite has no infectivity alleles. We can also derive the effective basic reproductive ratio, R_{EFF}^j, which is the average number of secondary infections produced for any composition of

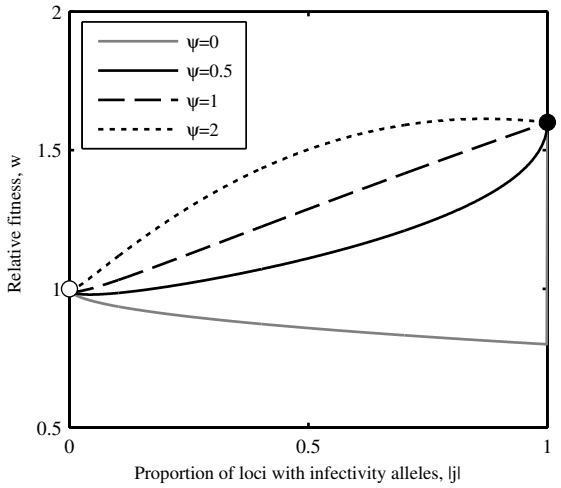

Figure 2.2: Relative fitness (w; equation 2.7) of parasites in a fixed host population for different values of the epistasis parameter, ψ. The white circle indicates the fitness of the wild type parasite (no infectivity alleles) and the black circle indicates the fitness of parasites with a full complement of infectivity alleles. When epistasis is strongly positive (grey), parasites with a complete set of infectivity alleles may have the highest fitness, but intermediate genotypes could perform worse than the wild type, so parasites may struggle to accumulate infectivity alleles (note the step-change when $q_j = 1$ for $\psi = 0$). When epistasis is weaker (solid black) or non-existent (dashed), parasites with an incomplete set of infectivity alleles are likely to experience an immediate increase in fitness, which may allow range expansion to occur. When epistasis is negative (dotted), an incomplete set of mutations may be optimal due to the presence of fitness costs, which outweigh the benefits of broader infectivity ranges. Parameters: $\phi_P = 0.5$; $c_P = 1.25$.

hosts:

$$R^j_{EFF} = \frac{\beta_j H_j}{\bar{\mu} N + \bar{\alpha}} \tag{2.6}$$

where $H_j = \sum_i Q_{ij} S_i / N$ is the proportion of the host population that the parasite is able to infect. For a given composition of hosts, parasite j should initially perform better than the wild type (no infectivity alleles, H_0) provided $H_0 = 0$ and $H_j > 0$, or, for $H_0 > 0$:

$$w_j = \frac{R^j_{EFF}}{R^0_{EFF}} = \frac{\left(1 + \left(\frac{1}{c_P} - 1\right) q_j^{\phi_P}\right) H_j}{H_0} > 1 \tag{2.7}$$

Figure 2.2 shows several examples of how this fitness function varies with the number of infectivity alleles in the parasite for different values of the epistasis parameter, ψ. While figure 2.2 represents an idealised scenario where the host population is held constant (50% wild type, 50% maximal resistance), it does reveal some interesting patterns. In particular, parasites with an intermediate number of infectivity alleles may perform worse than the wild type if epistasis is positive. This suggests that broad infectivity ranges are unlikely to emerge if hosts make sudden jumps in phenotype. Clearly, the composition of the host population will alter selection among parasites and vice versa, which prevents further algebraic analysis of this system. In addition, real populations are subject to stochasticity, which can have a considerable influence on the accumulation of rare mutations. However, based on the above analysis, we can make two predictions to test numerically: (i) selection for parasites with broad infectivity ranges should peak for low values of ψ, as complete or near complete sets of mutations are required to overcome host resistance; (ii) parasites in the ASYM scenario will struggle to accumulate mutations when demographic stochasticity is included if epistasis is strongly positive, as intermediate genotypes may perform worse than the wild type.

2.3.2 Numerical analysis

2.3.2.1 Strong positive epistasis selects for parasites with broad ranges

Numerical analysis of the deterministic model revealed that infectivity range peaks when epistasis is positive ($\psi < 1$), but declines rapidly if epistasis is negative ($\psi > 1$; figure 2.3). The lack of an extinction threshold in the deterministic model means that the optimal genotype is always able to emerge, so there are no qualitative differences between the SYM and ASYM scenarios.

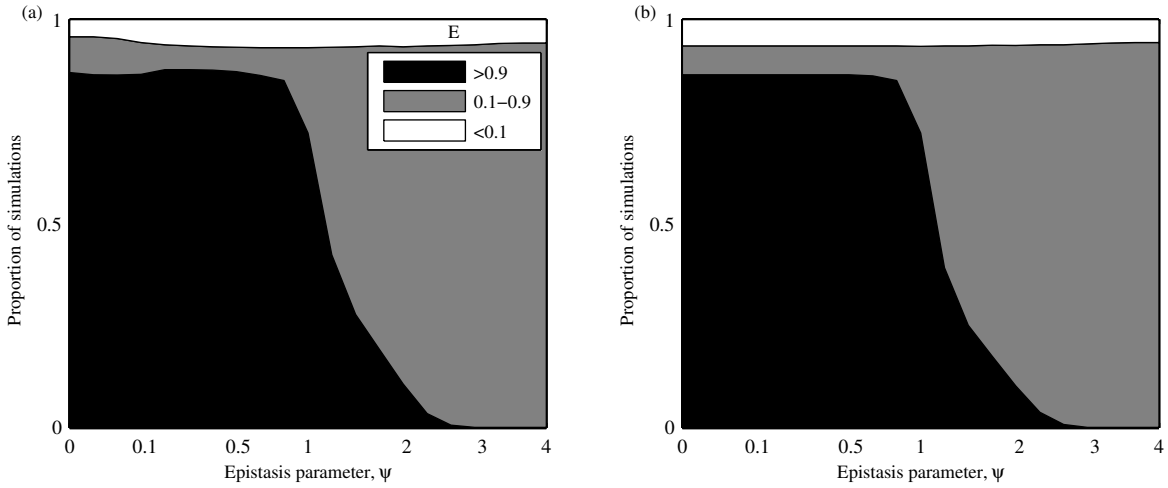

Figure 2.3: Proportion of deterministic simulations where peak infectivity range (E; equation 2.4) was greater than 0.9 (black), less than 0.1 (white) or between these values (grey), for (a) gradual (SYM) and (b) sudden (ASYM) changes in host phenotype. The parameter ψ controls the type and strength of epistasis between infectivity alleles, ranging from strong positive ($\psi << 1$), through weak positive ($\psi < 1$), none ($\psi = 1$) and finally, negative ($\psi > 1$) epistasis. The fittest genotype always emerges in a deterministic framework with no extinction threshold, so the SYM and ASYM scenarios produce almost identical outputs.

2.3.2.2 Stochasticity constrains infectivity range when hosts exhibit sudden changes in phenotype

The pattern of parasite evolution in the stochastic SYM scenario was broadly similar to that described for the deterministic model: infectivity ranges peaked for strong positive epistasis ($\psi << 1$) and decreased with greater ψ (figure 2.4a). Yet, unlike the deterministic version of the SYM scenario, the stochastic version exhibited very little, if any, selection for broader infectivity ranges for negative epistasis ($\psi > 1$). This is due to reduced selection for resistance in the host (figure A.5c). When epistasis between infectivity mutations is negative, resistance is largely ineffective as hosts require multiple mutations to achieve a significant reduction in susceptibility. Hence, hosts do not tend to evolve broader resistance ranges when demographic stochasticity

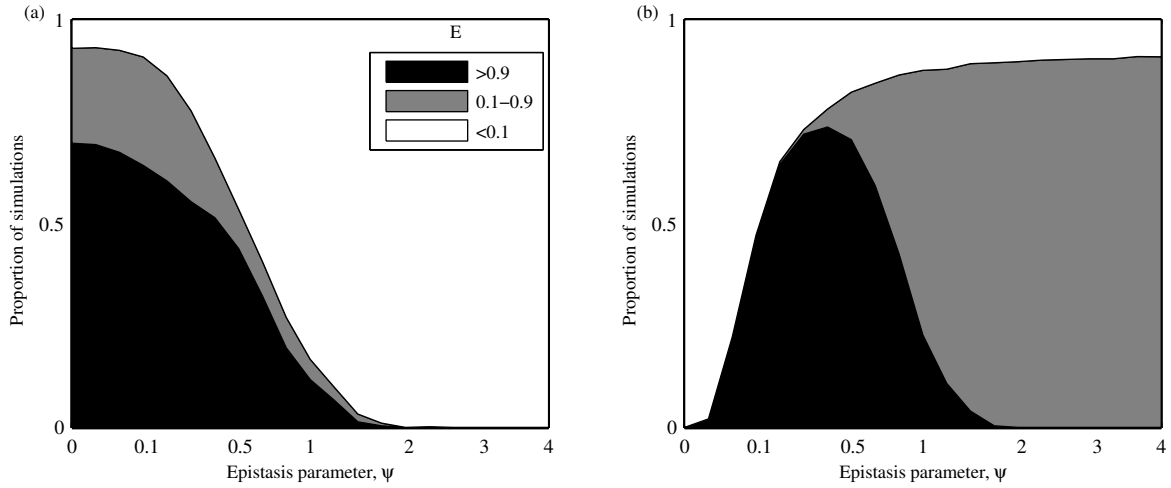

Figure 2.4: Proportion of stochastic simulations where peak infectivity range (E; equation 2.4) was greater than 0.9 (black), less than 0.1 (white) or between these values (grey), for (a) gradual (SYM) and (b) sudden (ASYM) changes in host phenotype. The parameter ψ controls the type and strength of epistasis between infectivity alleles, ranging from strong positive ($\psi << 1$), through weak positive ($\psi < 1$), none ($\psi = 1$) and finally, negative ($\psi > 1$) epistasis. The patterns in the SYM scenario are broadly similar to those in the deterministic version of the model (figure 2.3a), but infectivity range is predicted to peak for weak positive epistasis in the ASYM scenario. This disparity can be explained by the low probability of fixing multiple mutations in the presence of strong positive epistasis.

is included and consequently parasites do not need to evolve broader infectivity ranges under these conditions.

In contrast, the stochastic version of the ASYM scenario produced markedly different outcomes to both the deterministic model and the SYM scenario (figure 2.4b). Specifically, broad infectivity ranges were extremely rare for strong positive epistasis and instead peaked for weaker interactions (intermediate values of ψ). This pattern was consistent for different types of fitness cost, variations in the rate at which fitness costs increased (i.e. accelerating, decelerating or linear), different numbers of loci and density-dependent transmission (see appendix A).

We extended the duration of 10% of our stochastic simulations to $200,000$ time

steps (a tenfold increase) to ensure that the differences between the two versions of the ASYM scenario were not attributable to faster evolution in the deterministic setting. However, longer simulations did not change the overall pattern of our results and only led to relatively minor quantitative differences (average change in $E : 0.03$). Notably, parasites that experienced very strong positive epistasis ($\psi = 0$) were still unable to accumulate infectivity alleles, even over this longer time period.

The disparity between the deterministic and stochastic versions in the ASYM scenario can be explained by the low probability of fixing multiple mutations that were characterised by strong positive epistasis (i.e. parasites were trapped at a local fitness peak). Thus, although a complete set of infectivity alleles may have been optimal for low values of ψ (figures 2.2, 2.3b), parasites struggled to accumulate mutations that were not immediately beneficial. This situation did not occur in the SYM scenario, as individual mutations conferred an immediate increase in fitness due to the presence of genetically intermediate hosts (hosts exhibited gradual changes in phenotype). Figure 2.5 shows contrasting dynamics from the two scenarios. When hosts exhibit gradual changes in phenotype (SYM; figure 2.5a), the fitness of parasites with incomplete sets of infectivity alleles is greater than the wild type, which allows individual mutations to become fixed. When hosts exhibit sudden changes in phenotype (ASYM; figure 2.5b), parasites with the broadest ranges still have the highest fitness, but an incomplete set of infectivity alleles is costly, so mutations are unlikely to accumulate when demographic stochasticity is included.

Although broad infectivity ranges were much more common in the stochastic version of the SYM scenario than in the corresponding ASYM scenario for strong positive epistasis ($\psi << 1$), the converse was true for weak positive epistasis ($0.1 < \psi < 1$). This pattern can again be explained by less effective resistance in the host as epistasis between infectivity alleles weakens (figure A.5c). Gradual changes in phenotype are less advantageous to the host as ψincreases, reducing the likelihood that resistance

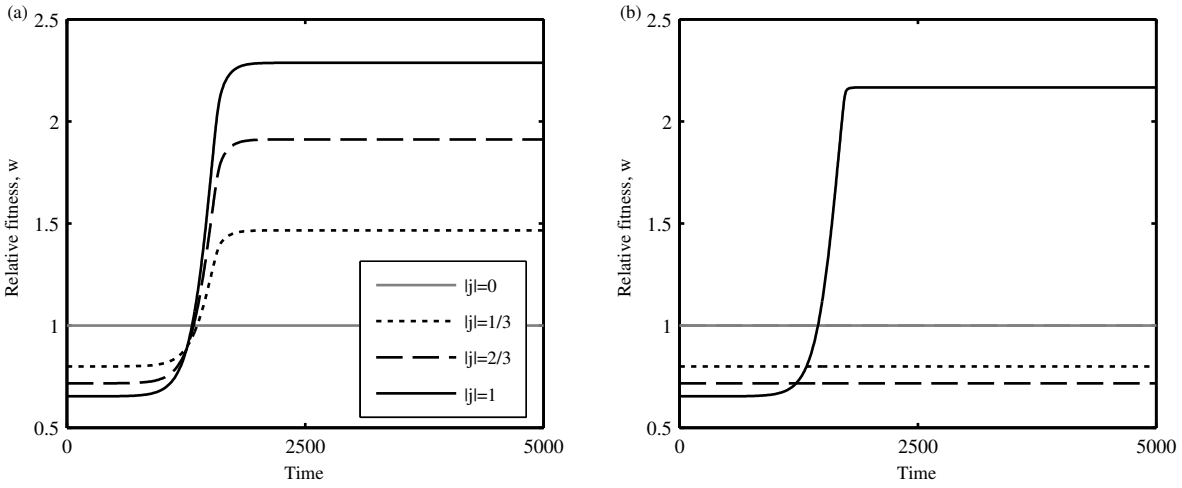

Figure 2.5: Example dynamics showing the relative fitness (w; equation 2.7) of parasites with different proportions of loci containing infectivity alleles ($q_j = 0$ (grey), 1/3 (dotted), 2/3 (dashed) and 1 (solid black)) for (a) gradual (SYM) and (b) sudden (ASYM) changes in host phenotype (deterministic version). (a) When resistance begins to spread in a host population that exhibits gradual changes in phenotype (around 1000 time units) parasites with incomplete sets of infectivity alleles experience an immediate increase in fitness, which allows these mutations to be fixed, leading to the evolution of broad infectivity ranges. (b) If the host makes a sudden change in phenotype, then parasites with incomplete sets of infectivity alleles may not experience an immediate increase in fitness. Hence, infectivity alleles may not accumulate in a stochastic environment, even though parasites with a full complement of infectivity alleles have the highest fitness. Parameters: $\alpha = 0.01$, $\beta = 0.25$, $\epsilon_H = 10^{-6}$, $\epsilon_P = 10^{-6}$, $\phi_H = 0.5$; $\phi_P = 0.5$, $\psi = 0$, $c_H = 1$, $c_P = 1.25$, $K = 10^6$, $n = 3$ and $r = 0.05$.

alleles will become fixed in a stochastic setting, which in turn reduces selection for parasites with broader infectivity ranges. Thus, for weak positive epistasis, overall levels of resistance will tend to be lower if hosts are restricted to gradual (SYM) rather than sudden (ASYM) changes in phenotype (SYM).

2.4 Discussion

Our study was inspired by recent experiments where gradual changes in host phenotype provided optimal conditions for parasite evolution, presumably by facilitating the accumulation of multiple mutations in quick succession, whereas sudden changes in host phenotype prevented mutations from being fixed (Benmayor et al., 2009; Paterson et al., 2010; Hall et al., 2011a; Meyer et al., 2012). However, we hypothesised that the presence of strong positive epistasis between mutations was likely to be a key factor in these experiments and that alternative forms of epistasis could lead to different evolutionary outcomes. Using a theoretical approach, we explored how the rate of phenotypic change in the host (gradual versus sudden) and the type and strength of epistasis shape the evolution of broader infectivity ranges. Our findings support empirical observations that gradual changes in host phenotype promote broad infectivity ranges, provided epistasis is strongly positive (Paterson et al., 2010; Hall et al., 2011a; Meyer et al., 2012). Moreover, by comparing deterministic and stochastic models, we have shown why sudden changes in host phenotype can restrict parasite evolution (low probabilities of fixing individual mutations that confer no immediate increase in fitness). However, we have also shown that this prediction does not hold for weak positive epistasis: parasites are more likely to evolve broad infectivity ranges if hosts exhibit sudden changes in phenotype. These results demonstrate that the nature of epistasis can be crucial for shaping parasite evolution and that gradual changes in host phenotype may not always provide the optimal conditions for broad infectivity ranges to evolve.

When epistasis is strongly positive, parasites require a complete (or near-complete) set of mutations to overcome resistance. These are difficult to accumulate if the host makes sudden changes in phenotype (e.g. due to the loss of a key receptor) as individual mutations may carry an intrinsic fitness cost without conferring any benefits with

regards to increased infectivity. If, however, hosts experience gradual changes in phenotype (e.g. reduced expression of a key receptor), then parasites may have access to hosts that are phenotypically intermediate between ancestral and future populations. Individual mutations may then confer an immediate fitness advantage, dramatically increasing the probability that multiple mutations will become fixed. Hence the symmetric (SYM) scenario in our model, which featured intermediate hosts, was much more favourable to the emergence of broad infectivity ranges under strong positive epistasis than the asymmetric (ASYM) scenario, which did not allow intermediate hosts to evolve. Conversely, when epistasis is negative there is a diminishing benefit associated with the acquisition of multiple infectivity alleles regardless of the potential presence of intermediate hosts. In other words, parasites can infect a reasonably broad set of hosts with a single mutation (e.g. Jonah et al., 2003; Brault et al., 2004) and the advantages of a full complement of mutations are outweighed by associated fitness costs. Between these extremes (i.e. for weak positive epistasis), individual infectivity alleles may confer a slight increase in fitness, allowing parasites to accumulate successive mutations. These mutations are likely to be under strong selection when hosts exhibit sudden changes in phenotype, but will be less beneficial if hosts evolve more gradually. Hence, infectivity range peaks for weak positive epistasis in our ASYM scenario and is more common than in the SYM scenario under these conditions.

The results for the ASYM scenario can also be interpreted in terms of a balance between mutation supply and selection for broader infectivity ranges. If parasites perform poorly on current hosts, then selection for broader infectivity ranges is strong, but the mutation supply is constrained due to a lack of suitable hosts. This problem is accentuated when epistasis between infectivity alleles is strongly positive, as multiple mutations that may be costly in isolation are required before parasites can infect large numbers of hosts. Hence, strong positive epistasis results in a low mutation supply

and broader infectivity ranges are unlikely to evolve. Conversely, if parasites perform well on the current host population, then the mutation supply is much greater, but selection for broader infectivity ranges is inevitably weaker. When epistasis is negative, parasites with few infectivity alleles can perform reasonably well on hosts that have invested in resistance; accumulating further (costly) infectivity alleles is unlikely to be beneficial under these circumstances, so again, broader infectivity ranges do not evolve. Between these extremes, parasites can infect sufficient hosts to maintain a modest supply of mutations while experiencing fairly strong selection for broader infectivity ranges. Hence, broader infectivity ranges are most likely to evolve in the stochastic ASYM scenario for intermediate values of ψ, which corresponds to weak positive epistasis.

The discrepancies between the deterministic and stochastic versions of the ASYM scenario are striking, but were not attributable to different rates of evolution, as simulations that were allowed to run for much longer time periods produced very similar results (although over infinitely long time scales or with simultaneous mutations, the stochastic model should eventually produce pathogens with complete sets of infectivity alleles when epistasis is strongly positive). If hosts make sudden jumps in phenotype, then parasites that experience stochasticity and strong positive epistasis between infectivity alleles are likely to get stuck at a local fitness peak. In the deterministic model, parasites are able to explore the entire fitness landscape, which always allows the globally-optimal phenotype to emerge. The contrasting outcomes in the deterministic and stochastic models highlight the need to compare different modelling approaches when studying host-parasite coevolution, as optimal phenotypes may not always emerge when demographic stochasticity is included (see also Ashby et al., 2014b). It is important to note that our approach differs from most other theoretical studies of host-parasite coevolution, which typically omit ecological feedbacks and stochasticity, assume that population sizes are infinite and focus only

on changes in gene frequencies (Sasaki, 2000; Agrawal and Lively, 2002, 2003; Fenton and Brockhurst, 2007; Fenton et al., 2009, 2012)). As a consequence, these studies do not explicitly model interactions between mutations, so it is not clear how ecological processes (and stochasticity) affect how they accumulate.

Our results were robust to changes in a number of modelling assumptions, which indicates that our findings are likely to be quite general. Still, it is somewhat surprising that different forms of fitness costs (accelerating, decelerating or linear) did not produce markedly different results, as different trade-off shapes are often associated with contrasting outcomes in studies of host and parasite evolution (Kisdi, 2006; Best et al., 2010b). This suggests that epistasis plays a dominant role in shaping host-parasite coevolution in our model.

The findings presented herein are consistent with recent empirical work showing that infectivity evolution tends to proceed faster in the presence of coevolving antagonists that exhibit gradual changes in phenotype (Poullain et al., 2008; Paterson et al., 2010; Schulte et al., 2010; Hall et al., 2011a; Morran et al., 2011; Zhang and Buckling, 2011). The host-parasite systems in these experiments appear to feature multiple reciprocal genetic adaptations, comparable to the symmetric (SYM) scenario in our model. Furthermore, some studies have focused specifically on the evolution of broad infectivity ranges, with parasites experiencing either coevolving or constant mixtures of host genotypes (Benmayor et al., 2009; Hall et al., 2011a). The latter treatment is analogous to our asymmetric (ASYM) scenario, as parasites must adapt to a large phenotypic change in the host and do not have access to intermediate populations. Using the host bacterium *Pseudomonas fluorescens* and the lytic phage Φ2, Benmayor et al. (2009) manipulated the ratio of sensitive to resistant hosts and observed an apparent trade-off between mutation supply and selection for broader infectivity ranges, in much the same way as epistasis affected our ASYM scenario. Building on this work, Hall et al. (2011a) demonstrated that phages evolved broader infectivity

ranges when hosts were allowed to coevolve. This microbial system is known to exhibit strong positive epistasis between infectivity mutations (Scanlan et al., 2011), so these findings are in excellent agreement with the results of our stochastic simulations. In another recent study involving bacteria and viruses (Meyer et al., 2012), a lytic mutant of phage lambda required four mutations to infect resistant *Escherichia coli*. These mutations showed all-or-nothing (i.e. strongly positive) epistasis, but in contrast to studies on *P. fluorescens*, the acquisition of these mutations resulted from fitness benefits on a sub-population of bacteria that had reverted to susceptibility. As such, broader infectivity ranges were again promoted by a gradual change in host phenotype, but this was facilitated by host polymorphism.

Several studies have explored the evolution of broader infectivity ranges, using either explicit genetics (as here; see also Sasaki, 2000; Agrawal and Lively, 2002, 2003; Fenton and Brockhurst, 2007; Fenton et al., 2009, 2012) or by treating infectivity as a quantitative trait (e.g. Best et al., 2010b). However, few theoretical studies have examined how different forms of epistasis influence parasite evolution. As an exception, Fenton and Brockhurst (2007) explored the role of accelerating, decelerating and linear fitness costs on coevolutionary dynamics in a GFG framework. Our study complements this work by focusing on the effects of epistasis on infectivity, while still allowing qualitatively different forms of fitness costs to exist. Similarly, studies focusing specifically on genetic factors that influence the evolution of infectivity ranges are rare. Poullain and Nuismer (2012) recently demonstrated that frameworks of infection genetics with overlapping ranges (e.g. GFG; Flor, 1956; Sasaki, 2000) are better suited to allow adaptation to a novel host than if ranges are disjoint (e.g. matching-alleles; Hamilton, 1980; Penn and Potts, 1999), but the authors did not consider the effects of multilocus interactions, epistasis, or the rate of phenotypic change in the host on the evolution of infectivity ranges, as have been explored in the present study.

We have focused on how epistasis influences parasite evolution, as the emergence

of broader infectivity ranges is of greater biological relevance to epidemiology and public health. Still, it is also important to understand how epistasis shapes host evolution, as this can elucidate patterns of selection among parasites (figure A.5). For example, hosts did not tend to become resistant in the stochastic SYM scenario when epistasis was negative, which explains why broader infectivity ranges did not evolve under these conditions. Host-parasite coevolution can lead to a variety of coevolutionary outcomes with respect to infectivity range, including competitive exclusion, stable polymorphism and fluctuating selection (Sasaki, 2000; Fenton and Brockhurst, 2007). Indeed, our models produced fluctuations in range under certain conditions, but the primary purpose of our study has been to highlight conditions that promote or inhibit the accumulation of mutations that confer broader infectivity ranges, rather than whether they are evolutionarily stable. Hence, we chose to measure peak rather than final infectivity range, as the latter would have been heavily influenced by the choice of simulation length when fluctuations occurred. While evolutionary stability is often an important consideration, there are many circumstances where the initial emergence of a trait is more significant. For example, the chief concern in the context of emerging infectious diseases is whether a parasite will be able to cause an epidemic in a new population, rather than if it will survive over longer timescales. Compensatory mutations may also arise after a trait has initially spread, which could allow parasites to offset associated fitness costs. We hope to address questions surrounding the evolutionary stability of infectivity ranges under various genetic and ecological conditions in future work.

Understanding the effects of epistasis in real biological systems represents a significant challenge for empiricists and we are not aware of many non-microbial systems where the strength of interactions between infectivity mutations have been accurately measured (Hall and Ebert, 2013). Finding a suitable host-parasite system to test the effects of weaker epistasis under different ecological conditions may prove to

be difficult, but some of the predictions from our model should be relatively straightforward to verify experimentally with current approaches that utilise coevolving and non-coevolving communities.

Chapter 3

Spatial structure mitigates fitness costs in host-parasite coevolution

Ben Ashby[1], Sunetra Gupta[1] and Angus Buckling[2]

Abstract

The extent of population mixing is known to influence the coevolutionary outcomes of many host and parasite traits, including the evolution of generalism (the ability to resist or infect a broad range of genotypes). While the segregation of populations into interconnected demes has been shown to influence the evolution of generalism, the role of local interactions between individuals is unclear. Here, we combine an individual-based model of microbial communities with a well established framework of genetic specificity that matches empirical observations of bacterium-phage interactions. We find the evolution of generalism in well-mixed populations to be highly sensitive to the severity of associated fitness costs, but the constraining effect of costs on the evolution of generalism is lessened in spatially structured populations. The contrasting outcomes between the two environments can be explained by different scales of competition (i.e. global versus local). These findings suggest that local interactions may have important effects on the evolution of generalism in host-parasite interactions, particularly in the presence of high fitness costs.

[1]Department of Zoology, University of Oxford, Oxford, UK
[2]Biosciences, University of Exeter, Cornwall Campus, Penryn, UK

3.1 Introduction

Antagonistic coevolution between hosts and parasites is often associated with the emergence of generalism, where populations develop the ability to resist or infect a broad range of genotypes. This means that contemporary populations may be well-adapted to ancestral lineages, but perform poorly against future populations (Buckling and Rainey, 2002a; Mizoguchi et al., 2003; Scanlan et al., 2011). The fundamental principles of these 'coevolutionary arms races' are captured by the gene-for-gene (GFG) framework, in which hosts can avoid infection by accumulating resistance alleles at multiple loci, but parasites can counter these adaptations by gaining infectivity alleles at matching loci (Flor, 1956; Sasaki, 2000). Hence there is a 'gene-for-gene' correspondence between resistance and infectivity alleles. (Note that the literature often refers to these parasite adaptations as 'virulence' alleles, but to avoid confusion with disease severity we will refer to them as 'infectivity' alleles instead). Under the GFG framework, parasites must match or exceed the host's resistance alleles at each locus to have a high probability of causing an infection, which naturally leads to the evolution of generalism in the form of broader resistance and infectivity ranges. These dynamics have been observed in a variety of real host-parasite relationships, including bacterium-phage (Bohannan and Lenski, 2000; Buckling and Rainey, 2002a; Mizoguchi et al., 2003; Brockhurst et al., 2006; Forde et al., 2008; Scanlan et al., 2011), plant-pathogen (Flor, 1956; Thompson and Burdon, 1992; Thrall and Burdon, 2003) and nematode-bacterium systems (Schulte et al., 2010). Recent studies of bacterium-phage coevolution have found that infectivity range is correlated with the number of amino acid changes in tail fibres relative to the ancestral genotype (Scanlan et al., 2011), providing further support for the GFG framework. However, coevolutionary arms races are unlikely to be maintained indefinitely as fitness costs associated with generalism (usually in the form of lower growth/infectivity rates) can reduce selection

for broad ranges (Chao et al., 1977; Webster and Woolhouse, 1999; Sasaki, 2000; Bohannan et al., 2002; Lopez-Pascua and Buckling, 2008; Poullain et al., 2008). Sasaki (2000) predicted that fitness costs will lead to fluctuations between specialism (narrow range) and generalism (broad range), but empirical observations suggest that fitness costs may instead lead to fluctuating selection among genotypes with similar ranges (Hall et al., 2011b).

Fitness costs clearly have considerable influence on the extent of range expansion among hosts and parasites, but other factors are known to have an equally profound impact on coevolutionary dynamics. In particular, it is well established that spatial structure affects both epidemiological dynamics and the scale of competition within a population (Thrall and Burdon, 2003; Forde et al., 2004, 2007; Morgan et al., 2007) and can allow polymorphism to be maintained even in the absence of fitness costs (Damgaard, 1999). In a spatially structured environment the optimal genotype for a particular location will depend on local selection pressures, which may differ between locations and from what would be considered the globally-optimal genotype in a well-mixed population (Thompson, 1994). Experiments with the bacterium *Pseudomonas fluorescens* and the lytic phage $\Phi 2$ suggest that while limited population mixing will slow down the rate of coevolution (Brockhurst et al., 2003), it may also provide more stable conditions for coexistence (Brockhurst et al., 2006). Most empirical studies exploring the effects of spatial structure on range expansion have focused on scenarios where the population is split into interconnected demes, mainly to address questions associated with local adaptation (Burdon and Thrall, 1999; Thrall and Burdon, 2002, 2003; Forde et al., 2004, 2007; Morgan et al., 2007). Similarly, theoretical studies have generally been limited to metapopulation analyses (Frank, 1993a; Gandon et al., 1996; Damgaard, 1999; Gandon et al., 2008), which incorporate a certain degree of spatial structure, but do not capture local interactions between individuals within subpopulations, which are known to be critical in many epidemiological scenarios

(Rand et al., 1995; Rhodes and Anderson, 1996; Keeling et al., 2001; Eames and Keeling, 2002). Individual-based models are able to capture local interactions and have been used to study a diverse set of biological phenomena including the evolution of life histories and virulence (Boots and Sasaki, 1999; Haraguchi and Sasaki, 2000; Read and Keeling, 2003; Heilmann et al., 2010), altruism (Jansen and van Baalen, 2006), and various other aspects of coevolution (Hartvigsen and Levin, 1997; Kerr et al., 2006; Mitchell et al., 2006; Best et al., 2011; Haerter et al., 2011; Zaman et al., 2011; Heilmann et al., 2012). However, the role of local interactions on range expansion has yet to be determined. Here, we attempt to address this gap in the literature by adapting an individual-based model of bacteria and phages first proposed by Heilmann et al. (2010). Although the model was originally used to explore the evolution of virulence in spatially structured populations, it can be readily adapted to serve our focus on range expansion by incorporating the multilocus GFG framework of Sasaki (2000). The model implements spatial structure by situating hosts and parasites on a two dimensional grid, which is of particular relevance to bacteria (Kerr et al., 2006; Hellweger and Bucci, 2009) as colonies often live attached to surfaces in biofilms (Matz et al., 2005; Faruque et al., 2006), providing potential spatial refuges to infection by phages (Levin and Bull, 2004; Gallet et al., 2009).

Our primary focus in the present study is to explore how the impact of fitness costs associated with range expansion is affected by the degree of population mixing. We show that global competition in well-mixed populations leads to rapid selective sweeps, preventing range expansion at high fitness costs. In spatially structured environments however, we find that local competition and spatial clustering can maintain selection for broader ranges even when fitness costs are high.

3.2 Methods

3.2.1 Model description

3.2.1.1 Genetic specificity

The genetic specificity of our model is based on the multilocus gene-for-gene (GFG) framework proposed by Sasaki (2000). Host and parasite genotypes are represented by binary strings of length n ($h_1^i...h_n^i$ for host genotype i and $p_1^j...p_n^j$ for parasite genotype j), where each locus corresponds to the presence (1) or absence (0) of a resistance or infectivity allele. For example, a string of 000 represents a highly-susceptible host (or a specialist parasite), whereas a string of 111 represents a highly-resistant host (or a generalist parasite). We follow Sasaki (2000) by assuming a resistance allele at a particular locus is only effective against parasites that do not have a corresponding infectivity allele at that location and each effective resistance allele reduces the probability of infection by a factor of σ. The parameter σ represents the strength of resistance conferred by each locus: when $\sigma \approx 0$, the acquisition of a single resistance allele will lead to a strong reduction in susceptibility, but when $\sigma \approx 1$ each allele has only a mild effect. We define Q_{ij} to be the infectivity of parasite j on host i, such that:

$$Q_{ij} = \sigma^{d_{ij}}, d_{ij} = \sum_{k=1}^{n} h_k^i \left(1 - p_k^j\right) \tag{3.1}$$

where d_{ij} is the sum of effective resistance alleles.

3.2.1.2 Simulation rules

We adapt the bacterium-phage model proposed by Heilmann et al. (2010) to incorporate the GFG framework outlined above, thus allowing the evolution of varying de-

grees of generalism. We conduct simulations on a square grid of side length $N = 100$, where boundary effects are removed by wrapping the grid around the surface of a torus, so that all grid sites have exactly four orthogonal neighbours. A maximum of one host is allowed per grid site, so that each location is either empty or contains an infected or uninfected host; there are no restrictions on parasite density. The initial grid consists of uninfected hosts at every site and 500 parasites at one location; both populations start without any resistance or infectivity alleles (i.e. $\sum_k h_k^i = \sum_k p_k^j = 0$). The grid is updated synchronously at the end of each time step. We implement two versions of our model (spatial and well-mixed) based on the following rules, which mostly follow those of Heilmann et al. (2010):

Host replication. *Spatial version*: A host is only able to replicate if it satisfies the following criteria: (i) The host is uninfected; and (ii) At least T time steps have elapsed since the host's previous replication event (tracked by individual replication timers). If these criteria are satisfied, then replication proceeds with probability $Ec_H(i)$, where E is the proportion of empty grid sites adjacent to the host and $c_H(i)$ is a fitness cost associated with resistance, given by:

$$c_H(i) = \exp\left(-\eta_H \left|i\right|\right) \tag{3.2}$$

with $|i| = \sum_k h_k^i$ equal to the total number of resistance alleles for host genotype i and η_H a scaling parameter for the strength of the fitness cost. Note that fitness costs were not included in the original model by Heilmann et al. (2010) due to the absence of host range expansion. Offspring are placed in a randomly chosen empty grid site adjacent to their parent; if multiple offspring attempt to occupy the same grid location, then one is chosen at random to survive and the others are removed from

the population. The replication timers for successful parents and offspring are reset following this procedure. Mutations occur with probability ϵ_H at each locus, with the restriction that parents and offspring can only differ by one bit. *Well-mixed version*: As per the spatial version, except that: (i) the probability of replication is equal to $\bar{E}c_H(i)$, where $\bar{E}$ is the proportion of sites across the entire grid that are empty; and (ii) new offspring are placed at randomly chosen empty grid sites.

Infection. *Both versions*: We modify the overall probability of infection derived by Heilmann et al. (2010) to allow competition between multiple host and parasite genotypes. Given a probability of infection (α) and decay (δ) per free parasite, the probability that genotype j is able to infect host genotype i is given by:

$$\rho_\alpha(i,j) = 1 - \exp\left(-\alpha P(j) c_P(j) Q_{ij} \left(\frac{1-\exp(-\delta)}{\delta}\right)\right) \tag{3.3}$$

where Q_{ij} is the strength of interaction between host and parasite and $P(j)$ is the local density of the parasite. Broader infectivity ranges are associated with fitness costs, which reduce the probability of infection, captured here by $c_P(j) = \exp\left(-\eta_P |j|\right)$, where η_P scales the strength of the fitness cost and $|j| = \sum_k p_k^j$ is the total number of infectivity alleles for the parasite. The probability that at least one parasite is able to infect the host is given by:

$$z_1(i) = 1 - \prod_k \left(1 - \rho_\alpha(i,k)\right) \tag{3.4}$$

If a uniform random number, $RAND_1 \in (0,1)$, satisfies $RAND_1 < z_1(i)$, then one parasite strain is chosen at random to infect the host. The proba-

bility of parasite j causing the infection is then equal to $\rho_\alpha(i,j)/\sum_k \rho_\alpha(i,k)$. We assume that coinfection does not occur.

Parasite decay. *Both versions*: Free parasites decay with probability $1-\exp(-\delta)$ and are immediately removed from the environment.

Parasite diffusion. *Both versions*: We assume that parasites move between adjacent grid sites with probability proportional to the negative concentration gradient between those locations, with diffusion constant D. Parasites are restricted to one grid site movement per time step.

Parasite-induced host mortality. *Spatial version*: Infected hosts are killed after a fixed number of time steps (latent period), τ, which results in the release of β new parasites into the environment. New parasites are placed at the same grid site as the newly deceased host. If a uniform random number, $RAND_2 \in (0,1)$, satisfies $RAND_2 < n\epsilon_P$, then one of the new parasites mutates at a random locus, gaining or losing an infectivity allele accordingly. *Well-mixed version*: As per the spatial version, except parasites are distributed to randomly chosen grid sites.

Natural host mortality. *Both versions*: All hosts die with probability μ per time step. Parasites are not released into the environment if an infected host dies before the full latent period has expired.

Host mixing. *Well-mixed version only*: Hosts are randomly assigned new grid sites at the end of each time step.

3.2.2 Analysis

We draw all parameters from uniform random distributions (table B.1), except for the strength of the fitness cost for hosts (η_H), the probability of natural death (μ) and the

number of loci (n). A total of 500 simulations are conducted for each combination (s) of these parameters in both spatially structured and well-mixed environments. We run simulations for a maximum of 10,000 time steps and parameter combinations where hosts or parasites die out in either environment are discarded from further analysis. We allow a burn-in period for the first half of each simulation before measuring the mean resistance, $R^*(s)$, and infectivity, $I^*(s)$, ranges over the final 5,000 time steps. Each measure varies between 0 and 1, where 0 indicates no range expansion by any members of the population and 1 corresponds to full range expansion by all members of the population. The resistance range of the host population at time step t is given by $R(s,t) = \sum_{i=1}^{n} ix(i,s,t)/n$, where $x(i,s,t)$ is the proportion of the host population that has i resistance alleles. Similarly, the infectivity range of the parasite population at time step t is given by $I(s,t) = \sum_{i=1}^{n} iy(i,s,t)/n$, where $y(i,s,t)$ is the proportion of the parasite population that has i infectivity alleles. We $\log_{10}$-transform the data and use analysis of covariance (ANCOVA) to explore the extent to which spatial structure influences range expansion. In addition, we define the following two pairwise comparisons to visualize differences between the environments:

$$C_R(s) = \frac{R^*_{\text{mixed}} - R^*_{\text{spatial}}}{R^*_{\text{mixed}} + R^*_{\text{spatial}}} \tag{3.5}$$

$$C_I(s) = \frac{I^*_{\text{mixed}} - I^*_{\text{spatial}}}{I^*_{\text{mixed}} + I^*_{\text{spatial}}} \tag{3.6}$$

where the subscripts mixed and spatial correspond to the two environments. These measures vary between -1 and 1, indicating whether ranges are on average broader (> 0) or narrower (< 0) in well-mixed environments. Values close to the extremes correspond to a large disparity in evolutionary outcomes between the environments.

Finally, we measure the 'patchiness' of spatially structured populations to assess how the distribution of host patch sizes varies with time. At a given time step, each

patch, q, is defined by a unique set of orthogonally connected hosts sharing the same genotype. We measure the richness and evenness of these patches using the Shannon index, $H(s,t) = -\sum_q a(q,s,t)\ln(a(q,s,t))$, where $a(q,s,t)$ is the number of hosts in patch q (Shannon, 1948). This index increases as the number of patches grows and/or the distribution of patch sizes becomes more even. We define the 'relative patchiness' of the population to be $H'(s,t) = H(s,t)/\bar{H}(s)$, where $\bar{H}(s)$ is the average value of $H(s,t)$ over the final 5,000 time steps of the simulation. Hence, if H' is consistently close to 1, then the patchiness of the population has reached a stable distribution.

3.2.3 Probabilistic cellular automata

We also explore a simplified version of our primary model using probabilistic cellular automata (PCA), which does not permit the existence of free-living parasites in the environment (Bak and Chao, 1990). The model structure and analysis for the PCA are largely similar to the methods described above for our primary model; the full PCA simulation rules, analysis and results are detailed in appendix B.3.

3.3 Results

For the sake of brevity we only present data for $n = 3$ here, but similar results were obtained for other numbers of loci (see appendix B.2).

3.3.1 Epidemiological dynamics

We began our investigation by qualitatively comparing epidemiological dynamics in the presence and absence of spatial structure. Overall, 63% of simulations led to host-parasite coexistence for 10,000 generations in both environments. We observed that well-mixed environments were prone to large amplitude epidemic cycles, whereas spatially structured populations typically demonstrated much steadier dynamics (fig-

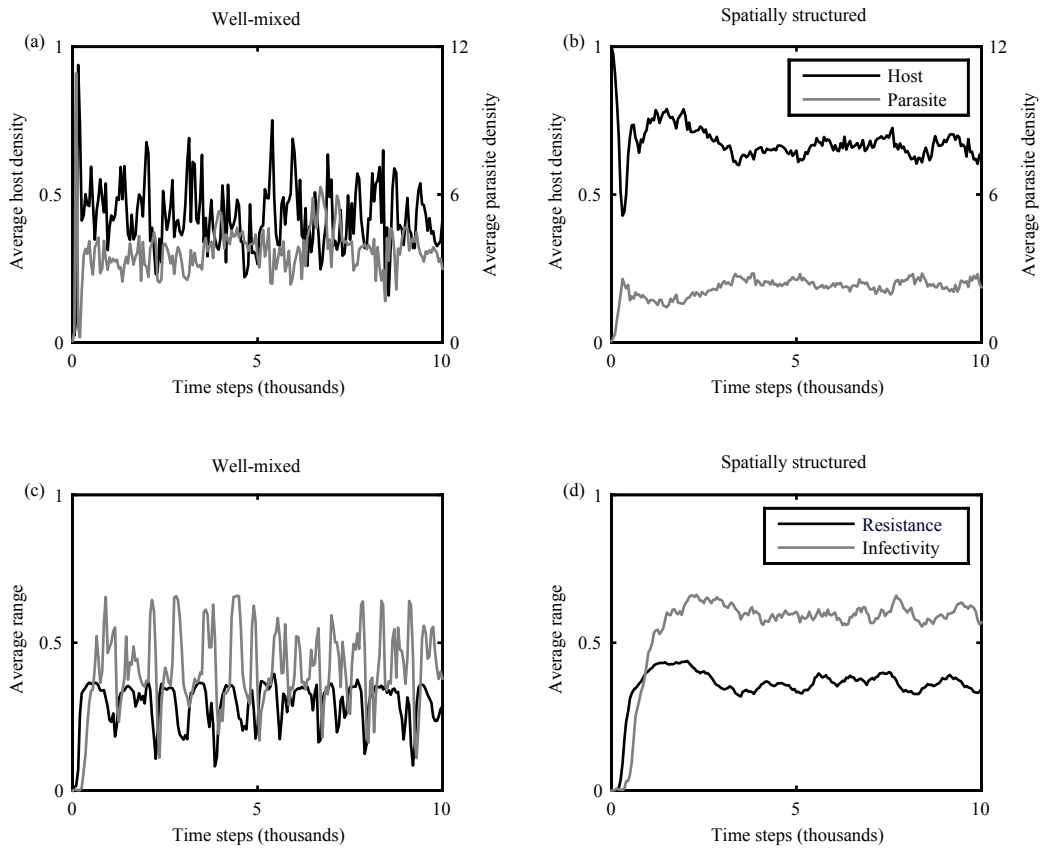

Figure 3.1: Example coevolutionary dynamics from well-mixed (left column) and spatially structured (right column) populations. (a-b) Average density of host (black) and parasite (grey) populations. (c-d) Average resistance (black) and infectivity (grey) ranges. Well-mixed populations are prone to rapid fluctuating dynamics, whereas spatially structured populations generally produce more stable dynamics. Parameters: $\alpha = 0.15$; $\beta = 50$; $\delta = 0.75$; $\epsilon_H = 0.001$; $\epsilon_P = 0.001$; $\eta_H = 1$; $\eta_P = 0.4$; $\mu = 0.001$; $\sigma = 0.25$; $\tau = 3$; $D = 0.5$; $N = 100$; $n = 3$; $T = 3$.

ure 3.1). The patchiness of spatially structured populations (H) typically increased sharply during the first 1,000 time steps of a simulation, before settling down to a stable distribution (figures 3.2-3.3). The snapshots in figure 3.2 demonstrate how these populations are typically structured into clusters of identical genotypes.

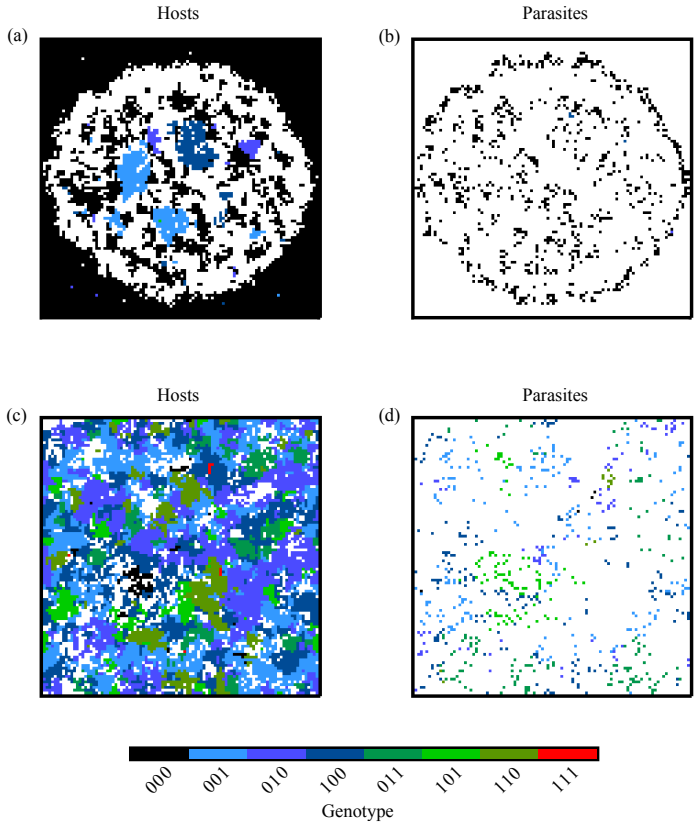

Figure 3.2: Snapshots of a simulation from the spatially structured population in figure 1 at two time points: (a-b) 250 generations; (c-d) 1,000 generations. Hosts (left hand column) and intracellular parasites (right hand column) are coloured according to their genotype (i.e. number of resistance/infectivity alleles, where $n = 3$), with: black equal to no alleles (000); blues equal to 1 allele (001, 010 or 100); greens equal to 2 alleles (011, 101, or 110); red equal to 3 alleles (111). White regions indicate no hosts/intracellular parasites. The Shannon index for the host population, $H(t)$, is equal to: (a) 2.19 and (b) 5.19 at these two time points.

3.3.2 Resistance

We found that the evolution of broad resistance ranges was highly constrained by the severity of fitness costs (η_H) and to a lesser extent by the probability of natural mortality (μ) (figure 3.4a). When $\mu = 0.001$, R^* was on average 80% lower in the presence of high fitness costs ($\eta_H = 2$) compared to no fitness costs. This difference increased to 88% for $\mu = 0.01$ and to 98% for $\mu = 0.05$, as higher values of μ led to

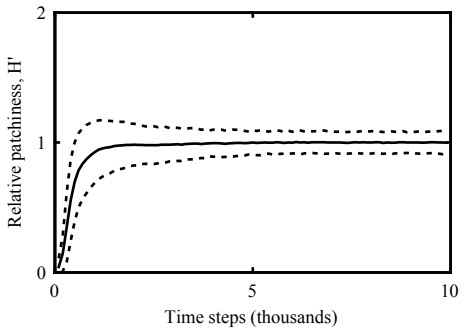

Figure 3.3: Relative patchiness of spatially structured populations as a function of time (mean ± standard deviation) for $n = 3$. Patchiness is defined by the Shannon index (as described in the main text), which measures the richness and evenness of patch sizes at a given time point; relative patchiness is equal to this value divided by the average over the final 5,000 generations, giving an indication as to how the spatial structure of the population tends to change during the course of a simulation. The first 1,000 generations are characterized by a rapid increase in patchiness, which then plateaus. The standard deviation does not increase over the final 5,000 generations, which implies that patchiness remains roughly constant within each simulation over this period.

a faster turnover in the host population, reducing the benefits associated with broad resistance ranges. Whilst both spatially structured and well-mixed populations experienced a decrease in R^* with η_H, the magnitude of the constraint was generally much greater in the absence of spatial structure (figures 3.4b, 3.6a; ANCOVA, environment $\times$ fitness cost interaction for $\mu = 0.001$: $F_{1,3234} = 194$, $p < .0001$; see table B.3 for full ANCOVA results). The difference in evolutionary outcomes in the two environments is highlighted by the pairwise comparison (C_R) in figure 3.4b, which shows that spatially structured populations tended to evolve broader resistance ranges than comparable well-mixed populations and that this disparity increased with greater fitness costs, provided the turnover rate of the host population was not too fast. For sufficiently large values of μ, the two environments exhibited broadly similar levels of resistance in the presence of high fitness costs.

We examined the robustness of our results by varying the period during which

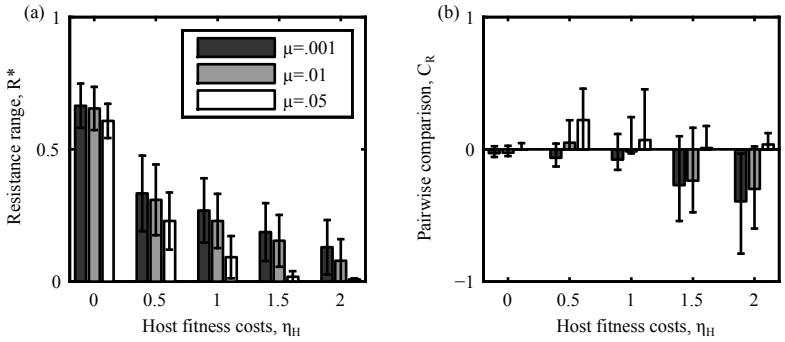

Figure 3.4: The effects of fitness costs, spatial structure and the natural mortality rate on the resistance range of the host population for $n = 3$. (a). The average resistance range (taken over both environments), R^*, tends to decrease with the host fitness cost parameter, η_H, and the natural mortality rate, μ. (b). A pairwise comparison, C_R, of resistance range expansion in the two environments. If $C_R > 0$, then the mixed environment tends to produce broader resistance ranges than the spatially structured environment and vice versa if $C_R < 0$. Spatially structured populations generally exhibit broader resistance ranges than well-mixed populations as the associated fitness costs increase. However, high natural mortality rates can negate this effect, as resistance becomes less beneficial in both environments.

resistance ranges were recorded. The median difference in resistance ranges between time steps 5,001-7,500 and 7,500-10,000 was 2.8%, indicating that the host population had reached a quasi-steady state prior to the beginning of the measurement window. Minor oscillations were often observed about the mean values of R^* (as shown in figure 3.1), with slightly larger fluctuations more likely to occur in well-mixed populations (median difference between R^* and the peak value of R: 4.6% for well-mixed and 2.7% for spatially structured populations).

3.3.3 Infectivity

Parasite evolution was closely linked to changes in the host population, which meant that host fitness costs and natural mortality had similar effects on I^* to those described above for R^* (figure 3.5a). This is unsurprising given that parasites also

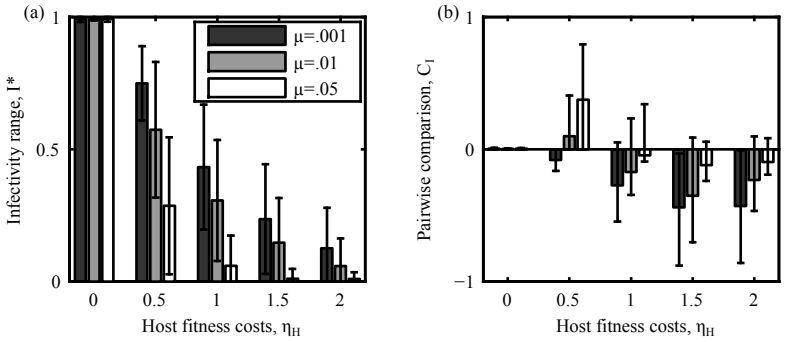

Figure 3.5: The effects of fitness costs, spatial structure and the natural mortality rate on the infectivity range of the parasite population for $n = 3$. (a). The average infectivity range (taken over both environments), I^*, tends to decrease with the host fitness cost parameter, η_H, and the natural mortality rate, μ. (b). A pairwise comparison, C_I, of infectivity range expansion in the two environments. If $C_I > 0$, then the mixed environment tends to produce broader infectivity ranges than the spatially structured environment and vice versa if $C_I < 0$. Spatially structured populations generally exhibit broader infectivity ranges than well-mixed populations as the fitness costs for hosts increase. However, high natural mortality rates can negate this effect, as selection for infectivity decreases in both environments.

experienced a fitness cost associated with generalism (η_P), so that broader infectivity ranges were unlikely to be selected for unless resistance was widespread. Hence, I^* decreased with η_H and μ in both environments, but parasites in spatially structured populations tended to exhibit broader ranges than those in well-mixed populations as η_H increased, provided the host population did not exhibit rapid turnover (figures 3.5b, 3.6b; ANCOVA, environment $\times$ fitness cost interaction for $\mu = 0.001$: $F_{1,3234} = 157$, $p < .0001$; see table B.3 for full ANCOVA results).

As observed in the host population, average ranges among parasites were found to be stable over different measurement windows (median difference of 3.9% between time steps 5,001-7,500 and 7,501-10,000) and minor oscillations about the mean values of I^* were common (median difference between I^* and the peak value of I within each simulation: 5% for well-mixed and 4.7% for spatially structured populations).

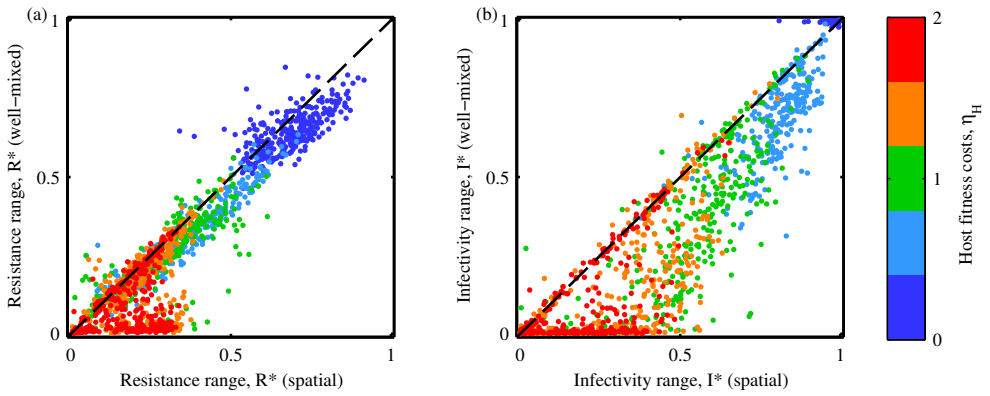

Figure 3.6: Scatter plots showing (a) resistance and (b) infectivity ranges in the two environments for each simulation. Colours correspond to the host fitness cost parameter, η_H. Spatially structured populations consistently exhibit broader ranges than well-mixed populations, particularly when fitness costs are high. Parameters: $\mu = 0.001$; $n = 3$.

3.3.4 Probabilistic cellular automata

To test the generality of our results, we created a simplified version of our model based on probabilistic cellular automata (PCA; see appendix B.3 for full description and results). We found that both the epidemiological dynamics and coevolutionary outcomes were qualitatively similar to those observed in the primary model, especially when moving between environments. Most importantly, our key finding was replicated using this alternative approach: spatially structured populations tended to produce broader resistance and infectivity ranges than well-mixed populations as fitness costs increased.

3.4 Discussion

Using an individual-based model of host-parasite coevolution, we explored the effects of spatial structure on the expansion of resistance and infectivity ranges. Consistent with previous theoretical (Hochberg and van Baalen, 1998; Sasaki, 2000; Agrawal

and Lively, 2002) and empirical studies (Bohannan and Lenski, 2000; Forde et al., 2004; Buckling et al., 2006; Forde et al., 2008; Lopez-Pascua and Buckling, 2008; Lopez-Pascua et al., 2012), we found resistance and infectivity range evolution to be constrained by fitness costs. Our novel result is that the magnitude of the constraint imposed by host fitness costs on range expansion is highly dependent on the degree of mixing in the population. More specifically, we found that high host fitness costs severely restricted range expansion in well-mixed environments, but the effect was much weaker in the presence of spatial structure. While host fitness costs did not have a direct impact on parasite fitness, we assumed that range expansion was still costly for parasites and so would only be selected for in the presence of resistant hosts (see Sasaki, 2000). Thus, parasite infectivity behaved in a qualitatively similar way to resistance in response to greater host fitness costs.

To test the generality of our results we constructed a second model, replacing the free-living parasite population with one transmitted via direct contact between hosts (PCA; appendix B.3). We found the results to be broadly consistent, despite the stark contrast in transmission dynamics between the two models, which suggests that our findings are likely to have general applicability to a variety of host-parasite systems. Our study is closely linked to the work of Best et al. (2011), who reported an important effect of spatial structure on the coevolution of quantitative resistance and infectivity (as opposed to the resistance and infectivity ranges explored here): high resistance and low virulence were favoured when dispersal was limited, but this trend was reversed in well-mixed environments.

The contrasting dynamics observed in the two environments can be explained by two processes. First, individuals in poorly-mixed populations only compete with a subset of the population at very small spatial scales, whereas individuals in well-mixed environments are in competition with the entire population. In structured environments the fitness of a particular genotype will vary depending on the neighbouring

population, which means that different genotypes may be optimal at different locations and spatial scales (Thompson, 1994; Gandon et al., 1996). Second, clustering and limited dispersal can limit the extent to which a genotype can spread in spatially structured populations, even if it is optimal for a given location (Campos et al., 2008).

As a specific example of the impact of these two related processes, first consider a well-mixed population initially composed of sensitive hosts and specialist (narrow range) parasites in a single locus GFG framework. If a resistant host emerges, it is likely to have a huge advantage over the sensitive population, and will therefore rapidly increase in prevalence. Specialist parasites will then be at a disadvantage, so generalists that can infect both host types will eventually dominate. Assuming resistance is associated with slower growth, but provides no protection against the generalist parasite, the sensitive population will then gradually increase relative to the resistant population. Similarly, generalist parasites have a lower fitness than specialist parasites when the host population is not resistant. Hence, the specialist parasite increases in prevalence and the cycle repeats, giving fluctuations in range. In a spatially structured environment, the emergence of resistant hosts and generalist parasites will initially follow a similar pattern. Although sensitive hosts will then have the globally-optimal genotype, they may not be able to realize their growth advantage as clustering may limit the extent to which they can spread before being wiped out by parasites. In addition, if small numbers of specialist parasites are maintained by these sensitive patches then resistance could still be the locally optimal trait. Thus, local competition and clustering provide ephemeral refuges for globally sub-optimal genotypes, which make spatially structured populations less likely to exhibit fluctuations in range. Similar dynamics emerge in our multilocus framework, as shown in figure 3.1. Resistance initially spreads in both environments, but they respond differently to the emergence of generalist parasites, with broader ranges persisting in the presence of spatial structure. We found that the difference in coevolutionary

outcomes between the two environments was dependent on the probability of natural death for the host, given by the parameter μ. Higher values of μ correspond to faster turnover rates in the host population, which increases the severity of fitness costs to the point where resistance is no longer beneficial in either environment. In addition, a faster population turnover rate will reduce the effects of clustering, allowing sensitive hosts with faster growth rates to re-establish themselves in the presence of generalist parasites.

The genetic specificity in our model was based on a well-established multilocus GFG framework that can produce arms race coevolutionary dynamics as well as fluctuations in range (Sasaki, 2000; Fenton et al., 2009). While there is considerable evidence for coevolutionary arms races taking place among bacteria and phages (Bohannan and Lenski, 2000; Buckling and Rainey, 2002a; Mizoguchi et al., 2003; Brockhurst et al., 2006; Forde et al., 2008; Scanlan et al., 2011) and various other host-parasite systems (Little et al., 2006; Schulte et al., 2010), there is limited evidence of fluctuations in range except for some plant-fungus interactions (e.g. Thrall and Burdon, 2003). Recent work with *P. fluorescens* and lytic phages has shown that fluctuating selection between genotypes with similar ranges is possible, either following (Hall et al., 2011b) or in the absence of a coevolutionary arms race (Gomez and Buckling, 2011). In addition, the frequent occurrence of local adaptation indicates that there may be multiple routes to generalism (Buckling and Rainey, 2002b; Morgan et al., 2005; Vos et al., 2009; Koskella et al., 2011). These data indicate that the GFG framework may only be capturing part of the genetic interactions between bacteria and phages, which has led others to propose more complex specificities (Agrawal and Lively, 2002, 2003; Weitz et al., 2005; Forde et al., 2008; Fenton et al., 2012).

Furthermore, some systems appear to be based on other forms of specificity that do not permit generalism. For example, Carius et al. (2001) observed that the bacterium *Pasteuria ramosa* specialises on different lineages of the freshwater crustacean *Daph-*

nia magna, which can lead to fluctuating selection between different genotypes rather than an escalatory arms race (Decaestecker et al., 2007). Similarly, coevolutionary dynamics between the freshwater snail *Potamopyrgus antipodarum* and trematode parasites of the genus *Microphallus* appear to be governed by fluctuating selection between specialists, a process which has been linked to the maintenance of sexual reproduction in the host population (Lively, 1987; King et al., 2009). Whilst generalists have not yet been observed in these systems, it is possible that the fitness costs associated with broad ranges are simply too high.

Our work complements the growing body of research on the effects of spatial structure on coevolutionary dynamics (Hartvigsen and Levin, 1997; Boots and Sasaki, 1999; Haraguchi and Sasaki, 2000; Read and Keeling, 2003; Jansen and van Baalen, 2006; Kerr et al., 2006; Mitchell et al., 2006; Heilmann et al., 2010; Best et al., 2011; Haerter et al., 2011; Zaman et al., 2011; Heilmann et al., 2012). There are also strong links between this study and a variety of ecological models on victim-exploiter relationships. Of particular relevance is the work on host and parasitoids, in which variation in dispersal rate can lead to a range of complex dynamics, with high rates of dispersal increasing extinction risk and leading to fluctuations in population sizes, as observed here (Hassell et al., 1991; Comins et al., 1992; Pascual, 1993). Together, these studies highlight the important role that spatial structure plays in shaping both ecological and evolutionary dynamics.

Chapter 4

A symmetric gene-for-gene framework for host-parasite coevolution

Ben Ashby[1] and Angus Buckling[2]

Abstract

The development of realistic models of host-parasite infection genetics (specificity) is key to improving our understanding of coevolutionary dynamics. We introduce a novel framework for modelling specificity, which produces dynamics that are consistent with a number of empirical observations of coevolution. We refer to this type of specificity as a 'symmetric gene-for-gene' (SGFG) interaction, as both hosts and parasites are able to trigger multi-step coevolutionary arms races, unlike other gene-for-gene models where one population is reactive. We then extend the SGFG framework to incorporate a second component that is governed by 'matching alleles' (MA) specificity. This two step model (SGFG-MA) provides plausible explanations as to why coevolutionary arms races eventually give way to fluctuating selection in real populations and why different environments can produce contrasting coevolutionary dynamics. The majority of previous studies of specificity have used population

[1]Department of Zoology, University of Oxford, Oxford, UK
[2]Biosciences, University of Exeter, Cornwall Campus, Penryn, UK

genetics approaches that lack ecological feedbacks; here, we compare coevolutionary outcomes in population genetics and population dynamics models and find that the inclusion of ecological feedbacks is crucial for producing realistic outcomes.

4.1 Introduction

Many host-parasite relationships show a clear genetic basis for resistance and infectivity, with each parasite genotype only able to infect a particular set of host genotypes (Allison, 1954; Dybdahl and Lively, 1998; Carius et al., 2001; Thrall and Burdon, 2002; Scanlan et al., 2011). Such genotype-by-genotype interactions can produce a variety of coevolutionary dynamics, including arms races (directional selection for broader resistance and infectivity ranges; Buckling and Rainey, 2002a; Jones and Dangl, 2006) and fluctuating selection (negative frequency-dependence; Dybdahl and Lively, 1998; Decaestecker et al., 2007; Gomez and Buckling, 2011), which has led to much interest in models of specificity (Frank, 1993a, 1994, 1996, 2000; Parker, 1994; Sasaki, 2000; Agrawal and Lively, 2002, 2003; Fenton et al., 2009, 2012). Although models of specificity are typically used to interpret qualitative differences in coevolutionary dynamics (i.e. arms races or fluctuating selection), they are also vital for understanding the maintenance of static trait diversity (Boots et al., 2014), how populations diverge and become locally adapted (Thrall and Burdon, 2002; Morgan et al., 2005; King et al., 2009) and if it is possible to predict evolutionary trajectories (i.e. forecasting which mutations will become fixed) (Meyer et al., 2012; Russell et al., 2012; Koskella and Brockhurst, 2014). Most studies of specificity have neglected population dynamics and have focused solely on population genetics (but see May and Anderson, 1983; Gandon and Day, 2009; Best et al., 2010b; Boots et al., 2014), yet ecological feedbacks are known to play a fundamental role in host-parasite relationships. For example, ecological feedbacks have been shown to limit fluctuations in gene

frequencies governing specificity and to reduce selection for sexual reproduction (May and Anderson, 1983; Ashby and Gupta, 2014). It is therefore crucial that frameworks of specificity are analysed in ecologically-realistic models.

Two general frameworks have been proposed for modelling host-parasite specificity. The 'matching alleles' (MA) approach assumes that parasites are forced to specialise on distinct subsets of host genotypes, which typically leads to fluctuating selection due to negative frequency-dependence as no genotype is intrinsically better than any other (Hamilton, 1980; Frank, 1993b). In contrast, 'gene-for-gene' (GFG) approaches assume that hosts and parasites exhibit varying degrees of generalism, with certain genotypes only able to resist or infect a narrow subset of the population (specialists) and others able to perform well against a broad range of genotypes (generalists) (Flor, 1956; Sasaki, 2000). These assumptions can lead to arms race dynamics (Best et al., 2010b), but may also produce fluctuating selection between specialists and generalists if broader resistance and infectivity ranges are associated with fitness costs (Sasaki, 2000; Fenton et al., 2009). Several host-parasite systems, including water snails and trematodes (Dybdahl and Lively, 1998) and water fleas and bacteria (Carius et al., 2001; Decaestecker et al., 2007) are prone to fluctuating selection and appear to conform to MA infection genetics, while others such as plants and pathogenic fungi (Thrall and Burdon, 2002; Jones and Dangl, 2006) produce arms races, in line with the GFG theory. However, the distinction between the two models is not necessarily clear-cut, with some authors noting that MA and GFG systems represent the extremes of a continuum (Parker, 1994; Agrawal and Lively, 2002).

There is no reason to assume that one framework of specificity is representative of all populations, but the infection genetics for any given host-parasite system must be consistent across all ecological conditions. Yet bacteria and phages have been shown to exhibit both GFG-like arms races and MA-like fluctuating selection depending on the environment and the duration of coevolution (Gomez and Buckling, 2011; Hall

et al., 2011b), which points towards more complex models of specificity (Agrawal and Lively, 2003; Forde et al., 2008; Fenton et al., 2012). Agrawal and Lively (2003) presented a two-step model of specificity involving MA infection genetics at one stage and GFG at the other and showed that the strength of fitness costs associated with the GFG component affected which loci were under selection, in much the same way that different ecological conditions can alter the dominant mode of selection among bacteria and phages (Gomez and Buckling, 2011). However, Fenton et al. (2012) noted that parasites in this system must only avoid host defences at one stage during the infection process, which means that parasites with narrow ranges (specialists) can still infect broadly resistant hosts (generalists), but this is not representative of how many hosts and parasites interact (e.g. bacteria from later time points in coevolutionary experiments are typically able to resist ancestral phages with narrow ranges; Buckling and Rainey, 2002a; Mizoguchi et al., 2003; Scanlan et al., 2011).

Fenton et al. (2012) proposed an alternative framework combining initial recognition of the host by the parasite (using an inverse gene-for-gene (IGFG) framework; Fenton et al., 2009) followed by recognition of the parasite by the host (using a GFG framework; Sasaki, 2000) and demonstrated the potential for fluctuating selection in functionally independent pathways. This model better suits our mechanistic understanding of two-step infection processes, but the dynamics are not entirely consistent with empirical observations of bacteria and phages. Existing models of GFG/IGFG-based specificity predict that parasites should evolve broad ranges and hosts should only experience selection for one or two resistance alleles (parasites must 'hedge their bets'; Sasaki, 2000), but real populations exhibit multi-step arms races with much greater variation in resistance (Scanlan et al., 2011). In addition, GFG/IGFG-based models typically produce fluctuations in range (i.e. cycling between specialists and generalists; Sasaki, 2000; Agrawal and Lively, 2002, 2003; Fenton et al., 2009, 2012), but microbial communities tend to exhibit arms races followed by fluctuation selection

between genotypes of similar ranges (Hall et al., 2011b).

Here, we introduce a novel approach to infection genetics, combining aspects of the GFG, IGFG and MA frameworks. We begin by relaxing some of the assumptions made by the GFG and IGFG frameworks, leading to a new form of specificity that we refer to as a 'symmetric gene-for-gene' (SGFG) interaction. We then incorporate a second component to the infection process, based on MA infection genetics, and explore the consequences of this two-step model for coevolutionary dynamics under a range of ecological conditions. Crucially, we also explore the importance of ecological feedbacks by comparing the behaviour of both population genetics and population dynamics models.

4.2 A symmetric gene-for-gene (SGFG) framework

4.2.1 Description of the SGFG framework

We propose a novel framework for modelling host-parasite genetic specificity, which is primarily based on empirical observations of bacteria and phages (Buckling and Rainey, 2002a; Mizoguchi et al., 2003; Hall et al., 2011b,a; Scanlan et al., 2011). Specificity is governed by interactions at n biallelic loci; a '1' at a particular locus corresponds to the presence of a resistance or infectivity allele and a '0' corresponds to its absence. Hence, genotypes for both populations can be represented by binary strings of length n ($h_1^i \dots h_n^i$ for host genotype i and $p_1^j \dots p_n^j$ for parasite genotype j). Previous studies have assumed that resistance alleles are only effective when parasites do not have a corresponding infectivity allele at the same locus (gene-for-gene, GFG; Sasaki, 2000), or vice versa (inverse gene-for-gene, IGFG; Fenton et al., 2009), but here we assume that neither population has an inherent advantage in a 'symmetric gene-for-gene' (SGFG) interaction. Instead, the presence (1) of a resistance allele at a locus where the parasite doesn't have an infectivity allele (0) results in a reduction in

(a) GFG

	P_{00}	P_{01}	P_{10}	P_{11}
H_{00}	1	1	1	1
H_{01}	σ	1	σ	1
H_{10}	σ	σ	1	1
H_{11}	σ^2	σ	σ	1

(b) IGFG

	P_{00}	P_{01}	P_{10}	P_{11}
H_{00}	σ^2	σ	σ	1
H_{01}	σ^2	σ^2	σ	σ
H_{10}	σ^2	σ	σ^2	σ
H_{11}	σ^2	σ^2	σ^2	σ^2

(c) SGFG

	P_{00}	P_{01}	P_{10}	P_{11}
H_{00}	σ^2	σ	σ	1
H_{01}	σ^3	σ^2	σ^2	σ
H_{10}	σ^3	σ^2	σ^2	σ
H_{11}	σ^4	σ^3	σ^3	σ^2

Figure 4.1: Genetic specificity in the (a) gene-for-gene (GFG), (b) inverse gene-for-gene (IGFG) and (c) symmetric gene-for-gene (SGFG) frameworks. Interactions between hosts (H) and parasites (P) are shown for a two-locus system, where subscripts correspond to the presence (1) or absence (0) of a resistance or infectivity allele at a given locus. Interactions at each locus can reduce the probability of infection by a factor of σ ($0 < \sigma < 1$). The SGFG framework is obtained by taking the element-wise product of the interaction matrices given by the GFG and IGFG frameworks.

infectivity by a factor of σ ($0 < \sigma < 1$), whereas the presence (1) of an infectivity allele at a locus where the host doesn't have a resistance allele (0) results in an increase in infectivity by a factor of $1/\sigma$ (figure 4.1). We define Q_{ij} to be the infectivity of parasite j on host i, such that:

$$Q_{ij} = \sigma^{n+d_{ij}} \tag{4.1}$$

where $d_{ij} = \sum_{k=1}^{n} \left(h_k^i - p_k^j\right)$. Hence Q_{ij} is minimised when hosts have a complete set of resistance alleles and parasites have no infectivity alleles ($Q_{ij} = \sigma^{2n}$) and is maximised when the reverse is true ($Q_{ij} = 1$).

4.2.2 Population genetics approach

4.2.2.1 Model description

We begin our exploration of the SGFG framework by adapting the population genetics approach taken by Sasaki (2000) and Fenton et al. (2009), which is based on the following fitness functions for hosts (w_H)

$$w_H(i) = f_H(i) \exp\left(-\beta_H \sum_j Q_{ij} P_j\right) \tag{4.2}$$

and parasites (w_P)

$$w_P(j) = f_P(j) \exp\left(\beta_P \sum_i Q_{ij} H_i\right) \tag{4.3}$$

where H_i and P_j are the frequencies of host and parasite genotypes i and j, β_H and β_P are the costs and benefits of infection for hosts and parasites, respectively, and $f_H(i)$ and $f_P(j)$ are cost functions associated with generalism:

$$f_H(i) = \exp\left(-c_H |i|^{\kappa_H}\right) \tag{4.4}$$

$$f_P(j) = \exp\left(-c_P |j|^{\kappa_P}\right) \tag{4.5}$$

The parameters c_H and κ_H (c_P and κ_P) modify the strength of fitness costs associated with broader ranges through the accumulation of resistance (infectivity) alleles, with higher values of κ_H (κ_P) increasing the rate at which fitness costs accelerate with greater generalism. The functions $|i| = \sum_k h_k^i$ and $|j| = \sum_k p_k^j$ are equal to the total number of resistance and infectivity alleles for host and parasite genotypes i and j, respectively. The frequencies of each genotype change according to the following set of coupled ordinary differential equations:

$$\frac{dH_i}{dt} = \left(\frac{w_H(i)}{\sum_k w_H(k) H_k} - 1\right) H_i \tag{4.6}$$

$$\frac{dP_j}{dt} = \left(\frac{w_P(j)}{\sum_k w_P(k) P_k} - 1\right) P_j \tag{4.7}$$

with sums taken over all genotypes. Note that population sizes are infinite and

mutations are ignored.

4.2.2.2 Results

Under what conditions will hosts and parasites broaden their respective ranges through the accumulation of resistance and infectivity alleles? Suppose that hosts initially do not possess any resistance alleles and that parasites possess $v \in [0, \ldots, n]$ infectivity alleles. By comparing the fitness of all host genotypes (equation 4.2), we see that hosts can accumulate $u \in [0, \ldots, n]$ resistance alleles if:

$$\frac{c_H u^{\kappa_H}}{\beta_H \sigma^{n-v}} + \sigma^u < 1 \tag{4.8}$$

Similarly, if parasites initially do not possess any infectivity alleles and hosts possess u resistance alleles, parasites can accumulate v infectivity alleles if:

$$\sigma^{-v} - \frac{c_P v^{\kappa_P}}{\beta_P \sigma^{n+u}} > 1 \tag{4.9}$$

The system has a coevolutionarily stable state (CoESS) at $(\bar{u}, \bar{v})$ if the following conditions are satisfied for all $u \neq \bar{u}$ and $v \neq \bar{v}$:

$$\sigma^{\bar{u}} < \sigma^u - \frac{c_H \left(\bar{u}^{\kappa_H} - u^{\kappa_H}\right)}{\beta_H \sigma^{n-\bar{v}}} \tag{4.10}$$

$$\sigma^{-\bar{v}} < \sigma^{-v} + \frac{c_P \left(\bar{v}^{\kappa_P} - v^{\kappa_P}\right)}{\beta_P \sigma^{n+\bar{u}}} \tag{4.11}$$

If no values of $\bar{u}$ and $\bar{v}$ satisfy both conditions, then the system exhibits coevolutionary cycling (fluctuations in range). The conditions in equations 4.10 and 4.11 can be used to construct phase diagrams that show how the relative strength of fitness costs affects coevolutionary dynamics, revealing several interesting patterns (figure 4.2). First, selection tends to favour parasites with extreme ranges ($\bar{v} = 0$ or $\bar{v} = n$), whereas

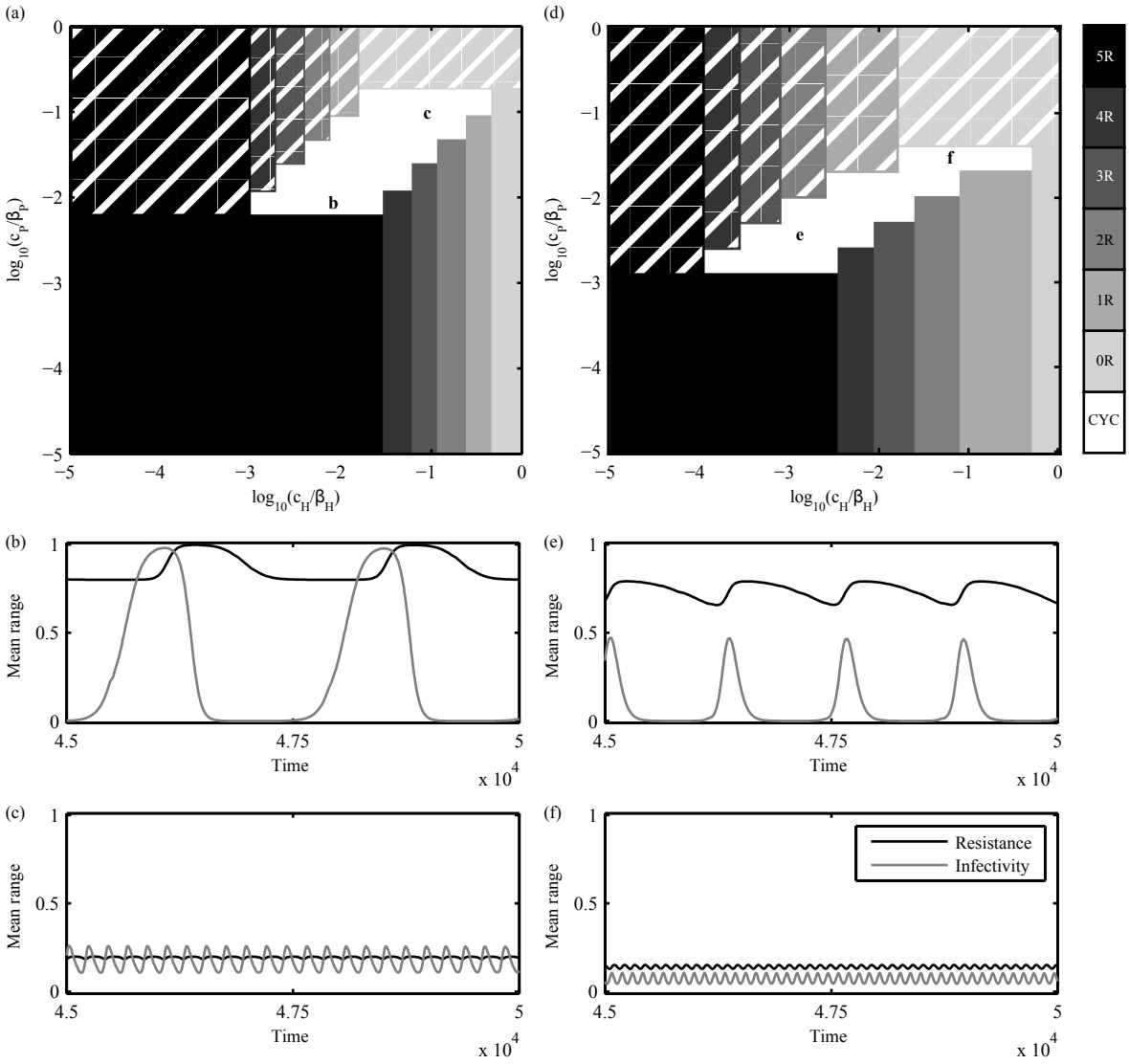

Figure 4.2: Phase diagrams for the population genetics implementation of the SGFG framework (equations 4.2-4.7) for (a-c) slowly ($\kappa_H = \kappa_P = 1$) and (d-f) rapidly ($\kappa_H = \kappa_P = 2$) accelerating fitness costs. (a) Coevolutionary outcomes for relative costs of resistance (c_H/β_H) and infectivity (c_P/β_P). Shading indicates the number of resistance alleles (R) at equilibrium. Solid shaded regions indicate that the parasite population has a complete set of infectivity alleles at equilibrium; hatched regions indicate that there are no infectivity alleles at equilibrium. The white region (CYC) corresponds to coevolutionary cycling, as shown in the time series plots. Parameters: $\beta_H = \beta_P = 1$, $\sigma = 0.5$ and $n = 5$.

hosts can experience selection for any level of resistance, depending on the relative cost of each allele and the degree of investment by the parasite. In fact, no intermediate values of $\bar{v}$ were obtained for $\kappa_H, \kappa_P \leq 2$, so if the system is at equilibrium, parasites must have either invested fully in becoming generalists, or not at all. When fitness costs accelerate very rapidly ($\kappa_H, \kappa_P > 2$) the model dynamics are largely dominated by selection against generalism, as even intermediate levels of resistance or infectivity are very costly. Second, fluctuations in range are reasonably common and tend to have larger amplitudes when fitness costs are relatively low (figures 4.2b, e). Third, the SGFG framework permits a much greater variety of coevolutionary dynamics than allowed by other single-step models of genetic specificity (Sasaki, 2000; Fenton et al., 2009). Figures 4.2a and 4.2d show 13 distinct coevolutionary outcomes for the SGFG framework, whereas corresponding phase diagrams for the GFG (Sasaki, 2000) and IGFG (Fenton et al., 2009) frameworks have only been shown to contain 4 or 5 regions, respectively. Furthermore, the GFG and IGFG frameworks tend to produce hosts with low numbers of resistance alleles, whereas parasites usually have to evolve broad ranges as a bet-hedging strategy. By contrast, the SGFG framework can select for broad ranges in both populations.

This basic analysis of the SGFG framework reveals that it is capable of producing a variety of coevolutionary outcomes that have not been captured by previous models of genetic specificity. However, the population genetics approach that has been used in this section lacks many features of real populations that can play a crucial role in coevolutionary dynamics, including finite population sizes, ecological feedbacks, mutations and stochasticity. As such, we shift our focus to an epidemiological model, which captures more realistic population dynamics.

4.2.3 Population dynamics approach

4.2.3.1 Model description

We incorporate the genetic specificity of the SGFG framework into an epidemiological model that is based on bacterium-lytic phage interactions. We split the host and parasite populations into the following non-overlapping classes: susceptible hosts of genotype i (S_i), hosts infected by parasite genotype j (I_j) and parasites of genotype j in the environment (P_j). Lytic phages typically prevent host reproduction, so we do not track the genotypes of infected hosts. We assume that density-dependence limits the growth rate of the host population, so that offspring of genotype i are produced asexually at a rate $r_i S_i(1 - N_H/K)$, where r_i is the maximum per-capita growth rate of host genotype i, K modifies the strength of density-dependence and $N_H = \sum_i S_i + \sum_j I_j$ is the total host population size.

Mixing between hosts and parasites is random. If an uninfected host of genotype i meets a parasite of genotype j, then infection occurs at a rate of $\beta_j Q_{ij}$, where β_j is the adsorption rate for the parasite and Q_{ij} is as described in equation 4.1. For simplicity, we assume that coinfection does not occur and that hosts cannot recover from infection. Parasites that successfully cause an infection increase the rate of host death from μ (natural mortality) to $\alpha + \mu$, where α is the mortality rate due to parasitism. Hosts that are killed by disease release γ new parasites into the environment, which decay at a rate δ. Infected hosts that die naturally do not release parasites into the environment. We incorporate fitness costs associated with broader resistance and infectivity ranges using the same functional forms described in equations 4.4 and 4.5, to give the following genotype dependent growth (r_i)

$$r_i = r^* f_H(i) \tag{4.12}$$

and adsorption (β_j)

$$\beta_j = \beta^* f_P(j) \tag{4.13}$$

parameters, where r^* and β^* are base growth and adsorption rates, respectively.

In a deterministic setting, the population dynamics of our model can be represented by the following set of ordinary differential equations:

$$\left.\begin{aligned} \frac{dS_i}{dt} &= r_i S_i \left(1 - \frac{N_H}{K}\right) - S_i \left(\sum_j Q_{ij}\beta_j P_j + \mu\right) \\ \frac{dI_j}{dt} &= \beta_j P_j \sum_i Q_{ij} S_i - (\alpha + \mu) I_j \\ \frac{dP_j}{dt} &= \alpha\gamma I_j - \beta_j P_j \sum_i Q_{ij} S_i - \delta P_j \end{aligned}\right\} \tag{4.14}$$

which we also translate into a stochastic model using the τ-leap method proposed by Gillespie (2001). The τ-leap method uses a fixed step size, τ, and assumes that the number of events occurring within a time step is Poisson distributed. Since we are using arbitrary time units, we set $\tau = 1$ and scale the remaining model parameters appropriately. In the stochastic version of our model, the binary strings representing each host genotype (see section 4.2.1) are subject to mutation at a rate of ϵ_H at each locus, with the restriction that multiple mutations cannot arise simultaneously (i.e. the genotypes of parent and progeny never differ at more than one locus). Similarly, mutations among parasites occur at a rate of ϵ_P at each locus with the restriction that successive generations cannot differ at multiple loci. We begin both deterministic and stochastic simulations with K susceptible hosts and parasites in the environment; initial populations possess neither any resistance, nor any infectivity alleles.

4.2.3.2 Results

The coevolutionary dynamics of the epidemiological model are in many ways similar to the dynamics described for the population genetics model in section 4.2.2.2,

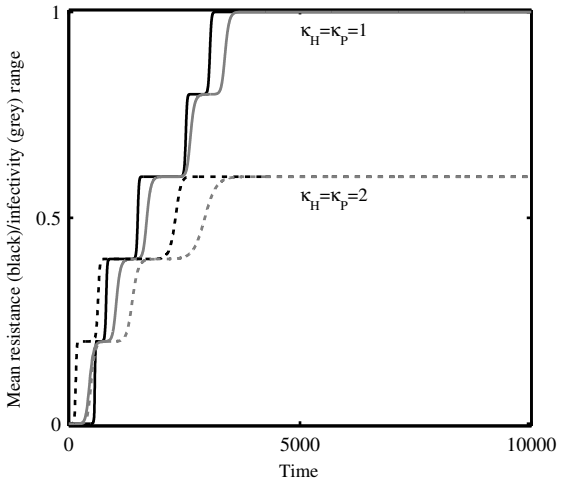

Figure 4.3: Example coevolutionary trajectories from the stochastic version of the SGFG epidemiological model (equation 4.14) for hosts (black) and parasites (grey) with slowly ($\kappa_H = \kappa_P = 1$; solid) and rapidly ($\kappa_H = \kappa_P = 2$; dotted) accelerating fitness costs. Mean ranges correspond to the average proportion of loci in the population that contain resistance and infectivity alleles. Coevolutionary dynamics are characterised by multi-step arms races, with no fluctuations in range. Parameters: $\alpha = 0.05$, $\beta = 0.05$, $\gamma = 50$, $\delta = 0.25$, $\epsilon_H = \epsilon_P = 10^{-7}$, $\mu = 10^{-3}$, $\sigma = 0.5$, $c_H = c_P = 0.1$, $K = 10^6$, $n = 5$ and $r^* = 0.5$.

with reciprocal expansions in resistance and infectivity ranges likely to occur over a broad set of parameters. However, there are also substantial differences between the dynamics of the two models. First, fluctuations in range are relatively common when using the population genetics approach (figure 4.2), but we have only detected stable equilibria in the epidemiological model, despite making extensive parameter sweeps with both deterministic and stochastic implementations. This is because the population genetics approach does not feature density-dependent processes (ecological feedbacks), which are less likely to produce cyclical dynamics (see chapter 5). Crucially, these dynamics are more consistent with our current understanding of host-parasite coevolution (Buckling and Rainey, 2002a; Mizoguchi et al., 2003; Thrall and Burdon, 2003; Hall et al., 2011b; Scanlan et al., 2011). Second, equilibria in the population genetics approach are characterised by parasites with extreme ranges

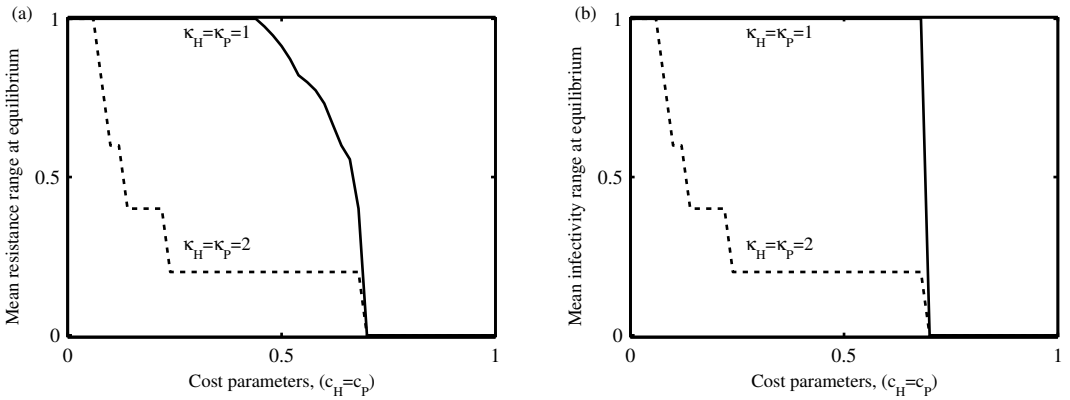

Figure 4.4: Mean (a) resistance and (b) infectivity ranges at equilibrium in the deterministic version of the epidemiological model for slowly ($\kappa_H = \kappa_P = 1$; solid) and rapidly ($\kappa_H = \kappa_P = 2$; dotted) accelerating fitness costs. The cost parameters c_H and c_P are varied along the x-axis in tandem (i.e. the curves correspond to a diagonal slice through the $c_H - c_P$ plane). Rapidly accelerating fitness costs can result in the fixation of some, but not all resistance and infectivity alleles; note that intermediate levels of infectivity are only stable in the population genetics model for $\kappa_P > 2$. Slowly accelerating fitness costs produce maximal or minimal levels of infectivity, in line with the population genetics model (figure 4.2). Parameters as specified in figure 4.3, with $\epsilon_H = \epsilon_P = 0$.

(i.e. either a full set of infectivity alleles or none) unless fitness costs accelerate very rapidly ($\kappa_P > 2$). By contrast, intermediate levels of infectivity are stable in the epidemiological model for more credible values of κ_P (figures 4.3-4.4). The fixation of intermediate numbers of resistance and infectivity alleles is particularly striking as this represents a plausible mechanism through which coevolutionary arms races can lead to population divergence, as different alleles can become fixed in segregated populations (figure 4.5). Furthermore, structured populations may experience heterogeneous environments with unequal fitness costs across space (e.g. due to variable patch quality), allowing different levels of resistance and infectivity to evolve in coevolutionary 'hot' and 'cold' spots, in accordance with the geographic mosaic theory of coevolution (Gomulkiewicz et al., 2000; King et al., 2009).

We have shown that the SGFG framework more closely matches empirical observa-

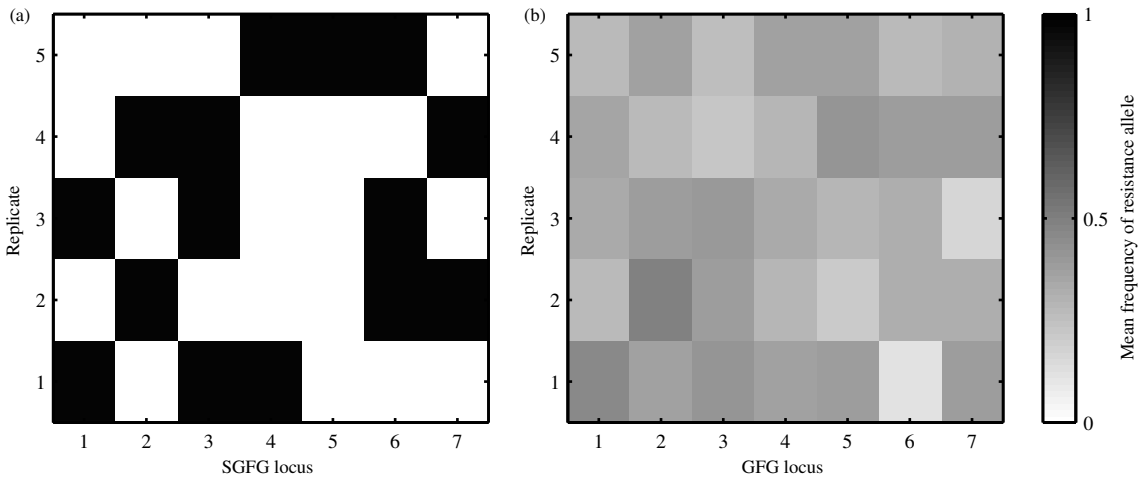

Figure 4.5: Example coevolutionary outcomes in multiple simulations of the stochastic epidemiological model with (a) SGFG and (b) GFG infection genetics. (a) Different alleles become fixed in the SGFG framework, allowing replicate populations to diverge (mean genetic distance: 0.57). (b) By contrast, alleles in the GFG framework ($Q_{ij} = \sigma^{D_{ij}}$, $D_{ij} = \sum_k h_k^i(1 - p_k^j)$) do not become fixed, so populations do not diverge (mean genetic distance: 0.10). The 'genetic distance' between two host populations is equal to the average difference in the frequency of each resistance allele. Parameters as specified in figure 4.3, with $\kappa_H = \kappa_P = 2$ and $n = 7$.

tions of host-parasite coevolution when implemented in a full epidemiological model, which highlights the importance of including ecological feedbacks when modelling coevolutionary dynamics. However, our epidemiological model is not entirely consistent with experimental data, as it does not appear to produce any form of coevolutionary cycling (Gomez and Buckling, 2011; Hall et al., 2011b). Yet, Hall et al. (2011b) showed that arms races between the bacterium *Pseudomonas fluorescens* and its associated lytic phage Φ2 eventually give way to fluctuating selection between genotypes with similar ranges; here, arms races tend towards equilibria, rather than producing fluctuating selection. When coevolution between bacteria and phages takes place in soil rather than in nutrient-rich media broth, arms races do not occur and fluctuating selection dominates instead (Gomez and Buckling, 2011). Together, these findings suggest that specificity is governed by a second component, which tends to dictate

coevolutionary dynamics when fitness costs associated with further range expansion are high. To explore this possibility, we extend the SGFG framework to include a second type of locus, which determines how well the parasite 'matches' the host.

4.3 Two-step infection genetics: symmetric gene-for-gene (SGFG) and matching-alleles (MA)

4.3.1 Description of the SGFG-MA framework

We follow previous two-step models of infection genetics by adding a second type of biallelic locus to the SGFG framework (Agrawal and Lively, 2003; Fenton et al., 2012). Interactions at the second type of locus are governed by 'matching alleles' (MA) infection genetics, which forces parasites to specialise on a subset of hosts (Hamilton, 1980; Penn and Potts, 1999). Genotypes are of the form $h_1^i \dots h_n^i / x_1^i \dots x_m^i$ for hosts and $p_1^i \dots p_n^i / y_1^i \dots y_m^i$ for parasites, where n is the number of SGFG loci and m is the number of MA loci. SGFG loci function as described in section 4.2.1. Each MA locus contains either an 'A' or a 'B' allele and mismatches between host and parasite result in a reduction in infectivity by a factor of ρ ($0 < \rho < 1$) per locus (figure 4.6). For example, the infectivity (Q) of parasite $00/AA$ on host $00/AA$ is σ^n (equation 4.1), but is reduced to $\rho\sigma^n$ on host $00/AB$ and to $\rho^2\sigma^n$ on host $00/BB$. Thus, SGFG loci affect the range of genotypes that a host may resist or that a parasite may infect, whereas MA loci affect the subset of the host population on which a parasite specialises. We analyse the SGFG-MA framework by adapting the epidemiological model from section 4.2.3.1 to incorporate this form of specificity. We assume that MA loci have no effect on fitness costs and that mutations are equally likely to occur at any locus.

SGFG-MA

	$P_{00/A}$	$P_{00/B}$	$P_{01/A}$	$P_{01/B}$	$P_{10/A}$	$P_{10/B}$	$P_{11/A}$	$P_{11/B}$
$H_{00/A}$	σ^2	$\rho\sigma^2$	σ	$\rho\sigma$	σ	$\rho\sigma$	1	ρ
$H_{00/B}$	$\rho\sigma^2$	σ^2	$\rho\sigma$	σ	$\rho\sigma$	σ	ρ	1
$H_{01/A}$	σ^3	$\rho\sigma^3$	σ^2	$\rho\sigma^2$	σ^2	$\rho\sigma^2$	σ	$\rho\sigma$
$H_{01/B}$	$\rho\sigma^3$	σ^3	$\rho\sigma^2$	σ^2	$\rho\sigma^2$	σ^2	$\rho\sigma$	σ
$H_{10/A}$	σ^3	$\rho\sigma^3$	σ^2	$\rho\sigma^2$	σ^2	$\rho\sigma^2$	σ	$\rho\sigma$
$H_{10/B}$	$\rho\sigma^3$	σ^3	$\rho\sigma^2$	σ^2	$\rho\sigma^2$	σ^2	$\rho\sigma$	σ
$H_{11/A}$	σ^4	$\rho\sigma^4$	σ^3	$\rho\sigma^3$	σ^3	$\rho\sigma^3$	σ^2	$\rho\sigma^2$
$H_{11/B}$	$\rho\sigma^4$	σ^4	$\rho\sigma^3$	σ^3	$\rho\sigma^3$	σ^3	$\rho\sigma^2$	σ^2

Figure 4.6: Genetic specificity in the SGFG-MA framework. Host (H_i) and parasite (P_j) genotypes are composed of both SGFG (alleles 0 and 1) and MA (alleles A and B) loci, which combine to give the specificity (infectivity, Q) of each genotype-by-genotype interaction. SGFG loci affect the range of genotypes that a host may resist or a parasite may infect, whereas MA loci affect the subset of the host population on which a parasites specialises. The presence (1) of an SGFG resistance allele at a locus where the parasite doesn't have a corresponding infectivity allele (0) results in a reduction in infectivity by a factor of σ ($0 < \sigma < 1$), whereas the presence (1) of an infectivity allele at a locus where the host doesn't have a resistance allele (0) results in an increase in infectivity by a factor of $1/\sigma$. Mismatches between host and parasite genotypes at MA loci result in a reduction in infectivity by a factor of ρ ($0 < \rho < 1$).

4.3.2 Coevolutionary dynamics of the SGFG-MA framework

The SGFG-MA framework produces coevolutionary dynamics that closely match empirical observations of bacterium-phage coevolution (figure 4.7), with fluctuating selection occurring between genotypes of identical ranges, either following an arms race (Hall et al., 2011b) or in its absence (Gomez and Buckling, 2011). When fitness costs are initially low hosts and parasites experience selection for broader ranges, but this process is ultimately restricted by physiological constraints (no further mutations exist that confer broader ranges; figure 4.7a) or rapidly accelerating fitness costs (figure 4.7b). Fluctuating selection between genotypes with identical ranges

then becomes dominant, as changes at MA loci allow hosts to avoid contemporary parasites without incurring a fitness cost (figures 4.7d-e). When fitness costs are initially high selection favours narrow ranges so arms races do not occur, but fluctuating selection at MA loci still takes place (figures 4.7c, f). In essence these scenarios are quite similar, in that populations quickly evolve towards optimal levels of generalism depending on the strength of fitness costs, but fluctuating selection at MA loci dominates the long-term behaviour of the system.

Interestingly, the amplitude and frequency of oscillations at MA loci are both minimised for intermediate costs (figure 4.8), even though intrinsic fitness is only dependent on alleles at SGFG loci. This somewhat surprising result is due to the effects of fitness costs on host-parasite contact rates. When fitness costs are very low, individuals with the broadest ranges have only marginally slower growth or adsorption rates than individuals with the narrowest ranges, so host-parasite contact rates remain high and hence fluctuating selection is strong (large amplitude and high frequency oscillations). When fitness costs are very high, individuals with narrow ranges are favoured; again, host-parasite contact rates remain high (no reduction in growth/adsorption rates), facilitating strong fluctuating selection. Intermediate fitness costs allow arms races to occur, but growth and adsorption rates are markedly lower for individuals with expanded ranges, which reduces host-parasite contact rates and hence the strength of fluctuating selection. Thus, intermediate fitness costs tend to produce low amplitude and low frequency oscillations at MA loci.

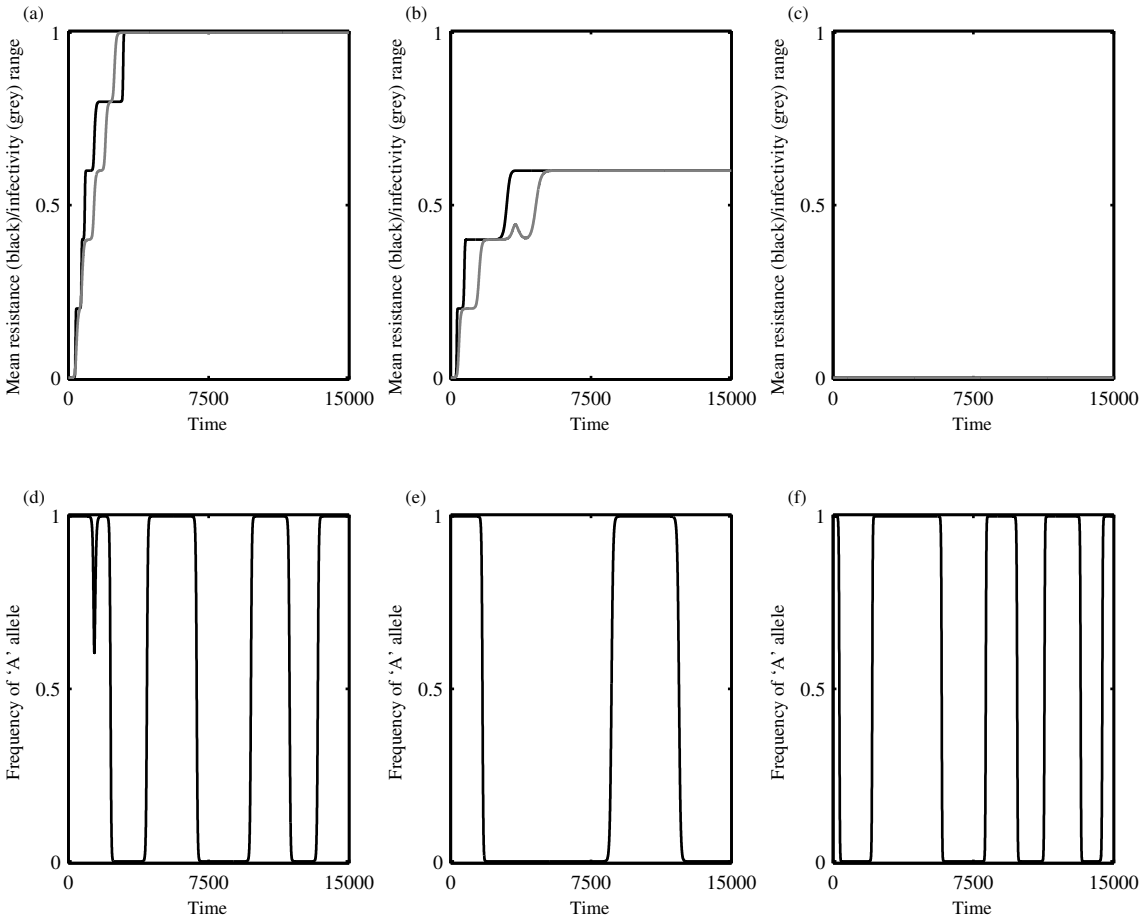

Figure 4.7: Example coevolutionary trajectories from the stochastic version of the SGFG-MA epidemiological model with a single MA locus ($m = 1$). Top row: mean ranges correspond to the average proportion of loci in the population that contain resistance (black) and infectivity (grey) alleles, for low initial fitness costs ($c_H = c_P = 0.1$) that accelerate (a) slowly ($\kappa_H = \kappa_P = 1$) or (b) rapidly ($\kappa_H = \kappa_P = 2$), and (c) high initial fitness costs ($c_H = c_P = 0.75$ and $\kappa_H = \kappa_P = 1$). Bottom row: frequency of the 'A' allele at the MA locus (plots (d)-(f) correspond to the same simulations as (a)-(c)). (a)-(b) Arms race dynamics are still possible when using SGFG-MA infection genetics, similar to the SGFG framework (figure 4.3). However, arms race dynamics eventually give way to fluctuating selection between genotypes with identical ranges (d)-(e), which closely matches empirical observations of bacteria and phages in nutrient-rich media broth (Hall et al., 2011a). When fitness costs are initially high ((c) and (f)), fluctuating selection dominates and no arms race occurs, as observed when bacteria and phages coevolve in soil (Gomez and Buckling, 2011). Remaining parameters as specified in figure 4.3, with $\rho = 0.7$ and $m = 1$.

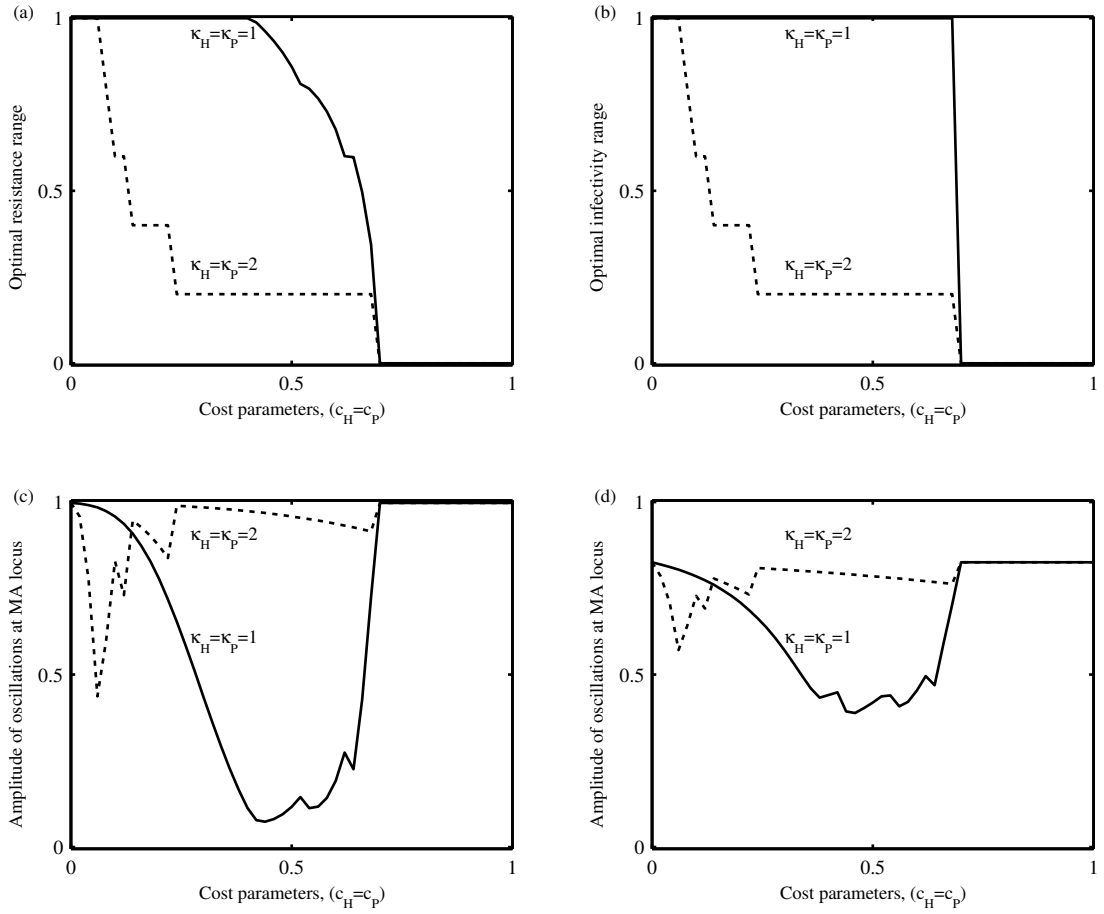

Figure 4.8: Effects of the cost parameters, c_H and c_P on coevolutionary dynamics in the deterministic version of the SGFG-MA framework, when fitness costs accelerate slowly ($(\kappa_H = \kappa_P = 1)$; solid) or rapidly ($\kappa_H = \kappa_P = 2$); dotted). The cost parameters c_H and c_P are varied along the x-axis in tandem (i.e. the curves correspond to a diagonal slice through the $c_H - c_P$ plane). (a)-(b) Range expansion is as described in figure 4.4 for the SGFG framework without MA loci. (c)-(d) Unlike SGFG loci, the MA component of specificity has no effect on the intrinsic fitness of hosts and parasites, yet cost parameters still lead to substantial variation in coevolutionary dynamics at these loci. The amplitude and frequency of oscillations at MA loci are highly correlated (Pearson's correlation coefficient, $r > 0.99$), so figures for the latter have been omitted. Both measures are minimised for intermediate fitness costs, as host-parasite contact rates are lowest in this region of parameter space, which reduces the strength of fluctuating selection (low costs have little effect on growth and adsorption rates and high costs select against range expansion). Parameters as specified in figure 4.3, with $\epsilon_H = \epsilon_P = 0$, $\rho = 0.7$ and $m = 1$.

4.4 Discussion

Empirical studies of host-parasite coevolution have revealed a complex relationship between genotype and phenotype, which can result in a variety of dynamics that are not accurately portrayed by current models of specificity. Examining precisely how different forms of specificity influence patterns of resistance and infectivity, and ultimately, how specificity shapes coevolutionary dynamics, is essential for improving our understanding of the relationship between genotype and phenotype in host-parasite systems. Here, we have developed a novel framework for modelling host-parasite specificity that is capable of producing a much wider variety of coevolutionary dynamics than have been obtained with previous models (Sasaki, 2000; Agrawal and Lively, 2002, 2003; Fenton et al., 2009, 2012). Crucially, other GFG-based models do not allow populations to diverge and only produce fluctuations in range or select for low levels of resistance and high levels of infectivity, whereas our model produces coevolutionary dynamics that closely match empirical observations of multi-step arms races and population divergence (Buckling and Rainey, 2002a; Mizoguchi et al., 2003; Thrall and Burdon, 2003; Hall et al., 2011b; Scanlan et al., 2011).

The SGFG framework behaves differently to other GFG-based models because it assumes that the accumulation of more resistance (infectivity) alleles always leads to a reduction (increase) in infectivity (Q) no matter how much the other population has invested in generalism, which tends to select for broader ranges in both hosts and parasites. This view of infection genetics is reminiscent of Van Valen's (1973) frequently-quoted Red Queen metaphor for antagonistic coevolution ("it takes all the running you can do, to keep in the same place"), as both hosts and parasites can improve their performance against an evolutionarily static population. In contrast, the original IGFG and GFG frameworks assume that hosts and parasites, respectively, are reactive, so arms races can only be triggered by the other population.

The emergence of resistance is often reported as the starting point for coevolutionary arms races, in line with the GFG framework (Chao et al., 1977; Lenski, 1984a; Jones and Dangl, 2006; Scanlan et al., 2011), but this is most likely due to the fact that hosts are victims, so are more likely to experience stronger selection than parasites (exploiters). However, parasites have been shown to evolve broader infectivity ranges on evolutionarily static hosts (Hall et al., 2011a), which indicates that the specificity of some host-parasite interactions may be more symmetric than previously thought (i.e. both populations are capable of triggering an arms race).

We explored an extension to the SGFG framework that incorporates a second type of genetic interaction, where parasites are forced to specialise on a subset of host genotypes (SGFG and matching alleles; SGFG-MA). We used this two-step model of specificity to provide theoretical support for two questions arising from recent experiments with bacteria (*P. fluorescens*) and phages (Φ2): (i) why do arms races eventually give way to fluctuating selection between genotypes of a similar range (Hall et al., 2011b), as opposed to fluctuations between specialists and generalists (Sasaki, 2000)? (ii) Why can environmental variation yield qualitative shifts in coevolutionary dynamics (Gomez and Buckling, 2011)? In both cases, the MA component of specificity, which facilitates fluctuating selection without affecting host or parasite range, dominates coevolutionary dynamics once the optimal number of SGFG alleles have become fixed for a given set of conditions. The optimal number of SGFG alleles depends on the strength of associated fitness costs, so different environments can produce contrasting levels of arms race dynamics. Hall et al. (2011b) speculated that a switch from arms race dynamics to fluctuating selection in microbial communities was most likely due to accelerating fitness costs; our model strongly supports this argument. Moreover, the authors found that phages initially diverged from the ancestral genotype at a very fast rate (∼8 amino acid substitutions (AAS) in the tail fibre gene SBWP25_0036 after 10 transfers), but this began to slow and eventually

stopped (∼13 AAS in SBWP25_0036 after 30 and 60 transfers), consistent with our simulation results: hosts and parasites diverged from their initial genotypes through the accumulation of SGFG alleles, but subsequent fluctuating selection at MA loci did not lead to further divergence. Gomez and Buckling (2011) observed fluctuating selection rather than arms race dynamics between bacteria and phages in soil, which the authors suggested was due to higher fitness costs in a poor quality environment; again, our model supports this theory. Note that contact rates between bacteria and phages will also be much lower in highly structured and heterogeneous environments, reducing selection for costly resistance and infectivity mechanisms.

We found that fluctuating selection at MA loci in our two-step model was indirectly affected by the strength of fitness costs at SGFG loci, with intermediate costs associated with low amplitude and low frequency oscillations. This result has important implications for a number of biological phenomena, including models based on the Red Queen Hypothesis for the maintenance of sex, which can be heavily influenced by the frequency of coevolutionary cycling (Barton, 1995; Peters and Lively, 1999; but see chapter 5). Our model suggests that the evolution of resistance or infectivity mechanisms that confer broader ranges may also alter the frequency of oscillatory dynamics, and in turn, selection for sex.

Our two-step model of infection using both SGFG and MA interactions is in many ways similar to the GFG-MA model proposed by Agrawal and Lively (2003), with both experiencing a switch in the type of loci that dominates selection as fitness costs increase. However, they also differ in a number of key aspects. Most notably, fluctuating selection in the SGFG-MA model only occurs between genotypes with identical ranges, whereas the GFG-MA model also exhibits fluctuating selection between specialists and generalists. Moreover, the switch from selection at one type of loci to another in the GFG-MA model is solely dependent on changes in fitness costs (i.e. between simulations); the SGFG-MA model also demonstrates a switch in the

mode of selection following arms race dynamics (i.e. within simulations), consistent with experimental data (Hall et al., 2011b). The two-step model proposed by Fenton et al. (2012) is more similar to our SGFG framework without the MA component, as both are based on a combination of GFG and IGFG infection genetics. However, Fenton et al. (2012) assumed that GFG and IGFG loci were functionally independent (in much the same way as SGFG and MA loci are in the two-step model presented herein), whereas we have effectively assumed that both GFG and IGFG specificities operate simultaneously at all loci. This means that no genotype in our model has the same value of Q for all pairwise interactions (i.e. no rows or columns of the infectivity matrix, Q, are constant; figure 4.1), which tends to limit fluctuations in range. In fact, we only observe fluctuations in range when ecological feedbacks are ignored (section 4.2.2.2), as density-dependent processes tend to suppress these dynamics. Coevolutionary cycling at MA loci is not suppressed by ecological feedbacks as negative frequency-dependent selection is generally easier to maintain than fluctuations in range (Agrawal and Lively, 2003). (Note that infected hosts are unable to reproduce in the present study so parasites are effectively castrators, which allows coevolutionary cycling to persist indefinitely; see chapter 5).

The contrasting behaviour of our population genetics and population dynamics models highlights the importance of including ecological feedbacks when modelling host-parasite coevolution, but the former approach is often favoured for its analytic tractability (but see May and Anderson, 1983; Gandon and Day, 2009; Best et al., 2010b; Boots et al., 2014). Although the two models produce a similar range of dynamics, fundamental disparities exist that have important consequences for the interpretation of experimental data (Gomez and Buckling, 2011; Hall et al., 2011b). Recently, some authors have moved towards adaptive dynamics approaches, which balance analytic tractability with ecologically-realistic assumptions (Best et al., 2010b; Boots et al., 2014). These studies have demonstrated that continuous-trait mod-

els are able to produce selection for broad ranges in both hosts and parasites, but the present study is the first to demonstrate such behaviour with explicit genetics (discrete traits).

Much of the empirical motivation for the models in the present study has arisen from experimental coevolution with bacteria and phages, as the infection genetics and coevolutionary dynamics of these populations are well-studied (Chao et al., 1977; Lenski and Levin, 1985; Buckling and Rainey, 2002a; Brockhurst et al., 2003; Mizoguchi et al., 2003; Poullain et al., 2008; Lopez-Pascua and Buckling, 2008; Paterson et al., 2010; Flores et al., 2011; Hall et al., 2011a,b; Scanlan et al., 2011; Meyer et al., 2012). In particular, mutations in the tail fibre gene of the lytic phage Φ2 have been identified that govern interactions with the bacterium *P. fluorescens*, with broader infectivity ranges associated with the accumulation of more amino acid substitutions (Scanlan et al., 2011). Similarly, changes to the structure of receptors on the surface of bacteria (e.g. outer membrane proteins) have been linked to resistance in several bacterial populations (Labrie et al., 2010). Nevertheless, our models are likely to be more broadly applicable. For example, *Drosophila melanogaster* and sigma virus show evidence of reciprocal adaptations that are characteristic of an arms race (Bangham et al., 2007), as do the nematode *Caenorhabditis elegans* and its microbial parasite *Bacillus thuringiensis* (Schulte et al., 2010). Although many plant-pathogens appear to exhibit a classic GFG interaction with their hosts, it is possible that an SGFG-like interaction may be revealed if they are allowed to evolve while host genotypes are held constant (as in Hall et al., 2011a).

Many host-parasite interactions appear to be characterised by MA infection genetics (e.g. water snails and trematodes, Dybdahl and Lively, 1998; or water fleas and their bacterial parasites, Carius et al., 2001; Decaestecker et al., 2007), but our two-step model demonstrates that arms races may still be possible, given the correct environmental conditions. However, our models suggest that arms races are likely

to be transient, with fluctuating selection between genotypes of similar ranges the default form of coevolutionary dynamics. This may explain why fluctuating selection appears to be more common than arms race dynamics in natural environments (e.g. Dybdahl and Lively, 1998; Decaestecker et al., 2007; Gomez and Buckling, 2011; but see Bangham et al., 2007). Our results suggest that artificially lowering fitness costs by improving environment quality or increasing contact rates between hosts and parasites could produce transient arms race dynamics in these systems (see also Lopez-Pascua and Buckling, 2008). Although the models presented herein match a broad range of empirical observations, further study of host-parasite infection genetics and coevolutionary dynamics across different environmental conditions will be required to determine the generality of the SGFG and SGFG-MA frameworks.

Chapter 5

Parasitic castration promotes coevolutionary cycling but also imposes a cost on sex

Ben Ashby[1] and Sunetra Gupta[1]

Abstract

Antagonistic coevolution between hosts and parasites is thought to drive a range of biological phenomena including the maintenance of sexual reproduction. Of particular interest are conditions that produce persistent fluctuations in the frequencies of genes governing host-parasite specificity (coevolutionary cycling), as sex may be more beneficial than asexual reproduction in a constantly changing environment. While many studies have shown that coevolutionary cycling can lead to the maintenance of sex, the effects of ecological feedbacks on the persistence of these fluctuations in gene frequencies are not well understood. Here, we use a simple deterministic model that incorporates ecological feedbacks to explore how parasitic reductions in host fecundity affect the maintenance of coevolutionary cycling. We demonstrate that parasitic castration is inherently destabilising and may be necessary for coevolutionary cycling to persist indefinitely, but also reduces the likelihood that sexually-reproducing individuals will find a fertile partner, which may select against sex. These findings

[1]Department of Zoology, University of Oxford, Oxford, UK

suggest that castrators can play an important role in shaping host evolution and are likely to be good targets for observing fluctuations in gene frequencies that govern specificity in host-parasite interactions.

5.1 Introduction

Parasites are near-ubiquitous in nature, sometimes causing severe damage to their hosts and creating strong selection for resistance (Dybdahl and Lively, 1998; Buckling and Rainey, 2002a; Thrall and Burdon, 2003; Decaestecker et al., 2007) or other disease-avoidance mechanisms (Simms and Triplett, 1994; Hamilton and Poulin, 1997). Some host-parasite systems are characterised by directional selection for increasingly sophisticated means of defence and counter-defence (Buckling and Rainey, 2002a; Thrall and Burdon, 2003), but others are prone to negative frequency-dependent selection where genotypes fluctuate in prevalence through time (Dybdahl and Lively, 1998; Decaestecker et al., 2007; Gomez and Buckling, 2011). Persistent fluctuations in host and parasite gene frequencies, often referred to as 'coevolutionary cycling', are of special interest to evolutionary biologists as they may help to resolve prominent questions such as why certain genes are highly polymorphic (e.g. those involved in the major histocompatibility complex; Penman et al., 2013) and why many organisms reproduce sexually. The evolution of sex has received considerable attention from both theoreticians and empiricists in the context of the Red Queen Hypothesis (RQH), which posits that costs associated with sex (e.g. the twofold cost of males, sexual conflict, etc) may be offset by more rapid adaptation to coevolving antagonists (Jaenike, 1978; Maynard Smith, 1978; Hamilton, 1980; Bell, 1982). Specifically, the RQH predicts that sex should be the dominant mode of reproduction if segregation and recombination lead to an increase in the frequency of rare gene combinations that confer resistance to contemporaneous parasites (Peters and Lively, 1999; Gandon and

Otto, 2007). Directional selection may also contribute to the maintenance of sex, as recombination breaks up linkage disequilibria that reduce additive genetic variance, allowing sexual populations to adapt at a faster rate (Barton, 1995; Peters and Lively, 1999).

Although many studies have explored conditions that sustain selection for novel gene combinations due to coevolutionary cycling, most have used traditional population genetics approaches that lack ecological feedbacks (Hamilton, 1980; Hamilton et al., 1990; Sasaki, 2000; Agrawal and Lively, 2002, 2003; Otto and Nuismer, 2004; Agrawal, 2009; Fenton et al., 2012). Ecological feedbacks such as time delays in the life cycle of the parasite, parasitic reductions in host fecundity and seasonal forcing are known to have a destabilising effect on population dynamics that can lead to sustained epidemic cycles (May and Anderson, 1978; Hudson et al., 1998; Boots and Norman, 2000; Smith et al., 2008; but see also White and Grenfell, 1997; Lively, 2006); improving our understanding of how these processes influence coevolutionary dynamics is central to determining the generality of the RQH (May and Anderson, 1983; Gandon and Day, 2009; Lively, 2010b; Gokhale et al., 2013). The present study explores how parasitic reductions in host fecundity affect the propensity for epidemiological systems to exhibit coevolutionary cycling.

Parasites may indirectly reduce the fecundity of their host by limiting an individual's ability to find or compete for suitable mates (Hamilton and Zuk, 1982; Boyce, 1990; Hamilton and Poulin, 1997; chapter 7 herein). However, parasites that directly target host reproductive tissues or hormone pathways have a much greater inhibitory effect on host fecundity (Baudoin, 1975). This infection strategy, which has been observed across a variety of taxa (Blower and Roughgarden, 1989; Crews and Yoshino, 1989; Lockhart et al., 1996; Agrios, 1997; Hudson et al., 1998; Ebert et al., 2004; Sarasa et al., 2011), can lead to different pathological outcomes for the host (e.g. gigantism; Ebert et al., 2004) and evolutionary outcomes for the parasite

(e.g. runaway virulence (full castration); O'Keefe and Antonovics, 2002) that are not normally observed for other parasites. By definition, castrators target reproductive mechanisms, but it is important to note that many castrators also increase the mortality rate of their host. For example, several microparasites that infect the waterflea *Daphnia magna* are capable of both castrating and killing their host (e.g. *Pasteuria ramosa*; Ebert et al., 2000) and parasitic trematodes (*Microphallus* sp.) are known to affect the foraging behaviour of water snails (*Potamopyrgus antipodarum*), which increases predation by waterfowl (Levri, 1995). Theoretical models that have explicitly incorporated parasitic reductions in host fecundity have tended to focus on epidemiological outcomes or predictions for the evolution of virulence, rather than coevolutionary dynamics (May and Anderson, 1978; Boots and Norman, 2000; O'Keefe and Antonovics, 2002; Smith et al., 2008; Ashby and Gupta, 2013; chapter 6 herein; although see Lively, 2010b). Yet, much of the empirical work surrounding the issue of coevolutionary cycling and the RQH has involved castrators (Lively, 1987; Dybdahl and Lively, 1998; Decaestecker et al., 2007; King et al., 2009).

Castrators may target both sexual and asexual members of a population, but a crucial difference exists between the two. Asexuals do not need to find mates and so are unaffected by the fertility of other individuals, but sexual reproduction is only successful if both individuals in a pair are fertile. Hence, we should expect parasitic castration to have a more severe effect on sexual than asexual populations. Despite this notable difference between the two modes of reproduction in the presence of castrators, theoretical studies of the RQH have yet to account for this additional cost of sex (Lively, 2010b). Here, we analyse a simple epidemiological model to determine how parasitic reductions in host fecundity affect coevolutionary dynamics and assess whether sexual reproduction is likely to be maintained when castration restricts the availability of fertile mates.

5.2 Model description

We use a deterministic model, where parasites are haploid with two biallelic loci (00, 01, 10 and 11) and hosts are diploid with two biallelic loci per haplotype (haplotypes: 00, 01, 10 and 11). The infectivity of parasite genotype j on host genotype i, Q_{ij}, is equal to the proportion of host haplotypes that are matched by the parasite at both loci. Consequently, homozygotes are only susceptible to one type of parasite ($Q_{ij} = 1$) and heterozygotes are susceptible to two, but with lesser infectivity ($Q_{ij} = 1/2$ for each parasite), which ensures there is no underdominance for the host on average. This type of framework is normally referred to as 'matching alleles' (MA) specificity and has traditionally been used to represent self/non-self recognition mechanisms among animals (Hamilton, 1980; Penn and Potts, 1999). In essence, the MA framework is akin to a lock and key mechanism, where parasites are only able to specialise on narrow subsets of the host population, which typically leads to coevolutionary cycling in traditional population genetics models due to negative frequency-dependent selection (Hamilton, 1980; Hamilton et al., 1990; Agrawal and Lively, 2002; Agrawal, 2009).

We split the population into susceptible (S_i^g) and infectious (I_{ij}^g) classes, where subscripts correspond to different host (i) and parasite (j) genotypes and superscripts denote asexual ($g = A$) and hermaphroditic sexual ($g = S$) members of the population. We assume that the population mixes randomly and all parasites have a transmission rate of β, so that the force of infection for parasite j is $\lambda_j = \beta \sum_i (I_{ij}^S + I_{ij}^A)$. For simplicity, we also assume coinfection does not occur and hosts are unable to recover once infected. All hosts have a maximum per-capita birth rate of $r > 0$ and a natural mortality rate of $\mu > 0$, but infected individuals may also experience a disease-associated mortality rate of $\alpha \geq 0$ and a reduction in fertility, f ($0 \leq f \leq 1$). When $f = 1$ the parasite has no effect on fertility ($rf = r$), whereas $f = 0$ represents

full castration of the host ($rf = 0$). Newborn hosts of genotype i are produced at the following rates for asexual

$$b_i^A = r\left(S_i^A + f\sum_j I_{ij}^A\right) \tag{5.1}$$

and sexual

$$b_i^S = \frac{r(1+\delta_i)F_pF_qz}{2\sum_k F_k} \tag{5.2}$$

members of the population, respectively, where $\delta_i = 1$ if i is heterozygous and $\delta_i = 0$ if i is homozygous, z is the probability of a sexual individual finding at least one fertile partner and F_p and F_q are the densities of haplotypes p and q that together form host genotype i, modified by the fertility of hosts with these haplotypes. In other words, if following recombination (which occurs at a rate ρ) there are u copies of haplotype k in susceptible individuals and v copies in infected individuals, then $F_k = u + fv$. We set $b_i^S = 0$ if $\sum_k F_k = 0$. We assume that sexual members of the population can mate multiply to increase their chances of finding a fertile partner. Each individual engages in m independent mating attempts per unit time, but is limited to bearing at most r offspring during this period. Biologically, this could be interpreted as a restriction on the rate at which hermaphrodites can produce eggs (r per unit time), with no limitation on the rate at which sperm are produced. Thus, individuals can continue to fertilise other members of the population even if they have temporarily exhausted their supply of eggs. The probability of finding a fertile mate in any attempt is therefore $\sum_k F_k/2N_S$, where $N_S = \sum_i \left(S_i^S + \sum_j I_{ij}^S\right)$ is the size of the sexual population. It follows that z is given by:

$$z = 1 - \left(1 - \frac{\sum_k F_k}{2N_S}\right)^m \tag{5.3}$$

(see figure 5.1). Note that the birth rates of sexual and asexual individuals are equal

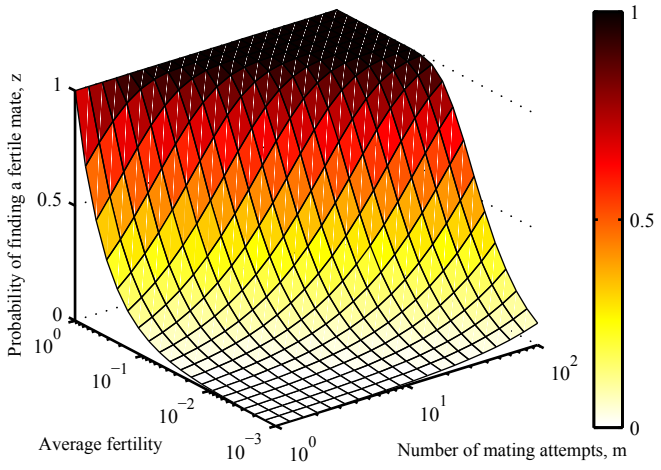

Figure 5.1: Probability of finding at least one fertile mate (z; equation 5.3) as a function of the number of mating attempts (m) and average fertility ($\sum_k F_k/2N_s$).

when parasites do not inhibit reproduction ($f = 1 \Rightarrow z = 1$, so $\sum_i b_i^S = \sum_i b_i^A = r$) and that $z \to 1$ as $m \to \infty$ provided some members of the sexual population have not been fully castrated ($\sum_k F_k > 0$). If parasites do inhibit reproduction ($f < 1$), then the overall birth rates of the two populations may still differ even when sexual individuals undergo an infinite number of mating attempts per unit time ($z = 1$), as certain haplotypes may be more prone to infection than others (e.g. $b_i^S = 0$ if $F_p = 0$ or $F_q = 0$, even if $z = 1$).

We implement our model using the following set of ordinary differential equations:

$$\begin{aligned} \frac{dS_i^g}{dt} &= b_i^g - S_i^g \left(\sum_j Q_{ij}\lambda_j + \mu \right) \\ \frac{dI_{ij}^g}{dt} &= S_i^g Q_{ij}\lambda_j - (\alpha + \mu) I_{ij}^g \end{aligned} \tag{5.4}$$

which is simply a generalisation of the standard SI epidemiological framework (see e.g. Anderson and May, 1991) to incorporate MA infection genetics. The infectious period is equal to $1/(\alpha+\mu)$ and the average number of secondary infections produced per infectious individual is given by $R_{EFF}^j = \beta \sum_i Q_{ij}(S_i^A + S_i^S)/(\alpha + \mu)$, which must exceed unity for the parasite j to increase in prevalence.

We take a similar approach to Otto and Nuismer (2004) in not including an explicit cost of sex in our model other than the cost of finding a mate as described above, as this allows us to determine the extent to which castrators disproportionately harm the sexual population by directly comparing sexual and asexual populations, all else being equal. In other words, if asexual reproduction is favoured under certain conditions in our model, then this outcome can be entirely attributed to differential effects of parasitism on sexual and asexual populations rather than to intrinsic costs of sex.

5.3 Results

5.3.1 Castration promotes persistent coevolutionary cycling

We begin by analysing the case where only homozygous asexual hosts are present, as this allows us to derive straightforward conditions under which coevolutionary cycling persists by simplifying the stability analysis of the system (homozygotes can only be infected by one type of parasite and asexual populations do not exhibit segregation and recombination). The reduced system can be represented by the following set of equations:

$$\begin{aligned} \frac{dS_i^A}{dt} &= r(S_i^A + f\sum_j \eta_{ij} I_{ij}^A) - S_i^A\left(\beta\sum_j \eta_{ij} I_{ij}^A + \mu\right) \\ \frac{dI_{ij}^A}{dt} &= \left(\beta\eta_{ij} S_i^A - \alpha - \mu\right) I_{ij}^A \end{aligned} \tag{5.5}$$

where $\eta_{ij} = 1$ if the parasite matches the (repeated) host haplotype at both loci and is otherwise equal to 0. The non-trivial fixed point ($S_i^{A^*}, I_{ij}^{A^*} > 0$, for i and j such that $\eta_{ij} = 1$) is given by:

$$\begin{aligned} S_i^{A^*} &= \frac{\alpha + \mu}{\beta} \\ I_{ij}^{A^*} &= \frac{(r-\mu)(\alpha+\mu)}{\beta(\alpha+\mu-rf)} \end{aligned} \tag{5.6}$$

which only exists provided $r > \mu$ (the birth rate is greater than the death rate) and $\alpha + \mu > rf$ (the parasite controls the population size). At this fixed point, the system has two repeated eigenvalues (Λ):

$$\Lambda = \frac{1}{2}\left(\frac{-rf(r-\mu)}{\alpha+\mu-rf} \pm \sqrt{\left(\frac{rf(r-\mu)}{\alpha+\mu-rf}\right)^2 - 4(r-\mu)(\alpha+\mu)}\right) \tag{5.7}$$

which must be complex for oscillations to occur. Oscillations will be damped if the real part of equation 7 is less than zero ($\mathrm{Re}(\Lambda) = -rf(r-\mu)/2(\alpha+\mu-rf)$), but will persist if $\mathrm{Re}(\Lambda) = 0$. Thus, only parasites that fully castrate their hosts ($f = 0$) are able to maintain coevolutionary cycling indefinitely. Although non-castrators ($f > 0$) still produce transient coevolutionary cycling provided $(rf/(\alpha+\mu-rf))^2 < 4(\alpha+\mu)/(r-\mu)$, these oscillations are damped because the supply of new susceptible hosts is less restricted, allowing the system to tend towards a stable equilibrium (e.g. figure 5.2b). Note that $\mathrm{Re}(\Lambda) \to 0$ as $f \to 0$, $\alpha \to \infty$ or $\mu \to r$, which means that partial reductions in host fertility or high mortality rates will slow the decay of oscillations, but will not allow these dynamics to persist indefinitely. The transmission rate, β, is absent from equation 5.7, so does not have any effect on the stability of the fixed point. The period of oscillations, T, is approximately equal to 2π multiplied by the reciprocal of the imaginary part of Λ. When $f = 0$, $T \approx 2\pi/\sqrt{\bar{r}\bar{\alpha}}$, where $\bar{r} = r - \mu$ and $\bar{\alpha} = \alpha + \mu$, so the period of oscillations will increase as $\bar{\alpha} \to 0$ or $\mu \to r$.

The reason why castrators cause this fundamental shift in the long-term behaviour of the system becomes clear if we set $f = 0$ and rewrite equation 5.5 as:

$$\begin{aligned} \frac{dS_i^A}{dt} &= S_i^A\left(\bar{r} - \beta\eta_{ij}I_{ij}^A\right) \\ \frac{dI_{ij}^A}{dt} &= -I_{ij}^A\left(\bar{\alpha} - \beta\eta_{ij}S_i^A\right) \end{aligned} \tag{5.8}$$

for i and j such that $\eta_{ij} = 1$. From a mathematical point of view, castrators create a perfect correspondence with a classical Lotka-Volterra predator-prey system, which

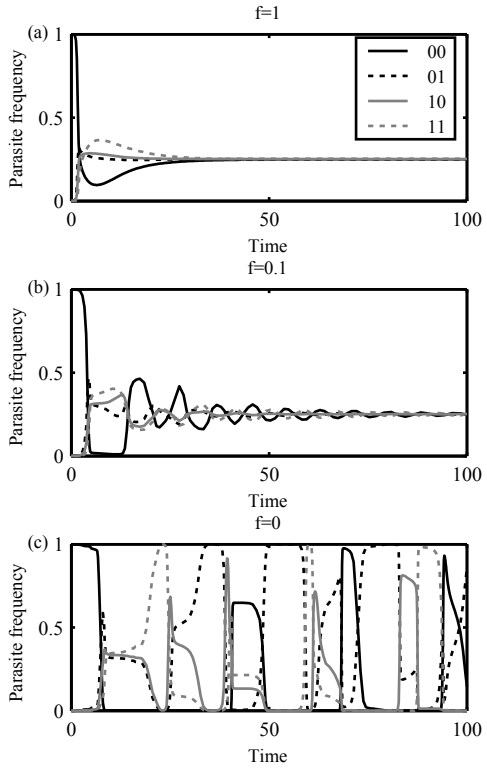

Figure 5.2: Example dynamics in an asexual population, with (a) $f = 1$ (no disease effects on reproduction); (b) $f = 0.1$ (partial reproduction inhibition); (c) $f = 0$ (full castration). Coevolutionary cycling is only able to persist in the presence of a castrator (equation 5.7). The dynamics for sexual and mixed populations are qualitatively similar. Parameters: $\alpha = 1$, $\beta = 0.05$, $\mu = 0.1$, $r = 1$.

is known to produce persistent oscillations (May, 1973). This is because infected individuals only contribute to the reproductive success of the parasite and cannot themselves reproduce in the presence of a castrator, which causes the host and parasite populations to suffer precipitous declines. In the context of the present study, hosts that are able to avoid contemporaneous parasites have the highest per-capita growth rate, but this will not last for long as parasites that target the most abundant host genotypes will be fittest, leading to coevolutionary cycling. The same principle allows non-castrators to produce transient oscillatory dynamics, but infected individuals still contribute to the reproductive success of the host ($f > 0$), which has a damping effect due to weaker selection against currently disadvantageous genotypes. Higher

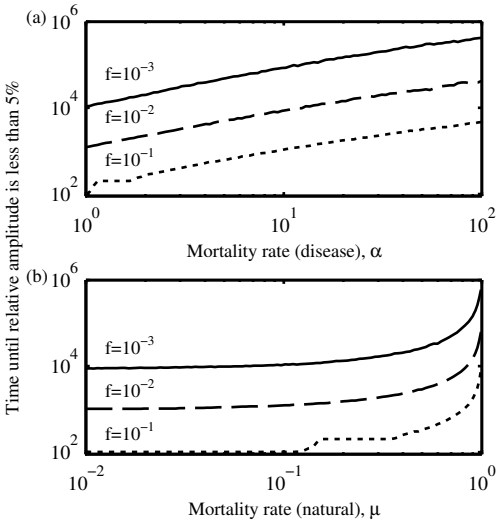

Figure 5.3: The effects of the disease-associated (α) and natural (μ) mortality rates on the decay of oscillations in sexual populations when $f = 10^{-3}$ (solid), $f = 10^{-2}$ (dashed) and $f = 10^{-1}$ (dotted). Simulations were terminated when fluctuations in the abundance of haplotypes were within 2.5% of the mean (i.e. relative amplitudes were less than 5%). Coevolutionary cycling persisted for longer for high values of α and μ, and low values of f. Higher recombination rates, ρ, have a relatively minor negative impact on the persistence of oscillations (log-log regression, $R^2 < 0.27$; table 5.1), but the effects of the transmission rate are negligible (log-log regression, $R^2 \leq 0.04$; table 5.1). Fixed parameters as specified in figure 5.2, with $\rho = 0.1$ and z held constant at 1 (i.e. $m \to \infty$).

mortality rates allow transient oscillations to persist for longer, but also reduce the infectious period, $1/(\alpha + \mu)$, limiting the size of an epidemic and hence the strength of selection on the host.

The stability analysis presented here only applies to asexual populations, but a numerical exploration of the parameter space indicates that patterns of coevolutionary cycling are qualitatively similar when hosts reproduce sexually, with full castration required for coevolutionary cycling to persist indefinitely. Oscillations tend to experience slower decay rates for small values of f, large values of α and as $\mu \to r$, as observed in asexual populations (figure 5.3). Higher recombination rates have a relatively minor negative impact on the persistence of oscillations, but the effects of

Persistence of oscillations, figure 5.3 (log-log)	$f = .001$	$f = .01$	$f = .1$
Disease-associated mortality rate, α	.934	.945	.971
Transmission rate, β^*	.002	.004	.039
Natural mortality rate, μ	.650	.665	.625
Recombination rate, ρ^*	.268	.203	.001
Period of oscillations, figure 5.4 (log-normal)		$f = 0$	
Disease-associated mortality rate, α		.909	
Transmission rate, β^*		$< .001$	
Natural mortality rate, μ		.282	
Recombination rate, ρ^*		.106	

Table 5.1: Coefficient of determination, R^2 (regression). *The recombination rate had a relatively minor negative impact on both the persistence and period of oscillations, but the effects of the transmission rate were negligible.

the transmission rate are negligible (log-log regression, $R_2 < 0.27$ and $R_2 \leq 0.04$, respectively; table 5.1).

As part of our numerical analysis, we also measured the average period of oscillations when coevolutionary cycling persists (i.e. when $f = 0$), as previous studies have shown that the frequency of oscillations can be crucial in determining whether sex is maintained (Barton, 1995; Gandon and Otto, 2007). We found that low disease-associated mortality rates and very high natural mortality rates are associated with lower frequency oscillations, as observed with asexual populations (figure 5.4). Again, the recombination rate has a very minor negative influence on the period of oscillations and the effects of the transmission rate are negligible (log-normal regression, $R^2 = 0.106$ and $R^2 =< 0.001$, respectively; table 5.1).

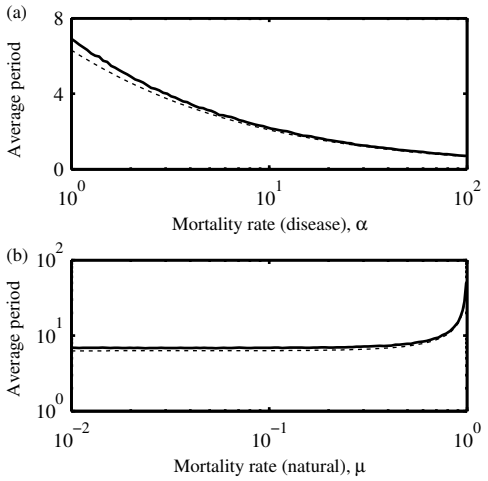

Figure 5.4: The effects of the disease-associated (α) and natural (μ) mortality rates on the period of oscillations in sexual (solid) populations when $f = 0$. Low values of α and high values of μ are associated with lower frequency oscillations. The recombination rate has a very minor negative influence on the period of oscillations (log-normal regression, $R^2 = 0.106$), but the effects of the transmission rate are negligible ($R^2 < 0.001$). Dotted lines show the period of oscillations in homozygous asexual populations. Parameters as specified in figure 5.3.

5.3.2 Castration disproportionately harms sexual populations

The previous section demonstrated that persistent coevolutionary cycling is only likely to be maintained in the presence of a castrating parasite. This finding is particularly relevant to the maintenance of sex according to the RQH, which requires a continually changing environment for sexual populations to outcompete their asexual counterparts. A naive interpretation of our model would therefore be that castrating parasites are more likely to select for sex than non-castrators, as only they are able to induce persistent coevolutionary cycling. However, this neglects a fundamental difference between sexual and asexual reproduction; a sexually-reproducing female is only able to produce offspring if both she and her partner are fertile, whereas asexual individuals are unaffected by the fertility of other members of the population. We capture this asymmetry in reproductive success using the term z in equation 5.3,

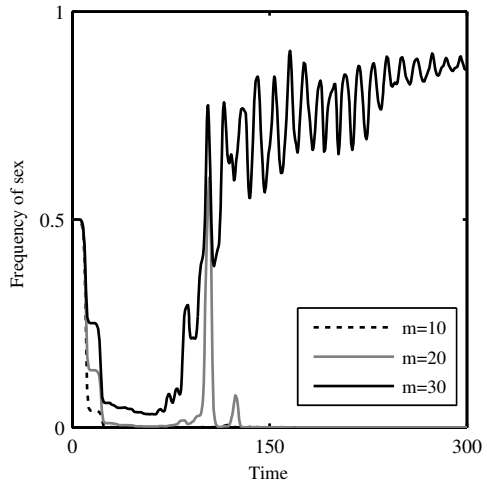

Figure 5.5: Example dynamics showing the frequency of sexually-reproducing individuals in the presence of castrating parasites ($f = 0$), for different values of m (mating attempts): $m = 10$ (dotted); (b) $m = 20$ (grey); (c) $m = 30$ (black). Although castrators induce coevolutionary cycling (figure 5.2), they also reduce the chances of sexually-reproducing individuals finding a fertile mate (equation 5.3). Mating multiply can counteract this implicit cost of sex, but may require a large number of sexual partners. Fixed parameters as specified in figure 5.2, with $\rho = 0.1$.

which gives the probability that a sexually-reproducing individual will find a fertile mate, given m mating attempts per unit time. Figure 5.1 shows how z depends on both m and the average fertility of potential mates. Even when average fertility is reasonably high, a large number of mating attempts is required before the probability of finding a fertile mate approaches certainty.

Figure 5.5 demonstrates the typical behaviour of our model in the presence of castrating parasites. When the number of mating attempts for sexuals is relatively low, asexual reproduction tends to dominate. Yet, if the number of mating attempts is sufficiently high, the implicit costs of castration are less severe, so sexual members of the population are not driven extinct. Instead, better avoidance of contemporaneous parasites by sexual individuals leads to the suppression of asexual reproduction. Hence, the number of mating attempts is crucial in determining whether sexual repro-

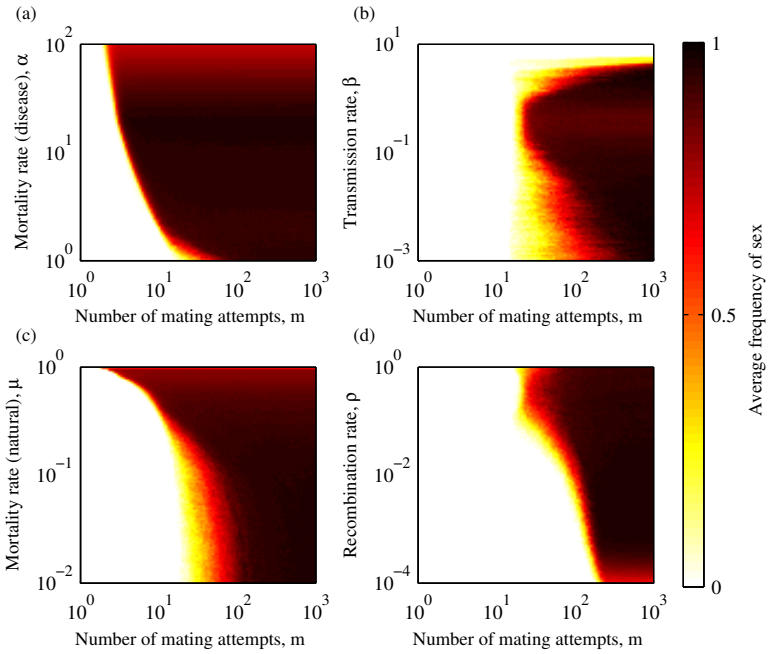

Figure 5.6: The effects of the number of mating attempts (m) and other model parameters on the maintenance of sex in the presence of castrating parasites ($f = 0$). Heatmaps show the average frequency of sex at equilibrium over 100 randomly chosen sets of initial conditions. Asexual populations dominate in the white areas and sex dominates in the dark areas. Both populations may coexist in the boundary between these two regions, but evolutionary outcomes are often highly dependent on initial conditions (see figure 5.7 for standard deviations), with sexual populations more likely to displace asexual populations in red areas (and vice versa in yellow areas). Fixed parameters as specified in figure 5.2, with $\rho = 0.1$ in (a)-(c).

duction is maintained. Other model parameters were also found to have an impact on the maintenance of sex, with high mortality (disease and natural), intermediate transmission, and high recombination rates increasing selection for sex (figures 5.6-5.7). Unlike previous studies (Barton, 1995; Peters and Lively, 1999; Gandon and Otto, 2007), we found that the period of oscillations had no discernible impact on the maintenance of sexual reproduction (figures 5.4, 5.6a, c). This is most likely attributable to ecological feedbacks in our model, as higher mortality rates reduce the average proportion of the population that is infected (equation 5.6), making it easier to find

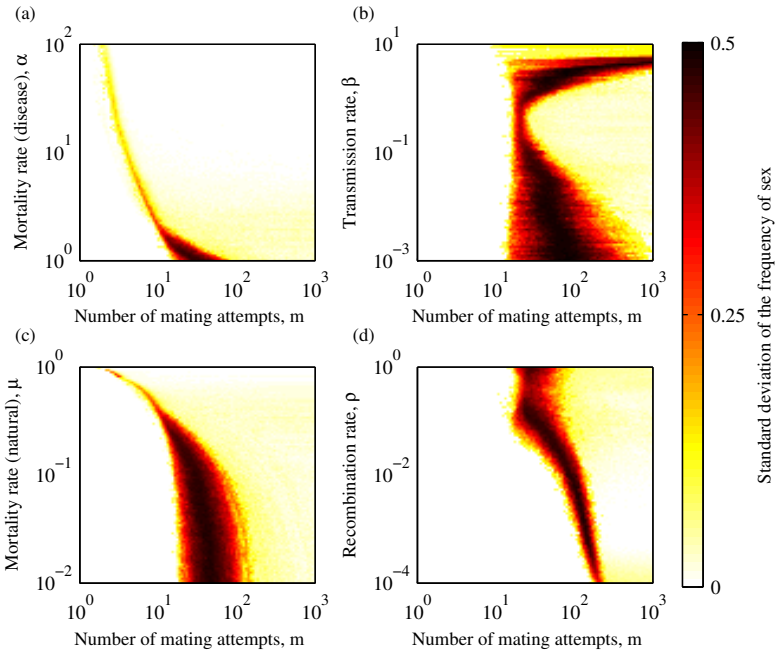

Figure 5.7: The effects of the number of mating attempts (m) and other model parameters on the maintenance of sex in the presence of castrating parasites ($f = 0$). Heatmaps show the standard deviation of the frequency of sex at equilibrium over 100 randomly chosen sets of initial conditions (see figure 5.6 for average values). Dark regions correspond to evolutionary outcomes that are highly dependent on initial conditions. Fixed parameters as specified in figure 5.2, with $\rho = 0.1$ in (a)-(c).

a fertile partner (z increases). Thus, it is the availability of fertile mates, rather than the period of oscillations that appears to be the dominant force in our model. High rates of recombination increase the likelihood that offspring will avoid infection by contemporaneous parasites, which reduces the number of mating attempts required for sexual populations to dominate (figure 5.6d), but the effects are less profound than for changes in mortality rates. The influence of the transmission rate on the maintenance of sex is less clear (figure 5.6b), but it appears that extreme values of β lead to high amplitude fluctuations in the proportion of the population that is infected, which periodically restricts the availability of fertile mates and consequently selects against sex.

5.4 Discussion

Fluctuations in gene frequencies caused by host-parasite interactions are thought to play a pivotal role in several biological phenomena, so it is important to understand conditions that promote or inhibit these dynamics. Using a deterministic epidemiological model with overlapping generations, we have demonstrated that coevolutionary cycling will only persist in the presence of parasites that fully castrate their hosts. Even extremely lethal parasites are unable to sustain cycles indefinitely in the absence of castration (although they do allow oscillations to persist for longer), as high disease-associated mortality rates cause a reduction in the average infectious period, which limits the size of an epidemic and consequently the overall strength of selection. Castration does not reduce the infectious period so a larger proportion of the population can be infected during an epidemic, leading to greater population crashes and inducing stronger selection for resistance, which are both more conducive to persistent oscillatory dynamics (May and Anderson, 1978; Otto and Nuismer, 2004).

Although castrators promote coevolutionary cycling, they also have a disproportionately harmful effect on sexual populations, as both members of a sexual partnership must be fertile for offspring to be produced. By contrast, the production of each asexual offspring is only dependent on the fertility of a single female. Taking this fundamental difference between the two modes of reproduction into account, we found that sex will only outcompete asexual reproduction provided multiple mating is very common. We did not include an explicit cost of sex in our model apart from the cost of finding a fertile mate, so that we could ascertain the severity of implicit costs of sex caused by castration, all else being equal. Still, we struggled to find conditions where sex is likely to outcompete asexual reproduction, either due to these implicit costs or a lack of coevolutionary cycling; thus conditions for sex to dominate could be even stricter than described herein.

A large number of mathematical models have been developed since the RQH was first proposed as a possible mechanism for the maintenance of sex (Hamilton, 1980; May and Anderson, 1983; Hamilton et al., 1990; Doebeli, 1996; Otto and Nuismer, 2004; Kouyos et al., 2007; Agrawal, 2009; Lively, 2010b; Gokhale et al., 2013), but our results differ from previous studies in a number of important ways. First, the MA framework has previously been shown to persistently oscillate under a broad range of conditions, whereas our model only maintains coevolutionary cycling in the presence of castrators. This difference is partially due to our inclusion of density-dependent processes, which are typically omitted from traditional population genetics models (Hamilton, 1980; Hamilton et al., 1990; Agrawal and Lively, 2002; Agrawal, 2009). Cycling is easier to maintain in the absence of density-dependence, as an increase in the prevalence of one allele must be exactly offset by decreases in alleles with below average fitness. In our model, castrators cause sudden population crashes for currently disadvantageous haplotypes, which then take a relatively long time to recover to sufficient levels for a new epidemic to occur ($R^j_{EFF} > 1$). Parasites that only partially reduce host fecundity have less of an impact on the supply of new susceptible hosts, so oscillations decay due to this ecological feedback. Crucially, an increase in the absolute prevalence of one allele does not need to be offset by changes in another, so cycling is not as readily maintained. The other reason that coevolutionary cycling is more common in previous models is due to their tendency to use discrete rather than overlapping generations (Hamilton, 1980; Hamilton et al., 1990; Doebeli, 1996; Otto and Nuismer, 2004; Agrawal, 2009; Lively, 2010b). Discretised generations synchronise various ecological and epidemiological processes, leading to persistent oscillations that would otherwise be damped if generations were to overlap. Had our model employed discrete generations, oscillations would have persisted over a much broader set of conditions (May, 1973). It is vital to establish whether oscillations in theoretical models are being propagated by fundamental aspects of a

host-parasite relationship or are simply arising due to the method of implementation. This question was explored by Kouyos et al. (2007), who found that moving from discrete to continuous time led to the loss of persistent oscillations of linkage disequilibria. Models that lack ecological feedbacks or overlapping generations may therefore be overestimating conditions that favour coevolutionary cycling, and hence sex. This point is particularly important, as much of the debate over the generality of the RQH may be fuelled by fundamental differences in modelling approaches (i.e. presence/lack of ecological feedbacks, continuous versus discrete time, diploid versus haploid hosts (see below)).

Second, Lively (2010b) has shown that sex can be maintained in the presence of a castrating parasite with or without coevolutionary cycling, but our model will only select for sex if cycling occurs. The contrasting outcomes are due to different initial conditions, as here we have assumed that sexual and asexual populations start with all genotypes present, but Lively (2010b) challenged established sexual populations with invasion by a single clonal lineage, which increased the benefits of sex. Our model has the additional requirement that sex will only be maintained in the presence of a castrator if individuals mate with a sufficient number of partners, as we explicitly account for a limited number of sexual contacts per individual rather than basing birth rates on the average fertility of the sexual population. Lively (2010b) took the latter approach, which again increases selection for sex.

Third, sexual populations are most successful in our model when recombination rates are high, but Otto and Nuismer (2004) have suggested that this is only likely to be true when selection is strong. However, the two findings are not directly comparable as Otto and Nuismer (2004) used a population genetics approach that did not include ecological feedbacks. Additionally, Otto and Nuismer (2004) modelled haploid hosts (as have most studies), whereas we have used diploids. Selection for sex will depend on the effects of both segregation and recombination, but the former

cannot be accounted for in haploids (Agrawal, 2009). It is important to note that our model assumes there is no underdominance or overdominance, but previous studies have shown that either can select against sex as segregation breaks up beneficial combinations of alleles (Otto, 2003; Agrawal and Otto, 2006). In addition, the period of oscillations appeared to have little impact on the maintenance of sex in our model, but others have shown that the period of fluctuating epistasis can be crucial, with sex only likely to be maintained for a narrow range of frequencies (Barton, 1995; Peters and Lively, 1999; Gandon and Otto, 2007). Again, this discrepancy is attributable to ecological feedbacks that were absent in other models; sex is maintained in our model even when the period of oscillations is very high, because these conditions are associated with a relatively low prevalence of infection, which reduces the cost of sex by increasing the probability of finding a fertile mate (figures 5.4, 5.6).

Although we have made our model as general as possible, it is likely that real host-parasite systems may differ in a number of aspects, some of which may alter selection for sex. For example, any synchronisation in ecological or epidemiological processes (e.g. annual populations or seasonal epidemics) could increase the likelihood of coevolutionary cycling being maintained indefinitely, as could other environmental effects (Wolinska and King, 2009). It is also possible that stochasticity in natural populations may be sufficient to counteract the damping effects caused by deterministic attractors or may lead to random linkage disequilibria (i.e. the Hill-Robertson effect), allowing coevolutionary cycling (and potentially sex) to persist under a wider range of conditions than observed here (Rohani et al., 1999; Kouyos et al., 2007). However, it has recently been demonstrated that stochasticity may have the opposite effect, increasing the likelihood that certain alleles will reach fixation rather than persistently oscillate (Gokhale et al., 2013). Finally, sexual populations may have access to regions of genotype space that are unavailable to asexuals, which may allow them to survive invasion by clonal lineages, even in the absence of coevolutionary cycling

(Lively, 2010b).

Some of these caveats may help to explain why empirical observations of certain host-parasite systems provide support for the RQH, most notably in the freshwater snail *Potamopyrgus antipodarum* (Lively, 1987; King et al., 2009). These snails are no doubt host to a variety of parasites, but it is the presence or absence of the parasitic trematode *Microphallus* that determines whether sex is maintained. Crucially, this parasite differs from others in that it fully castrates its host; our model suggests that it is this feature of the host-parasite relationship that induces persistent coevolutionary cycling, which is a prerequisite for sex to be maintained according to the RQH. These snails are also known to exhibit multiple mating (Soper et al., 2012), which will reduce the disproportional impact of castration on sexual populations, as demonstrated by our model. Note however that multiple mating could incur an opportunity cost, as individuals will have less time and energy available for foraging. Other factors such as spatial structure (offspring are more likely to experience the same parasites as their hosts if dispersal is limited), the complex life cycle of the parasite (synchronisation of ecological processes) or the accumulation of deleterious mutations were not captured by our model, but could conceivably contribute to the maintenance of sex (Howard and Lively, 1994; Keeling and Rand, 1995; West et al., 1999).

We are aware of only one other experimental system where recombination (sex) has been selected for in the presence of parasites, but this involved a highly virulent pathogen that killed its host within 24 hours (*Caenorhabditis elegans-Serratia marcescens*; Morran et al., 2011). We have shown that high disease-associated mortality rates allow coevolutionary cycling to persist for longer (figure 5.3), but our findings suggest that selection for recombination may not continue indefinitely in this system. Empirical tests of the RQH are generally difficult, as we are limited by the need for comparable sexual and asexual populations. Similarly, coevolutionary cycling is often hard to detect in real host-parasite systems, but has been observed

among invertebrates (Dybdahl and Lively, 1998; Decaestecker et al., 2007). It is interesting to note that the bacterium *Pasteuria Ramosa*, which undergoes coevolutionary cycling with its host *Daphnia magna*, is also a castrator (Decaestecker et al., 2007). In the absence of evidence from a more diverse set of systems, it is difficult to draw conclusions about the generality of the RQH from these observations alone. In a review of the RQH in the context of plant-parasite interactions, Clay and Kover (1996) came to the conclusion that

> "parasites that kill or sterilize [castrate] their hosts are the most likely players in the coevolutionary scenario envisioned by the RQH. Many lesion-forming parasites are unlikely to exert selection on hosts of a magnitude strong enough to generate cycles of gene frequencies."

In light of our results, it appears that parasitic castrators are a much more likely candidate for producing these dynamics, but the implicit costs imposed on sexually-reproducing individuals means that asexual populations may still win the evolutionary battle.

Chapter 6

Sexually transmitted infections in polygamous mating systems

Ben Ashby[1] and Sunetra Gupta[1]

Abstract

Sexually transmitted infections (STIs) are often associated with chronic diseases and can have severe impacts on host reproductive success. For airborne or socially-transmitted pathogens, patterns of contact by which the infection spreads tend to be dispersed and each contact may be of very short duration. By contrast, the transmission pathways for STIs are usually characterised by repeated contacts with a small subset of the population. Here we review how heterogeneity in sexual contact patterns can influence epidemiological dynamics, and present a simple model of polygyny/polyandry to illustrate the impact of biased mating systems on disease incidence and pathogen virulence.

6.1 Introduction

Evidence from anthropological and ethological studies suggests that there is much heterogeneity in sexual behaviour of humans and animals (Dixson, 1998; Sherman et al., 1988; Jennions and Petrie, 2000; Liljeros et al., 2001) both in rates of sexual activity

[1]Department of Zoology, University of Oxford, Oxford, UK

and in patterns of sexual contact. Polygynous and polyandrous mating systems are particular examples, where one sex tends to have a much higher variance in partner acquisition rate compared to the other sex. It is well-established that the structure of a mating system can have a profound influence on genetic diversity (Boomsma et al., 1999) and the evolution of sexually-selected traits (Emlen and Oring, 1977); here we discuss how such heterogeneities can influence epidemiological dynamics of sexually transmitted infections (STIs) and the evolution of associated pathogens.

We begin by reviewing how host heterogeneity in sexual behaviour can influence the epidemiological dynamics of STIs, mainly in the context of human diseases such as HIV-1 and gonorrhoea. We then introduce an individual based model for biased (polygynous and polyandrous) mating systems, where movement is based on perception of reproductive failure. In line with the results of Thrall et al. (2000), we observe that the polygamous sex exhibits much lower levels of infection than the monogamous sex and the difference tends to increase with greater variance in attractiveness; in addition, we show that the difference between the polygynous and polyandrous scenarios depends on the probability of sterility. Finally, we present an example of how the evolution of pathogen virulence can be explored within this framework by introducing a simple dichotomy in the trade-off between transmissibility and duration of infection. Within our system the less virulent pathogen tends to be favoured for high degrees of polygamy, demonstrating a clear link between mating patterns and pathogen evolution.

6.2 Influence of sexual contact patterns on epidemiology of STIs

From an epidemiological point of view, the transmission dynamics of STIs are fundamentally different to those of many other infectious diseases. First, sexual contact

rates are usually invariant to population size, which means that there is no critical population density required for a typical STI to persist. By contrast, the rate at which non-STIs spread is often dependent on the density of the host population (Anderson and May, 1991). Second, there is often considerable variation in sexual behaviour both within (Liljeros et al., 2001) and between (Blanchard, 2002) populations. Highly active members of the population (e.g. sex workers in human populations, alpha males in animal populations) will generally be at much greater risk of receiving and transmitting an infection than monogamous couples and so contribute disproportionately to the spread of disease as well as representing important targets for disease control.

In order to establish how heterogeneity in sexual behaviour can alter epidemiological dynamics, it is first useful to define the basic reproductive number, R_0, of an infectious agent in a well-mixed population:

$$R_0 = \beta D n \tag{6.1}$$

where β is the probability of transmission per contact, D is the average infectious period and n is the average number of contacts (see Anderson and May, 1991 for a more detailed discussion of R_0). The basic reproductive number is essentially the average number of secondary infections that a single infectious individual will produce in an entirely susceptible population. Hence the infection will tend to spread if $R_0 > 1$, but will go extinct if $R_0 < 1$. This formulation of R_0 is based on an idealised, randomly-mixing homogeneous population, but real mixing patterns are likely to be more complex due to spatial constraints and variations in host behaviour. If we imagine a continuum with well-mixed and highly-structured populations at the extremes, then most real populations will fall somewhere between the two. Note that the position of a population on this continuum is dependent on the transmission pathways of a particular infection; a population may be relatively well-mixed in terms of social contacts, but might demonstrate a high degree of heterogeneity in sexual

mixing patterns.

In general, casual contacts between humans tend to be ephemeral and non-repetitive, whereas sexual contacts are more stable (Edmunds et al., 2006). In addition, variation in close contact rates is likely to be much smaller than variation in sexual partner acquisition rates. For example, Mossong et al. (2008) found that adolescents had just over twice the number of close contacts than the elderly, but studies of sexual mixing patterns generally find power-law distributions in partner acquisition rates, sometimes ranging over three orders of magnitude (Woolhouse et al., 1997; Liljeros et al., 2001). Power-law distributions are also likely to be applicable to a variety of animal mating systems, particularly where a few members of one sex are dominant (i.e. polyandry or polygyny). Hence, the above formulation of R_0 may be a reasonably good indicator of epidemic spread for infections transmitted by close contact, but is likely to be a poor approximation for STIs. In addition, sexual contacts tend to be much less frequent than social contacts, lowering the value of R_0. Hence, one explanation as to why many STIs are associated with chronic, asymptomatic diseases is that this increases the value of D to compensate for lower contact rates.

Sexual transmission can be considered part of a much broader class of models with heterogeneous contact rates, usually referred to as 'super-spreader' models, where a few members of the population have a disproportionately large effect on disease spread (Anderson and May, 1991; Woolhouse et al., 1997). For super-spreader models, we can incorporate this heterogeneity into the formula for R_0 by compartmentalising the population according to contact rates, so that N_i is the proportion of the population that acquires i contacts per unit time (Hethcote and Yorke, 1984; Anderson et al., 1986). Retaining the assumption that mixing is random, this allows us to calculate an effective contact rate, c, over the distribution:

$$c = \frac{\sum_i i^2 N_i}{\sum_i i N_i} = m + \frac{\sigma^2}{m} \tag{6.2}$$

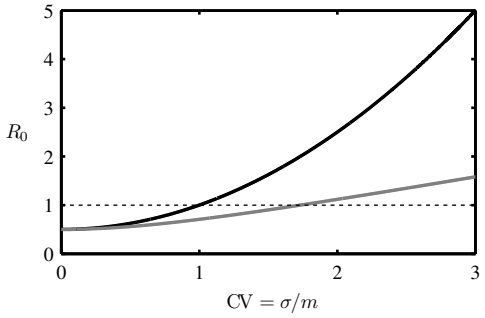

Figure 6.1: The relationship between the coefficient of variation, $CV = \sigma/m$, of the partner acquisition rate distribution and the basic reproductive number, R_0, for fixed values of β, D and m. ($\beta = 0.1$, $D = 5$yr and $m = 1\text{yr}^{-1}$, so that $\beta Dm = 0.5$). The black curve shows the relationship when there is no distinction between the sexes $R_0 = \beta Dm(1 + CV^2))$ and the grey curve shows the relationship when variation only occurs in one sex $R_0 = \beta Dm(1 + CV^2))^{\frac{1}{2}}$. The dotted line corresponds to the threshold for an epidemic ($R_0 > 1$). Adapted from May et al. (2001).

where m and σ^2 are the mean and variance in contact rates, respectively (Anderson et al., 1986). The formula for the basic reproductive number now becomes $R_0 = \beta Dc$; clearly, any heterogeneity in contact rates will increase the value of R_0 and hence the initial growth rate of the epidemic (figure 6.1).

In the context of sexual transmission, the second formulation of R_0 applies to a homosexual population, but it can be readily generalised for a heterosexual population by separating the effective partner acquisition (i.e. contact) rate into male (c_m) and female (c_f) components. If we also assume that there are differential transmission rates across the sexes (as is common with many STIs), then our equation for R_0 becomes:

$$R_0 = D\sqrt{\beta_m \beta_f c_m c_f} \tag{6.3}$$

where β_m is the transmission rate from males to females and β_f is the transmission rate from females to males (Anderson and May, 1991; Anderson et al., 1991). If $c_m = c_f$ then R_0 will asymptote towards quadratic growth with the coefficient of variation

($CV = \sigma/m$), but if variation is limited to one sex then R_0 will tend towards linear growth. Changes in the variance will be most significant when R_0 is close to unity (the epidemic threshold), as relatively small changes in the size or the behaviour of the core group can determine whether an epidemic will occur (figure 6.1).

The effective partner acquisition rates can also be used to estimate the ratio of cases in males (C_m) to females (C_f) during the early stages of the epidemic:

$$C_m/C_f = \sqrt{\beta_f c_f/\beta_m c_m} \tag{6.4}$$

(see Anderson and May, 1991, §11.3.9 for a more detailed discussion; also May et al., 2001). This work was originally motivated by the spread of HIV in Africa, but the principles can be applied to other populations that exhibit host heterogeneity. For example, the ratio C_m/C_f suggests that the dominant sex in a biased mating system (e.g. males in polygynous systems) will tend to exhibit lower than average levels of infection, although this could be counterbalanced by differences in transmission probabilities. Indeed, it is thought that $C_m/C_f \approx 1$ in many parts of Africa because the partner acquisition rates ($c_f > c_m$) are more or less balanced by differences in transmission rates ($\beta_f < \beta_m$) (Anderson and May, 1991; Anderson et al., 1991). Note that even if $C_m/C_f \approx 1$, the distribution of infection will still be biased towards those members of the population who are highly active.

Further complications will arise if the population does not mix homogeneously, for example where people tend to show a preference for mixing with similar individuals (assortativity). Mixing patterns have been found to vary considerably between human populations, ranging from highly assortative (Ghani and Garnett, 1998), to highly disassortative (i.e. showing preference for dissimilar individuals; Haraldsdottir et al., 1992) mixing. The degree of assortative mixing may also vary within a population: for example, Wylie and Jolly (2001) found that assortative mixing was common in linear components of a sexual contact network, but disassortative mixing was common

in radial components.

In order to model heterogeneous mixing, we can group individuals according to their level of sexual activity (i.e. partner acquisition rate) and describe interactions between groups using a 'mixing-matrix' (Hethcote and Yorke, 1984; Anderson et al., 1986; Jacquez et al., 1988; Gupta et al., 1989; Anderson et al., 1991). A simple mixing-matrix for a population split into high (H) and low (L) activity groups would be:

$$\mathbf{p}_{ij} = \begin{pmatrix} p_{LL} & p_{LH} \\ \\ p_{HL} & p_{HH} \end{pmatrix} \tag{6.5}$$

where p_{ij} is the proportion of sexual contacts that individuals from group i make with members of group j. For completely assortative mixing, $\mathbf{p}_{ij}$ is equal to the identity matrix ($p_{ii} = 1, p_{i,j} = 0$ for $i \neq j$). There is usually no single disassortative extreme however, as disassortativity is maximised whenever the elements of the main diagonal of $\mathbf{p}_{ij}$ are minimised (Gupta et al., 1989). For a given mixing matrix, we can measure the degree of assortativity, Q, in the population as:

$$Q = \frac{1}{g-1} \sum_{i=1}^{g} \lambda_i - 1 \tag{6.6}$$

where g is the number of activity groups and λ_i are the eigenvalues of $\mathbf{p}_{ij}$. Gupta et al. (1989) found that highly assortative mixing ($Q \approx 1$) tends to lead to more rapid epidemic growth and can produce multiple peaks in disease incidence. By contrast, highly disassortative mixing ($Q \approx -1/(g-1)$) is generally associated with slower epidemic growth, but will typically produce higher peaks in disease incidence (see figure 6.2). This method highlights the importance of host heterogeneity in the spread of STIs and suggests that targeting control measures at the core group is optimal, although the efficacy of such procedures will depend on the size of this group and the

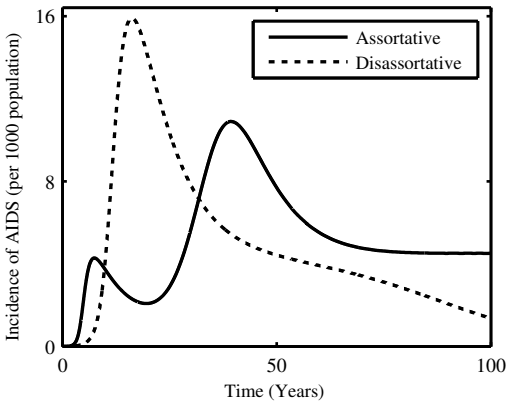

Figure 6.2: Incidence of AIDS as a proportion of population size in populations that exhibit highly assortative (solid curve) and highly disassortative (dotted curve) mixing. Highly assortative mixing tends to lead to rapid growth during the early stages of an epidemic and can produce multiple peaks in disease incidence. Highly disassortative mixing is usually characterised by slower initial growth, but a higher peak in the incidence of AIDS. The model is adapted from Gupta et al. (1989) (see appendix C).

degree of assortative mixing in the population.

While this approach is a useful way of capturing host heterogeneity, it cannot capture some of the complex interactions found in real populations that are imposed by other factors than level of sexual activity. Such mean-field approaches assume that sexual activity classes are well-mixed so that if an infectious individual mixes with a particular activity class, then all members of that class will have an equally increased risk of infection. In reality, the risk of infection will be limited to those who have sexual contact with the infectious individual rather than the entire activity class. An alternative method is to use a sexual contact network (SCN) which captures heterogeneity at the level of individuals and provides a means of replicating more realistic transmission pathways. This approach is particularly well suited to STIs, as transmission pathways are usually much more clearly defined (i.e. sexual contact) than for non-STIs. However, there are many problems associated with collecting data on real SCNs, including biases in reporting and difficulties with linking up components

in a larger network (Ghani and Garnett, 1998; Ghani et al., 1998), although some attempts have been made for small populations (Haraldsdottir et al., 1992; Woodhouse et al., 1994; Wylie and Jolly, 2001; Potterat et al., 2002; Bearman et al., 2004; De et al., 2004).

Mean-field models and SCNs may show good agreement over the main part of an epidemic, but fundamental differences in structure are likely to have significant consequences for the spread of infection during the early stages (Keeling, 2005). Keeling (2005) showed that in a general SCN, the probability that an index case i will fail to pass on the infection to any of their n_i contacts is:

$$P_{SCN}(i) = \left(1 - \frac{R_0(i)}{n_i}\right)^{n_i} \tag{6.7}$$

where $R_0(i) = \min(\beta D n_i, n_i)$ is the expected number of secondary infections to be caused by the index case. The corresponding probability of extinction after the first generation using a mean-field approximation is:

$$P_{MF}(i) = \exp(-R_0(i)) \tag{6.8}$$

which satisfies $P_{MF}(i) > P_{SCN}(i)$. Hence the probability of extinction during the first generation is always higher in a mean-field approximation than it is in the corresponding SCN, but as the number of contacts increases the two values converge ($\lim_{n_i \to \infty} P_{SCN}(i) = P_{MF}(i)$). Averaging these values over the entire population gives the probability that a randomly introduced infection will die out after the first generation. Note that in real populations, an index case is probably more likely to occur among highly active individuals, which will decrease the probability of extinction during the first generation.

One advantage that SCNs have over mean-field approximations is their ability to capture long-term partnerships that are commonly found in many human and ani-

mal populations. In particular, serially monogamous partnerships (common among birds as well as humans) cannot be modelled using traditional mean-field approaches. Computer-generated contact networks can be used to recreate mixing patterns observed in real populations (Bearman et al., 2004), by connecting individuals (nodes) to other members of the population preferentially, based on factors such as proximity, cluster size or assortativity. Studies of simulated epidemics on SCNs have revealed that concurrent partnerships are crucially important to the spread of many STIs (Watts and May, 1992; Morris and Kretzschmar, 1997; Eames and Keeling, 2004). For example, Morris and Kretzschmar (1997) demonstrated that the size of an epidemic grows exponentially with the relative number of concurrent partnerships in a population. Reducing the number of concurrent partnerships in a population is therefore likely to be an effective mechanism of disease control.

6.3 Exploring the role of mating system structure on epidemiological dynamics

We now introduce a simple SCN model to illustrate how epidemiological dynamics of STIs can be influenced by biased (i.e. polygynous or polyandrous) mating systems. Our model is similar to that of Thrall et al. (2000), which was used to explore how disease prevalence is affected in a general biased mating system with random movement between mating groups. The authors found that disease prevalence in the two sexes tends to diverge as variance in mating success increases and that less attractive members of the population may have higher lifetime reproductive success in the presence of a sterilising STI. We build on this study by varying the probability that an infection will cause host sterility and by basing movement decisions between mating groups on an individual's perception of reproductive failure. This simple 'stay-or-stray' decision introduces behavioural differences between polygynous

and polyandrous mating systems, as females are generally better placed to infer their reproductive success. Given that a range of complex mate choice behaviour has been observed, including the avoidance of parasitism, inbreeding and harassment (see Jennions and Petrie, 1997 for a review of mate choice behaviour), it seems reasonable that a simple binary decision of prior reproductive success or failure could influence mate choice. In fact, a meta-analysis of mate fidelity among 35 species of monogamous birds found that divorce rates were significantly higher among unsuccessful than successful pairs (Dubois and Cézilly, 2002), providing strong evidence that prior reproductive failure can reduce mate fidelity.

We consider a population of constant size, composed of N_m males and N_f females, where one sex is polygamous and the other is serially monogamous. We follow Thrall et al. (2000) in assigning members of the serially monogamous sex to mating groups consisting of a single member of the polygamous sex. Each polygamous individual, i, is assigned a fixed level of attractiveness, $a(i)$, according to a power-law distribution with shape-parameter α:

$$a(i) = \frac{i^{-\alpha}}{\sum_{k=1}^{N_m} k^{-\alpha}} \tag{6.9}$$

with $\sum_{k=1}^{N_m} a(i) = 1$. Members of the polygamous sex are then assigned non-overlapping line segments, $L(i)$, with lengths equal to their attractiveness:

$$L(i) = \begin{cases} [0, a(1)] & \text{if } i = 1 \\ \left[\sum_{k=1}^{i-1} a(k), \sum_{k=1}^{i} a(k)\right] & \text{otherwise} \end{cases} \tag{6.10}$$

For each serially monogamous individual j, a random number, $r(j) \in (0, 1)$, is then generated. The connections between males and females are given by the adjacency matrix $\mathbf{A}_{ij}$, where $A_{ij} = 1$ if $r(j) \in L(i)$ and is 0 otherwise. For small values of α there is little variation in attractiveness and so the network approaches serial monogamy for

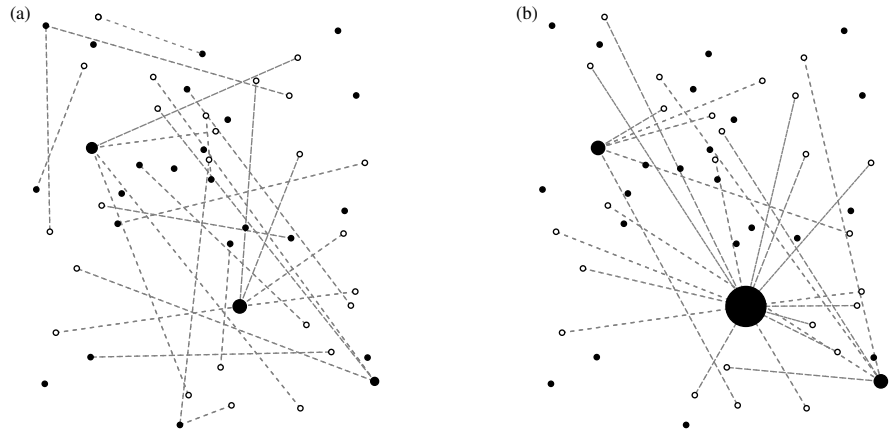

Figure 6.3: Example sexual contact networks (SCNs) in a biased mating system. Sexual contacts are indicated by dotted lines connecting the serially monogamous sex (white nodes) to the polygamous sex (black nodes). Members of the polygamous sex attract mates based on their relative 'attractiveness', which is based on a power-law distribution (equation 6.9) with (a) $\alpha = 1$ and (b) $\alpha = 2$. Lower values of α produce more balanced SCNs, with less variation in the number of partners (a). Higher values of α cause a few members of the population to dominate the network and leave many members of the polygamous sex without partners (b).

both sexes, although some concurrent partnerships may still occur (figure 6.3a). For large values of α, the network is dominated by a single polygamous individual who is connected to a large proportion of the serially monogamous population (figure 6.3b). We refer to these scenarios as having low and high degrees of polygamy respectively.

A randomly chosen member of the monogamous sex initiates the epidemic, after which susceptible individuals are infected with probability βI, where β is the probability of transmission per contact and I is the total number of infectious contacts for a given individual. For members of the polygamous sex, the number of infectious contacts will be equal to the total number of infected individuals in their mating group, whereas for the monogamous sex $I = 1$ if their mate is infected and is 0 otherwise. We also include an external force of infection, such that susceptible individuals are randomly infected with probability κ. Individuals recover from infection with probability σ, at which point they become susceptible to infection again. Infection causes

no increase in mortality, but does carry a risk of permanent sterility (probability γ).

We assume that serially monogamous individuals stay within a mating group unless they are certain that they have not successfully produced offspring in the previous mating season. For the purposes of our model, we assume that mating is only successful provided both partners are not sterile. If they are certain that they have been unsuccessful, then they will reassess their mate choice with probability ρ. In real populations, searching for a new mate may be risky (e.g. increased chance of predation, exposure to new pathogens or risk of exclusion; Jennions and Petrie, 1997) and so it is reasonable to assume that the serially monogamous sex will only leave a mating group if they are certain that they have not successfully reproduced. If an individual leaves a mating group, then they are immediately reassigned to a new mating group according to the original procedure described above. Deaths occur randomly with probability μ; we keep the population size constant, by assuming that there is either always a surplus of offspring, or that the immigration rate is sufficiently high to maintain this balance.

The description of polyandrous and polygynous mating systems has been identical up until this point, but they can be distinguished by the behaviour of sterile members of the monogamous sex. In a polygynous system, each female is able to independently determine whether she has successfully reproduced or not. If she has been unsuccessful, then she may opt to choose an alternate mate in future. In a polyandrous system however, success is based on the ability of the sole female in the mating group to reproduce. As long as there is a chance that a male has successfully reproduced (i.e. the female and at least one male in the group are both fertile), then the benefits of staying within a mating group may outweigh the costs of leaving, even if they are unable to detect who is the father of an infant.

Figure 6.4 shows example simulation dynamics for various degrees of polygamy (α) in polygynous (figures 6.4a-b) and polyandrous (figures 6.4c-d) mating systems. For

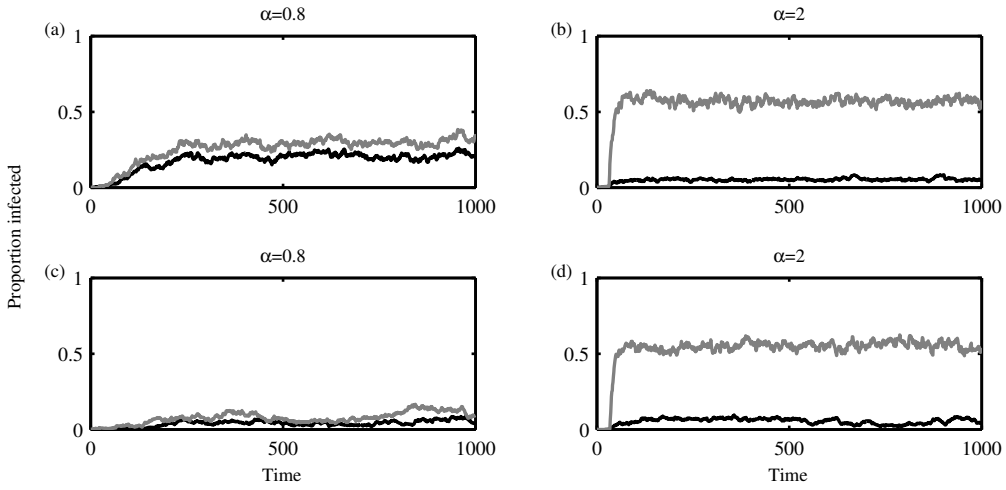

Figure 6.4: Dynamics of an SIS-type infection on sexual contact networks (SCNs) for varying degrees of polygamy (α). (a) and (b) are examples from a polygynous species, whereas (c) and (d) are examples from a polyandrous species. Higher values of α correspond to higher variation in the number of partners for the polygamous sex (equation 6.9). Infection is usually more prevalent in the serially monogamous sex (grey) than in the polygamous sex (black). Parameters: $N_m = N_f = 500$; $\beta = 0.1$; $\gamma = 0.5$; $\kappa = 10^{-4}$; $\mu = 0.01$; $\rho = 0.5$; $\sigma = 0.05$.

relatively low values of α, the infection is able to spread to a reasonably large proportion of a polygynous population, but is unable to spread extensively in a polyandrous population. For higher values of α, the infection is able to propagate through both polygynous and polyandrous populations, with little difference between the two scenarios. At this extreme, most of the monogamous population mates with a small set of individuals, leaving the rest of the polygamous sex disconnected from the network. This simultaneously increases the average exposure of the monogamous sex to infection, whilst decreasing the average exposure of the polygamous sex. As in Thrall et al. (2000), we observe that the polygamous sex tends to exhibit much lower levels of infection than the monogamous sex and the difference tends to increase with greater variance in attractiveness. This is further emphasised in figure 6.5, where the average prevalence of infection in the monogamous sex generally increases with larger

values of α, but peaks at intermediate values of α for the polygamous sex.

The number of cases in the monogamous sex is largely invariant to changes in the level of ρ; this is true for both polygynous and polyandrous scenarios (figures 6.5b, d). Similarly, the prevalence of infection in polyandrous females is only marginally influenced by ρ (figure 6.5c). In accordance with Thrall et al. (2000), the average prevalence of infection in polygynous males increases with ρ, but only for intermediate values of α (figure 6.5a). In fact, the average number of males infected in a highly mobile population ($\rho = 1$) is approximately double that when movement is more limited ($\rho = 0.1$). Higher values of ρ can lead to more mixing between groups and increases mating opportunities for less attractive members of the polygamous sex. Hence, equilibrium levels of infection tend to increase with ρ, but the polygamous sex is disproportionately affected. Mixing tends to be much less common in our polyandrous system, as males are less able to determine whether or not they have successfully produced offspring and so generally choose to stay in a mating group. The opposite is true in the polygynous scenario, as females are always able to distinguish success from failure. This may well be a double-edged sword: although polyandry may increase inertia and limit the spread of infection within the population as a whole, it may increase infection locally.

We find that the probability of sterility (γ) is also an important factor in determining disease prevalence when movement between groups is based on reproductive failure. In particular, the difference between the polygynous and polyandrous scenarios was found to be maximised for intermediate values of γ, but only for low to intermediate values of (α) (figure 6.6).

In section 6.2 we discussed how the ratio of male to female cases during the early stages of an epidemic could be predicted based on the transmission rates between sexes and partner acquisition rates (equation 6.4). Figure 6.7 compares this prediction (using attractiveness, $a(i)$, as a proxy for partner acquisition rates) with the actual

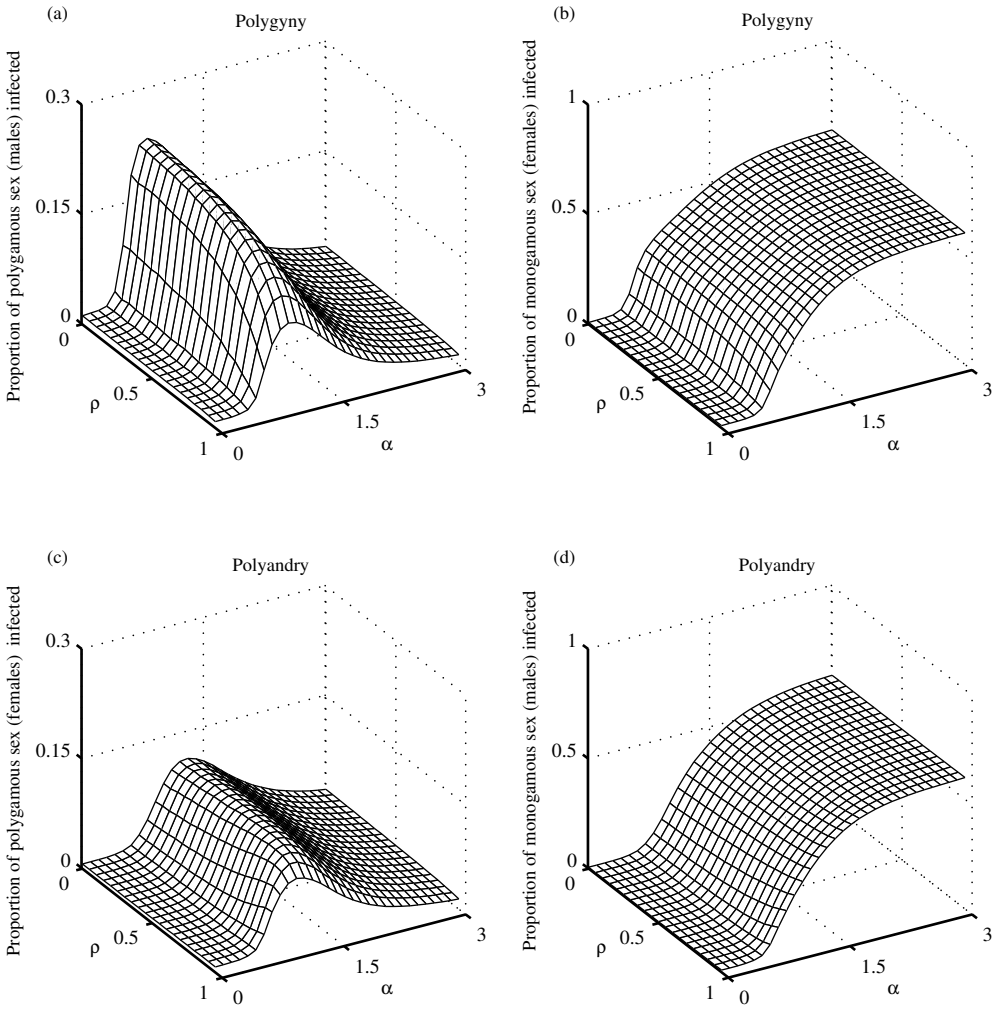

Figure 6.5: The average proportion of the (a, c) polygamous and (b, d) the monogamous sexes that are infected for varying degrees of polygamy (α) and movement probabilities (ρ). (a) and (b) show data from a polygynous species, whereas (c) and (d) show data from a polyandrous species. Higher values of α correspond to higher variation in the number of partners for the polygamous sex (equation 6.9). For the monogamous sex, the prevalence of infection tends to increase with higher degrees of polygamy and movement between groups. By contrast, the prevalence of infection in the polygamous sex peaks at intermediate values of α. This peak is higher for polygyny than for polyandry. Parameters as in figure 6.4.

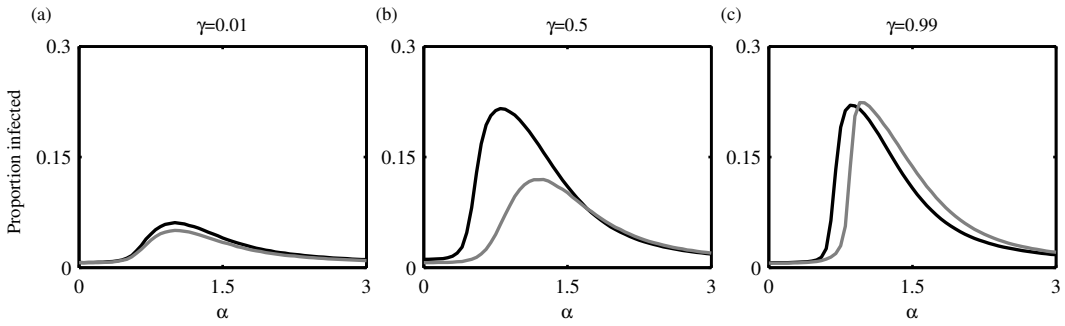

Figure 6.6: The average proportion of the polygamous sex that is infected for polygynous (black) and polyandrous (grey) mating systems. The difference in epidemic size between the two mating systems peaks at intermediate values of α (degree of polygamy) and γ (probability of sterilisation). Parameters as in figure 6.4.

ratio of cases between the monogamous sex (C_M) and the polygamous sex (C_P) for our model. It is clear that the polyandrous system tends to have a greater bias towards infection in the monogamous sex compared to polygynous systems. As discussed above, this is due to increased mixing in polygynous systems exposed to a sterilising pathogen. For low to moderate degrees of polygamy ($\alpha < 1.5$) there is very good agreement between the predicted and actual ratios for the polyandrous scenario, but is generally an overestimate for the polygynous scenario. For higher degrees of polygamy ($\alpha > 1.5$), the predicted and actual ratios tend to diverge, with the prediction increasingly underestimating the actual ratios (for the predicted values, the variance in attractiveness grows linearly with α, giving $C_M/C_P \sim O(\sqrt{\alpha})$; this is because the network is increasingly dominated by a very small number of individuals who are in contact with almost the entire monogamous population (i.e. highly disassortative mating).

Thus far, we have only been concerned with the epidemiological dynamics of our model. We now introduce a second pathogen strain into our model in order to consider the evolutionary implications for pathogens in biased mating systems. Each pathogen strain, p, has a transmission probability per contact β_p and recovery rate σ_p and it is

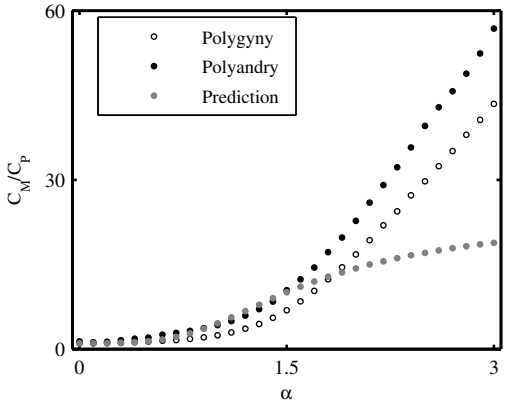

Figure 6.7: The ratio of cases in the monogamous sex (C_M) to the polygamous sex (C_P) during the early stages of the epidemic for varying degrees of polygamy (α). Empty circles correspond to a polygynous system, filled black circles correspond to a polyandrous system and grey circles correspond to the predicted ratio (as per equation 6.4). The ratio C_M/C_P is generally higher in polyandrous systems than in polygynous systems. For low to moderate values of α, the prediction and actual ratios are generally in good agreement, but this breaks down as α increases.

assumed that there is a trade-off between these two values, such that $\beta_p = r(\sigma_p + \mu)^s$ with $r, s > 0$ parameters describing the trade-off and μ equal to the natural death rate. For human populations, the trade-off for more transmissible strains could be interpreted as an increased likelihood of seeking medical treatment due to more visible signs of disease. For simplicity, we do not allow coinfection to occur: if an individual is challenged by two different pathogens in a single time step, then one pathogen is randomly chosen to establish an infection. Both strains are introduced at the start of each simulation and susceptible individuals are infected with probability $\beta_p I + \kappa$, where I is the total number of infectious contacts for a given individual and κ is the external force of infection, as before. We assume that the two strains are equally likely to cause sterility.

Figure 6.8 shows the probability that each strain will account for at least 95% of infections for various degrees of polygamy (α) in the polygynous and polyandrous scenarios when the trade-off between transmission probability per contact (β_p) and

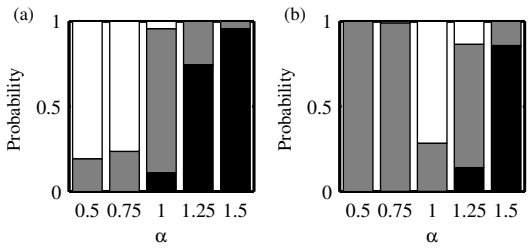

Figure 6.8: Probabilities of different outcomes when two strains compete, with varying degrees of polygamy (α) for (a) polygynous and (b) polyandrous mating systems, and a superlinear trade-off between transmission and recovery rates. Black bars correspond to strain 1 (less virulent) dominating, white bars to strain 2 (more virulent) dominating and grey bars to coexistence. A strain is defined to be dominating if it accounts for at least 95% of infections during the final 20% of a simulation. Parameters: $N_m = N_f = 500$; $r = 1.5$; $s = 1.1$; $\beta_1 \approx 0.07$; $\beta_1 \approx 0.56$; $\gamma = 0.5$; $\kappa = 10^{-4}$; $\mu = 0.01$; $\sigma_1 = 0.05$; $\sigma_2 = 0.4$.

recovery rate (σ_p) is superlinear ($s = 1.1$). Under polygyny (figure 6.8a), the less virulent pathogen (strain 1) tends to dominate for $\alpha > 1$, the more virulent pathogen (strain 2) dominates when $\alpha < 1$ and coexistence is most common when $\alpha = 1$. The pattern is similar for polyandry (figure 6.8b) but for $\alpha < 1$ neither strain is able to become widely established, making coexistence at very low levels highly likely. For linear and sublinear trade-offs, the less virulent pathogen tends to dominate, but coexistence is often still possible (figures C.1-C.2).

A possible reason that the less virulent pathogen tended to be favoured for high degrees of polygamy is that maintaining infection in the dominant member of the mating group is likely to significantly contribute to the survival of a particular strain. Hence long infectious periods (low σ) are likely to be favoured when the degree of polygamy is high, even if this results in a lower basic reproductive number (i.e. if $s > 1$). By contrast, a higher transmission rate may be favoured when the distribution of partners is much more even (small α), as maintaining infection in the most dominant individual in the population becomes less important.

Typically, when multiple pathogens compete for the same pool of hosts, it is predicted that the pathogen with the highest basic reproductive number (R_0) will drive all others to extinction (Anderson and May, 1991). Host heterogeneity can complicate this picture for two reasons. Firstly, different behavioural patterns between groups will select for different traits in pathogens (Eames and Keeling, 2006). Secondly, components of real SCNs are likely to remain unconnected for long periods of time, due to a combination of spatial constraints, assortative mixing and serial monogamy. Distinct components may increase divergent selection due to behavioural differences, but they may also provide spatial refuges for less competitive strains to persist.

6.4 Discussion

Host heterogeneity is known to play an important role in the epidemiological dynamics of many infectious diseases, but is particularly influential among STIs due to the potential for high variability in partner acquisition rates and sexual mixing patterns. Of particular interest are biased mating systems, where high variance in mating success for one sex may leave many individuals isolated from the sexual contact network. Such isolation has been observed in a variety of polygynous and polyandrous mating systems. For example, observations of elephant seals (*Mirounga angustirostris*) indicate that two thirds of males do not mate during a breeding season and that a high proportion of males are unlikely to breed at all during their lifetime (LeBoeuf, 1974). Similarly, approximately 75% of matings in the insect *Zorotypus gurneui* are carried out by dominant males, leaving most males with only sporadic opportunities for mating (Choe, 1994). Other mating systems may contain groups of males and females where mating opportunities are limited for subordinate members of the group, as observed in Tamarins (*Saguinus*; Garber, 1997) and white-winged trumpeters (*Psophia leucoptera*; Sherman, 1995); in both cases, subordinate females are much less likely

to copulate than the dominant female. Even in the absence of a strong dominance hierarchy, populations may still exhibit high variance in partner acquisition rates between the sexes. This is particularly evident in large parts of Africa, where male migrant workers are separated from wives and girlfriends for extended periods of time and are often concentrated in populations with highly unbalanced sex ratios, leading to widespread use of prostitutes (Hunt, 1989). Thus, the distribution of partner acquisition rates for males is usually characterised by a high mean and low variance, whereas the distribution for females will have a low mean and high variance.

In general, large skews in risk and reward for level of sexual activity will lead to high levels of host heterogeneity and will be conducive to the evolution of mate choice. Similarly, the complex decisions that govern mate choice and movement between mating groups are likely to be of crucial importance for the evolution of virulence in STIs. Although the extension of our model to investigate pathogen competition was fairly simplistic, it lends some credence to the notion of coevolution between mating systems and STIs. In particular, the finding that less virulent strains with long infectious periods may be favoured in highly skewed mating systems provides an interesting contrast to studies that have found virulence to increase with greater potential for pathogen dispersal (Lipsitch and Nowak, 1995; Boots and Sasaki, 1999; Eames and Keeling, 2006).

We made the assumption that individuals are generally averse to leaving a mating group, and will only do so if they are certain that they have not produced offspring. Inertia to change could be due to a number of social and environmental factors, such as competition for mates, exposure to new pathogens, increased predation risk or time and energy costs of mate searching (Jennions and Petrie, 1997). Conversely, we could have assumed that individuals only stay within a mating group if they are certain of their success. In this alternate scenario, we would expect levels of infection to be generally higher under polyandry than under polygyny.

We based movement between mating groups on a binary decision (perception of reproductive success/failure), but real systems are likely to demonstrate more complex decisions with regards to mate choice. For example, female fur seals (*Arctocephalus gazelle*) show preference for unrelated, heterozygous males, even if this choice requires increased movement (Hoffman et al., 2007) and monogamous bird pairs may still divorce even if they have successfully reared young together, indicating that other factors contribute to mate choice (Dubois and Cézilly, 2002). Still, one might speculate that by tending to cause sterility rather than mortality (Lockhart et al., 1996), sexually transmitted pathogens could increase divorce rates and movement between mating groups, thereby leading to a higher incidence of disease.

In our model we chose to hold the host population size constant, so as to keep the degree of polygamy (i.e. the distribution of $a(i)$, equation 6.9) fixed and to ensure that any variation in epidemiological dynamics could be unequivocally ascribed to behavioural differences between mating systems. While this was a reasonable assumption to make for the purposes of this paper, our model would need to be extended to explicitly incorporate demographic processes in order to study the evolution of host mating systems in the presence of STIs. It is possible that widespread sterilisation could put host populations under considerable pressure: for example, it has been suggested that sterilising *Chlamydia psittaci* infections may have contributed to declines in koala populations (*Phascolarctos cinereus*; Weigler et al., 1988). Similarly, parasitic nematodes (*Trichostrongylus tenuis*) can reduce host fecundity and are believed to be the primary cause of population crashes among red grouse (*Lagopus lagopus scoticus*; Hudson et al., 1998). Thus, it is probable that widespread sterility would lead to counter-adaptations in the host, such as resistance, tolerance and more advanced mate inspection.

It is widely believed that bright plumage and other sexually-selected ornaments could be indicators of resistance to parasitism, so that high-quality mates can be

readily identified (Hamilton and Zuk, 1982). In proposing the good-genes theory, Hamilton and Zuk (1982) avoid the lek paradox (the depletion of genetic variation) by arguing that genetic variation can be maintained by parasite counter-adaptation, resulting in coevolutionary cycling. An alternative resolution to the lek paradox is the pathogen avoidance argument, which suggests that there is a trade-off between obtaining high-quality genes (maximising the fitness of offspring) and risking exposure to infection (potential reproductive failure) (Hamilton, 1990; Loehle, 1997; Lombardo, 1998; Qvarnström and Forsgren, 1998; Thrall et al., 2000; see chapter 7 herein). Neither theory is likely to be universally applicable (Hamilton and Poulin, 1997), but it is conceivable that the presence or absence of virulent STIs could be used to distinguish between the two in some circumstances. STIs may be expected to have limited involvement in the good-genes theory, due to the tendency for STIs to be asymptomatic (Lockhart et al., 1996), although this is certainly not always the case (e.g. venereal pox; Metz et al., 1985). By contrast, STIs are of crucial importance to the pathogen avoidance theory. As such, one might expect to witness the good-genes theory in action if STIs result in obvious host deterioration, are uncommon or are avirulent, but pathogen avoidance behaviour should be more conspicuous among species with virulent, less visible STIs.

6.5 Future directions

Mathematical models have shown how heterogeneity in sexual behaviour can shape epidemiological dynamics (Hethcote and Yorke, 1984; Gupta et al., 1989; Anderson and May, 1991) and influence the efficacy of intervention programmes (Boily et al., 2002; Eames and Keeling, 2003; Eames, 2007). Still, there are many consequences of host heterogeneity that are yet to be fully understood. The models introduced here and elsewhere (Thrall et al., 1997, 2000; Gouveia-Oliveira and Pedersen, 2009),

suggest that different mating strategies between the sexes can lead to considerable variation in the dynamics of STIs, but many of the evolutionary consequences of such heterogeneity are yet to be determined. For instance, (Boots and Knell, 2002) explored a system where hosts exhibited either risky (highly-active) or safe (less active) mating strategies in the presence of a sterilising STI and found that both strategies are able to coexist for a wide range of parameters, provided the risky strategy carries a fitness benefit in the absence of disease. However, the authors did not explore how non-random mixing might affect the coexistence of risky and safe mating strategies, or whether coexistence is possible when there is a greater degree of host heterogeneity.

Simulated epidemics on SCNs are the most realistic models available for the spread of STIs in human and animal populations, but they are computationally intensive and are often difficult to parameterise. Pair approximations offer some of the realism of SCNs by tracking the formation and break-up of partnerships, making them analytically and computationally more manageable than SCNs, but at the cost of neglecting wider population structure (Eames and Keeling, 2002). Pair approximations are still in their infancy, but have been shown to exhibit dynamics similar to full simulations on SCNs (Eames and Keeling, 2002, 2004, 2006). Various degrees of host heterogeneity (Eames and Keeling, 2002) and assortative mixing (Eames and Keeling, 2006) have been modelled using this approach, but there is considerable scope for further research on these topics. In addition, pair approximations for polygynous/polyandrous mating systems are noticeably absent from the literature.

Various models incorporating heterogeneity in host contact structure have been used to study the evolution of pathogen virulence (Lipsitch and Nowak, 1995; Lipsitch et al., 1995; Read and Keeling, 2003; Eames and Keeling, 2006; Messinger and Ostling, 2009) and of antigenic diversity (Buckee et al., 2004, 2007), but these have not been widely applied to STIs. Models have also been used to explore the role of STIs in

the evolution of host-mating strategies (Thrall et al., 1997, 2000; Kokko et al., 2002), but there have been very few studies that have combined these approaches to explore coevolution between hosts and STIs. As an exception, Prado et al. (2009) explored how host sociality (i.e. contact frequency) and pathogen virulence may coevolve on a contact network. The authors found that high levels of sociality tend to benefit more virulent pathogens, but then selection will favour more cautious hosts and subsequent reductions in virulence, which can lead to coevolutionary cycling in these traits. An exciting avenue for future work in this area would be to explicitly incorporate host and pathogen genetics within a coevolutionary framework with some plasticity in mating strategies, particularly as the predictions of such models may be amenable to testing in a wide variety of animal systems.

Chapter 7

Coevolution of mate choice and virulence among sexually transmitted infections

Ben Ashby[1]

Abstract

Sexually transmitted infections (STIs) are likely to select for mating strategies that limit disease exposure, but this will in turn select for traits that allow STIs to avoid host defences, leading to subsequent adaptations in both populations in an ongoing coevolutionary struggle. Here, I explore how mate choice in the form of a preference for healthy individuals coevolves with STI virulence. I first show that this simple behavioural adaptation can constrain virulence in STIs that reduce host fertility. I then show that variation in costs associated with choosy behaviour and the strength of trade-offs between transmissibility and virulence can lead to a range of coevolutionary dynamics, including fluctuating selection, stable equilibria and runaway virulence. Hosts may also retain a preference for healthy mates even if STIs evolve to be avirulent, which has important implications for understanding the role that STIs play in sexual selection. The variety of dynamics that I observe demonstrates the need to consider host-STI interactions in a coevolutionary context.

[1]Department of Zoology, University of Oxford, Oxford, UK

7.1 Introduction

Sex is the dominant mode of reproduction among higher organisms, but can leave individuals at risk of infection due to sustained close-contact with sexual partners (e.g. transmission of ectoparasites) or through the transfer of genetic material (e.g. transmission of pathogens in seminal fluid) (Knell and Webberley, 2004). Thus, sexually transmitted infections (STIs), which are common in both plants (Mink, 1993) and animals (Lockhart et al., 1996), are predicted to play an important role in the evolution of mating strategies and secondary sex traits (Sheldon, 1993; Able, 1996; Thrall et al., 1997, 2000; Knell, 1999; Knell and Webberley, 2004; Boots and Knell, 2002; Kokko et al., 2002; Tybur and Gangestad, 2011; Ashby and Gupta, 2013). Although a large number of studies have explored the effects of parasitism on mate choice, few have considered how sexual selection in the host influences parasite evolution and it is currently unclear whether mate choice can persist if parasites evolve in response to sexual selection, as is likely to be the case (Knell, 1999). More generally, the processes that shape the coevolutionary dynamics of hosts and STIs are yet to be explored.

Hamilton and Zuk (1982) were first to suggest that parasitism may play a role in sexual selection, with bright plumage and other male ornaments existing to signal genetic quality in the form of resistance to parasitism, allowing females to identify mates that can pass on 'good genes' to their offspring. Although some support exists for basic aspects of this hypothesis (Borgia and Collis, 1990; Clayton, 1991), evidence from natural populations is generally lacking (Hamilton and Poulin, 1997) or is difficult to obtain (Reid et al., 2005; Balenger and Zuk, 2014) and a number of theoretical issues have been identified, such as its dependence on indirect rather than direct selection (Kirkpatrick and Ryan, 1991) and the need for associations between genes for showy traits and mate choice (Loehle, 1997). Several authors have proposed alternative hypotheses for parasite-mediated sexual selection (PMSS), most of

which can be classified as 'transmission-avoidance' theories that suggest that mating strategies and secondary sex traits may have evolved to limit the risk of infection (Freeland, 1976; Hamilton, 1990; Loehle, 1995; Able, 1996; Loehle, 1997). The two types of theory are not exclusive, as individuals with resistance genes are more likely to be free from infection, so will tend to experience greater mating success in both scenarios. Nevertheless, transmission-avoidance theories do not suffer from many of the problems associated with good genes models (e.g. selection is direct, fewer restrictive assumptions) and are more consistent with empirical observations (Borgia and Collis, 1990; Loehle, 1997). In addition, transmission-avoidance theories do not depend on showy traits for parasites to influence the evolution of mating strategies.

Models of mate choice based on transmission-avoidance have tended to focus on STIs, as they are more likely to influence mate choice than non-STIs due to their route of transmission (Loehle, 1997; Knell, 1999; Boots and Knell, 2002). Although this is somewhat self-evident, it is vital to make the distinction as STIs typically have different epidemiological dynamics (Anderson and May, 1991; Ashby and Gupta, 2013) and disease outcomes (sterility rather than mortality; Lockhart et al., 1996) to non-STIs. The propensity for STIs to inhibit host reproduction rather than increase mortality is particularly important in the context of virulence evolution, as these parasites do not experience a trade-off between transmission rate and longevity of infection, leading to predictions of runaway virulence (i.e. full castration) (O'Keefe and Antonovics, 2002; O'Keefe, 2005). Despite this rather ominous prediction, many STIs only cause partial reductions in host fertility (Lockhart et al., 1996) and the relative scarcity of full parasitic castrators in natural populations has yet to be explained. Limited dispersal (O'Keefe and Antonovics, 2002; O'Keefe, 2005), vertical transmission (Lipsitch et al., 1996) and tolerance mechanisms (Best et al., 2010a) have been shown theoretically to select against full castration, but it is likely that social exclusion and other behavioural traits that limit disease transmission will also constrain virulence (Boots

and Sasaki, 1999; Knell, 1999; de Roode and Lefevre, 2012; Ashby and Gupta, 2013). Hence, PMSS may help to explain both the evolution of mate choice and the relative scarcity of fully-castrating STIs. Knell (1999) suggested that parasites will evolve to become less virulent if hosts experience selection for disease-avoidance traits, in turn reducing selection in the host. This key insight has been overlooked in studies that only consider host evolution (Hamilton and Zuk, 1982; Loehle, 1997; Thrall et al., 1997; Boots and Knell, 2002; Kokko et al., 2002), so at first glance, it would appear that the potential importance of PMSS may have been overstated. However, while Knell (1999) recognised the importance of feedback for selection in the host, he did not account for a full coevolutionary interaction (i.e. feedback in both directions). Lower virulence may reduce the strength of sexual selection, but this will in turn enable STIs to exploit a lack of disease-avoidance traits, potentially allowing virulence to once again increase. Hence, it is necessary to use a coevolutionary approach to understand PMSS, so that evolutionary pressures in both populations and feedback between them can be taken into account.

Here, I theoretically explore the coevolution of mate choice and virulence among sterilising STIs. I model mate choice as a mechanism for transmission-avoidance, based on the following criteria: (i) parasites with higher transmission rates cause greater and more conspicuous damage to their hosts; and (ii) individuals preferentially choose mates that do not exhibit signs of infection (hereafter referred to as 'mate inspection'). The first assumption is reasonable given that a wide range of parasites appear to exhibit a trade-off between transmissibility and disease severity (reviewed in Ebert and Bull, 2003) and many parasites are also known to cause visible damage to their hosts (e.g. to plumage, Hamilton and Zuk, 1982; Hamilton and Poulin, 1997; or cloaca, Metz et al., 1985). Several species of birds carry out cloaca-pecking or inspection (e.g. *Prunella modularis*, Davies, 1983; *P. collaris*, Nakamura, 1990; *Passer domesticus*, Moller, 1987) and some organisms are known to show preference

for healthy social contacts, relying on visual, behavioural or olfactory cues to determine the condition of other individuals (Hart, 1990; Loehle, 1995; Kavaliers et al., 1999), giving credence to the second assumption. Using deterministic and stochastic models, I first demonstrate that fixed levels of mate inspection can reduce selection for virulence in sterilising STIs. I then allow the strength of mate inspection (preference for healthy mates) to coevolve with the parasite and show that a range of outcomes can occur (fluctuating selection, coevolutionarily stable strategies and runaway virulence) depending on costs associated with mate inspection and trade-offs between virulence and transmission.

7.2 Model description

I model the spread of a chronic STI through a serially monogamous population of hermaphrodites, where hosts either have one or no sexual contacts at any given time. I assume that parasites are transmitted from infectious individuals to their susceptible partners at a rate β and that there is no non-sexual transmission or superinfection. I also assume that disease causes a permanent reduction in host fertility, but has no effect on host mortality, with more transmissible parasites associated with greater virulence. I set the fertility, f, of uninfected hosts equal to 1 and the fertility of infected hosts equal to $\exp(-\alpha\beta)$, where α modifies the strength of the trade-off between the transmission rate and virulence $(1 - f)$.

Sexual partnerships are formed between currently unpaired individuals based on their preference for healthy mates. Specifically, two unpaired members of the population, i and j, form a sexual partnership at a rate $P(i, j) = pc^{\gamma_i+\gamma_j} f_i^{\gamma_j} f_j^{\gamma_i}/N_U$, where p is the base rate of pair formation, N_U is the number of potential partners, γ_i and f_i are the strength of mate inspection (preference for healthy mates) and the fertility for host i, respectively, and c is an associated cost of parasite avoidance,

which captures time that is wasted due to stricter standards of mate inspection (e.g. longer courtships). In the absence of mate inspection $P(i,j) = p/N_U$. Note that the overall pair formation rate is independent of the number of unpaired individuals, in line with standard frequency-dependent models of STIs (May and Anderson, 1979). Sexual partnerships end if one partner dies (all hosts have a mortality rate of μ), or if either member decides to terminate the partnership, which occurs at a rate d per pair. Individuals leaving a pair are immediately able to form new sexual partnerships. A partnership between individuals with fertilities f_i and f_j produces offspring at a rate of $\mathrm{r} f_i f_j (1 - N/K)$, where r is the maximum birth rate, N is the current host population size and K is the carrying capacity in the absence of mortality. Note that only paired individuals are able to produce offspring.

I implement two versions of my model. I analyse the epidemiological and (co)evolutionary dynamics of the system using a deterministic pair formation model (PFM), which (in the absence of genetic variation) can be represented by the following equations:

$$\left.\begin{aligned}
\frac{dS}{dt} &= b + (d+\mu)(2[SS] + [SI]) - \frac{pc^{2\gamma}S}{N_U}(S + f^{\gamma}I) - \mu S \\
\frac{dI}{dt} &= (d+\mu)(2[II] + [SI]) - \frac{pc^{2\gamma}f^{\gamma}I}{N_U}(S + f^{\gamma}I) - \mu I \\
\frac{d[SS]}{dt} &= \frac{pc^{2\gamma}S^2}{2N_U} - (d+2\mu)[SS] \\
\frac{d[SI]}{dt} &= \frac{pc^{2\gamma}f^{\gamma}SI}{N_U} - (d+2\mu+\beta)[SI] \\
\frac{d[II]}{dt} &= \frac{p(cf)^{2\gamma}I^2}{2N_U} - (d+2\mu)[II] + \beta[SI]
\end{aligned}\right\} \qquad (7.1)$$

where $[SI]$ indicates the number of susceptible- (S) infectious (I) pairs (similarly

for $[SS]$ and $[II]$) and classes without brackets represent unpaired members of the population. The total population size is $N = N_U + 2([SS] + [SI] + [II])$, with $N_U = S + I$ and the birth rate is given by $b = r(1 - N/K)([SS] + f[SI] + f^2[II])$. The PFM for polymorphic populations is omitted here for the sake of brevity, but is fully-described in appendix D. Either member of a pair may die or terminate the partnership, so each partnership ends at a rate of $(2\mu + d)$. Terms of the form $(d + \mu)(2[AA] + [AB])$ (where $A, B \in \{S, I\}$) represent individuals that have just become unpaired. It is straightforward to show that the pair formation terms (i.e. the sum of the third term in the equation for dS/dt and second term in the equation for dI/dt) balance with half the sum of the first terms in the 'paired' equations.

I use an adaptive dynamics approach with the PFM to establish the invasion success of 'mutant' hosts or parasites while the resident population is at equilibrium (Hofbauer and Sigmund, 1990; Dieckmann and Law, 1996). This amounts to a separation of ecological and evolutionary timescales so that evolutionarily and co-evolutionarily stable strategies (ESS/CoESS) can be determined. I use numerical methods as the complexity of the PFM makes it intractable to algebraic analysis.

I verify the results of the PFM using a stochastic individual-based model (IBM; based on the first-reaction method proposed by Gillespie, 1977), which relaxes the assumption of separate ecological and evolutionary timescales. In this version, the transmission rate (β) and the strength of mate inspection (γ) (co)evolve as quantitative traits. Newborn hosts inherit the strength of mate inspection from a randomly chosen parent, subject to mutation. Similarly, parasites inherit the transmission rate of the previous generation, subject to mutation. In both cases, the trait for the new generation X', is given by $X' = \max(0, X + \zeta\epsilon_k)$, where X is the parental trait value, ζ is a normally distributed random number with mean 0 and standard deviation 1, and ϵ_k scales the mutation rate ($k = H$ for hosts and $k = P$ for parasites). Simulations begin with monomorphic host and parasite populations, with all individuals

unpaired and a small number infected.

7.3 Results

7.3.1 Epidemiology

An important epidemiological parameter for characterising the success of any parasite is the basic reproductive ratio, R_0, which is equal to the average number of secondary infections produced by a single infectious individual in an otherwise susceptible population (Anderson and May, 1991). Hence if $R_0 > 1$, the parasite should spread, else it will go extinct. R_0 can be calculated for the models presented here by taking the product of: (i) the expected number of partners that the infectious individual will have in his/her lifetime; and (ii) the probability of infecting a partner. In a monomorphic population, the first term can be derived from a geometric progression involving the pair formation and dissolution rates (see appendix D), while the second term is simply equal to $\beta/(\beta + d + 2\mu)$, giving:

$$R_0 = \frac{pc^{2\gamma} f^{\gamma} \beta (d + 2\mu)}{\mu (pc^{2\gamma} f^{\gamma} + d + 2\mu)(\beta + d + 2\mu)} \tag{7.2}$$

It is not immediately obvious from this equation how the strength of mate inspection, γ, affects R_0, but note that it can be rewritten in the following form:

$$R_0 = \frac{a_0}{a_1 + a_2 a_3^{-\gamma}} \tag{7.3}$$

where a_k are positive coefficients and $a_3 = c^2 f$. Differentiating equation 7.3 with respect to γ gives:

$$\frac{dR_0}{d\gamma} = \frac{a_0 a_2 a_3^{\gamma} \ln(a_3)}{(a_1 + a_2 a_3^{\gamma})^2} \tag{7.4}$$

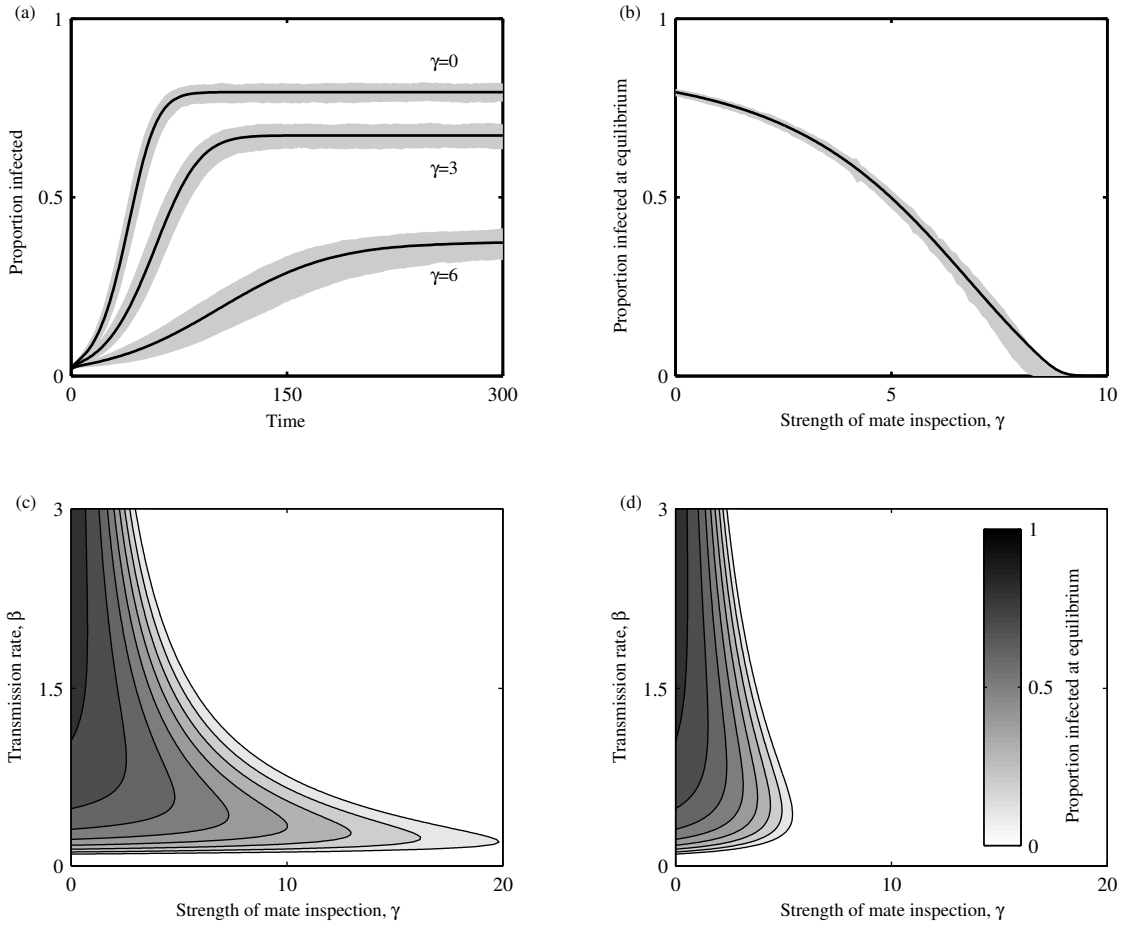

Figure 7.1: Epidemiological dynamics and equilibrium prevalence of infection in the presence of mate inspection. (a) Time series showing the proportion of the host population infected during the course of an epidemic for different levels of mate inspection (γ). Shaded areas correspond to the IBM (mean $\pm$ 1 standard deviation, 500 simulations); solid lines generated using the PFM. (b) The proportion of the population infected at equilibrium as a function of γ. Shading and lines as in (a). (c) Equilibrium prevalence of infection in the PFM as a function of the strength of mate inspection (γ) and transmission rate (β), with $c = 1$. (d) As in (c), but with $c = 0.9$. Parameters: $\alpha = 0.25$, $\beta = 1$, $\epsilon_H = \epsilon_P = 0$, $\mu = 0.05$, $d = 0.2$, $K = 500$, $p = 1$ and $r = 1$.

which is always negative when $f < 1$ and tends towards 0 as $\gamma \to \infty$. Thus, both the initial growth rate of an epidemic and the equilibrium prevalence of infection are constrained by stricter mate inspection (figure 7.1). Decreasing the parameter c

also reduces R_0 due to slower pairing rates on average (figure 7.1d). The non-linear relationship between R_0, β and γ means that relatively small changes in parameter values can lead to drastic changes in disease incidence. In particular, the close contour lines in figures 7.1c-d indicate that slight changes in the strength of mate inspection when transmission rates are high will rapidly suppress STI prevalence. Similarly, when mate inspection is weak a minor increase in transmission rate can lead to much larger epidemics. Together, these dynamics suggest that both populations are capable of exerting strong selection on each other.

7.3.2 Evolution of virulence

I examined the effects of fixed levels of mate inspection (γ) on the evolution of virulence using both the PFM and the IBM. In the absence of mate inspection ($\gamma = 0$), R_0 (equation 7.2) is an increasing function of β, so the STI should evolve to fully castrate the host (O'Keefe and Antonovics, 2002). If mate inspection is present ($\gamma > 0$), R_0 can be rewritten in the following form:

$$R_0 = \frac{a_4}{(a_5 + a_6 f^{-\gamma})\left(1 + \frac{a_7}{\beta}\right)} \tag{7.5}$$

where a_k are positive coefficients. Recall that $f = \exp(-\alpha\beta)$, so $f^{-\gamma} \to 1$ as $\beta \to 0$ and $f^{-\gamma} \to \infty$ as $\beta \to \infty$. In both cases, $R_0 \to 0$, which means that an intermediate transmission rate (and hence virulence) maximises R_0 (figure 7.2b). This argument holds for other functional forms of f, provided they are strictly decreasing functions of β with $f(0) = 1$. Here, the optimal level of virulence (ESS) depends on both α and γ, as shown in figure 7.2a, which compares predictions obtained using equation 7.5 with simulation data from the IBM. Although R_0 only describes the initial spread of an infection, it is very accurate in approximating the optimal level of virulence that evolves in the stochastic IBM. Note that stronger mate inspection reduces both

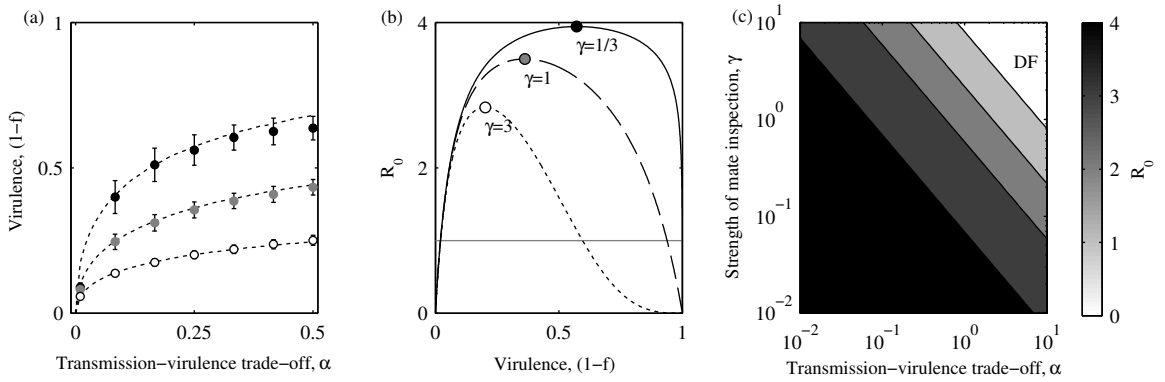

Figure 7.2: Evolution of virulence in the presence of mate inspection. (a) Optimal level of virulence $(1-f)$ as predicted by equation 7.2 (dotted lines) compared with simulation data from the IBM (mean $\pm$ 1 standard deviation; $\gamma = 1/3$ (black), $\gamma = 1$ (grey) and $\gamma = 3$ (white)). The predictions closely match simulation data, but are less accurate when parasites become highly virulent as the host population is more likely to go extinct. (b) R_0 as a function of virulence, for various levels of mate inspection. The optimal value of R_0 is indicated by the circle on each curve. (c) The optimal value of R_0 as a function of the transmission-virulence trade-off parameter (α) and the strength of mate inspection (γ). $R_0 < 1$ in the region labelled 'DF' (disease-free). Fixed parameters as in figure 7.1, with $\epsilon_P = 0.1$ and $c = 1$.

virulence and R_0 (figure 7.2b).

7.3.3 Coevolution

The results from the previous section showed that fixed levels of mate inspection produce an ESS with intermediate levels of virulence. However, if mate inspection is an inherited trait, we should expect variation in virulence to modulate selection in the host. Similarly, changes in the strength of mate inspection in the host will alter selection for virulence. Using an adaptive dynamics approach, I determined how these ecological feedbacks affect coevolutionary outcomes in the PFM.

Figure 7.3 shows how the strength of the trade-off between the transmission rate and virulence (α), and the cost associated with mate inspection (c) affect the coevo-

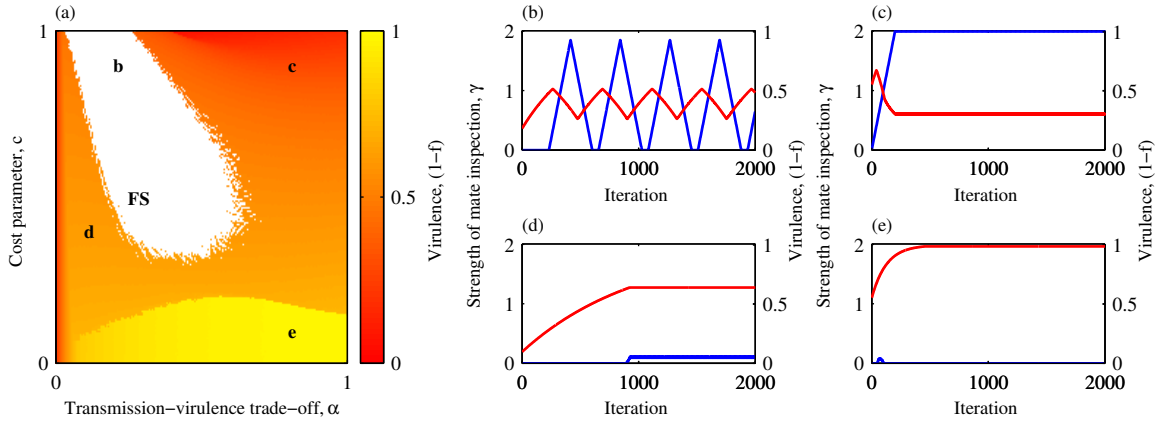

Figure 7.3: Coevolutionary dynamics in the PFM, obtained using an adaptive dynamics approach. (a) Evolved level of virulence, as a function of the virulence-transmission trade-off (α) and cost of mate inspection (c) parameters. Parameters in the white region produce coevolutionary trajectories that exhibit fluctuating selection (FS). Coevolutionary trajectories for the strength of mate inspection (γ; blue) and virulence ($1 - f$; red) at points b-e are shown in the corresponding subplots. (b) Fluctuating selection. (c) Mate inspection evolves to restrict virulence (non-monotonic trajectory for virulence). (d) A very low level of mate inspection still manages to constrain virulence. (e) Parasites evolve to become castrators (runaway virulence), leading to host extinction. Fixed parameters as in figure 7.1.

lution of these traits. Broadly speaking, the model produces four types of dynamics (figures 7.3b-e): (i) fluctuating selection; (ii) CoESS (non-monotonic approach); (iii) CoESS (monotonic approach); and (iv) host extinction due to runaway virulence. Fluctuating selection occurrs when the trade-off between transmission and virulence is not too strong relative to the cost of mate inspection (figure 7.3b). An initial lack of mate inspection allows higher virulence to evolve, leading to selection for mate inspection, which precipitates a reduction in virulence. At this point mate inspection is selected against as it is costly when the STI is avirulent, allowing the cycle to repeat. When the trade-off between transmission and virulence is very strong relative to the cost of mate inspection (high α, c), parasites exhibit an initial increase in virulence while mate inspection is weak (figure 7.3c). Increasingly strict mate inspection leads

to a reduction in virulence and the system tends towards a CoESS. Fluctuating selection does not occur under these conditions, because the costs of mate inspection are relatively low. When the costs of mate inspection are higher (low c), parasites evolve greater levels of virulence, which is either eventually curtailed by very weak mate inspection (figure 7.3d) or continues increasing until the host becomes extinct (figure 7.3e).

Qualitatively similar dynamics to those described for the PFM emerge in the stochastic IBM (see figure 7.4 for examples of fluctuating selection). Small values of α produce lower frequency oscillations, which are characterised by short intervals of strong mate inspection, interspersed with long intervals of weak mate inspection (virulence increases slowly due to a weak relationship with transmission; figures 7.4a-b), whereas larger values of α produce higher frequency oscillations (virulence changes rapidly due to the strong relationship with transmission; figures 7.4c-d). Oscillations tend to have a large amplitude when mate inspection is not too costly (high c; figures 7.4b, d), but their magnitude decreases with higher costs (low c; figures 7.4a, c).

I constructed pairwise invasion plots (PIPs) to analyse why the system produces these dynamics (figure 7.5). Each host PIP was generated by holding the trait value for the parasite population constant (and vice versa). PIPs were generated using an adaptive dynamics approach, by comparing how mutants performed in a resident population that was previously at equilibrium. If a particular mutant replaced the resident, then it was subsequently challenged by new mutants. Hence, PIPs were used to determine the direction of selection under different conditions.

I obtained three qualitatively different types of PIP during my analysis, which were classified as either stable (contained a single ESS; figures 7.5c, f), unstable (contained no ESSs; figures 7.5d, e) or bistable (contained two ESSs; figures 7.5a-b). Hosts produce only stable and bistable PIPs, whereas parasites produce stable

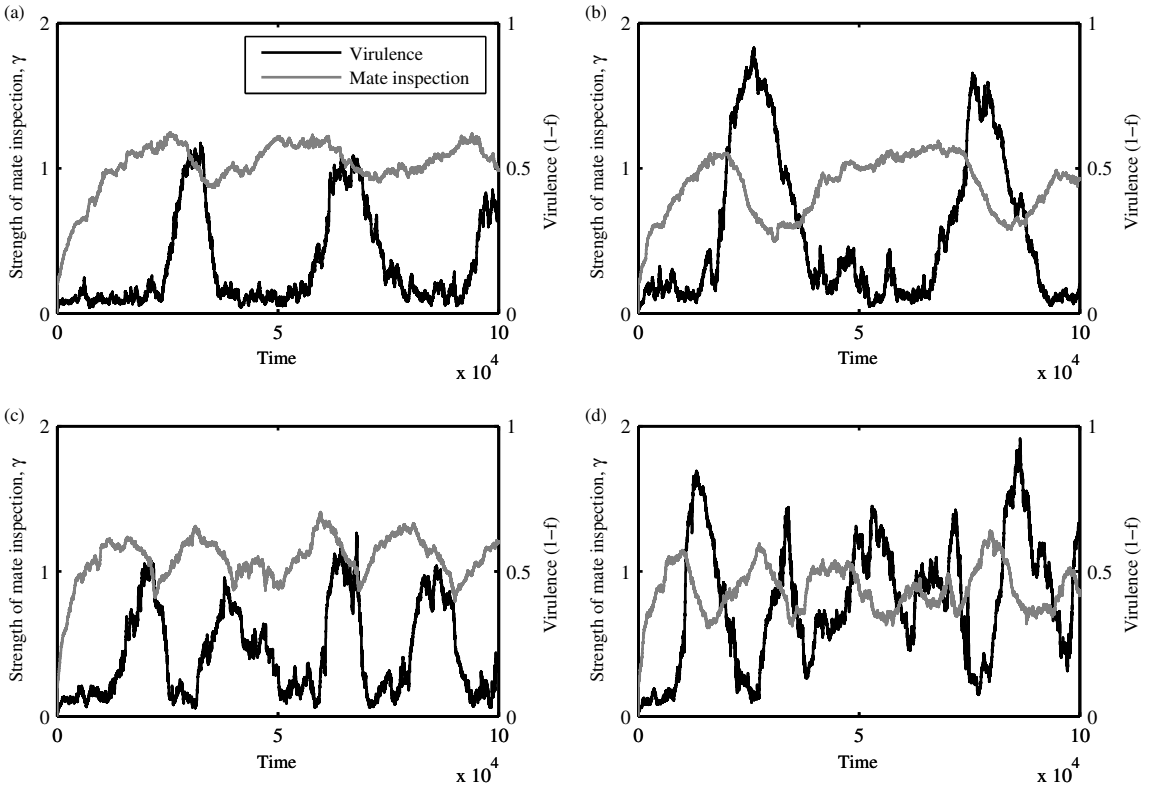

Figure 7.4: Coevolutionary trajectories in the stochastic IBM for (a) $\alpha = 0.3$, $c = 0.6$; (b) $\alpha = 0.3$, $c = 0.8$; (c) $\alpha = 0.5$, $c = 0.6$; (d) $\alpha = 0.5$, $c = 0.8$ (strength of mate inspection: black; virulence: grey). Small values of α produce lower frequency oscillations, which are characterised by short intervals of strong mate inspection, interspersed with long intervals of weak mate inspection (virulence increases slowly due to a weak relationship with transmission; (a)-(b)), whereas larger values of α can produce higher frequency oscillations (virulence changes rapidly due to the strong relationship with transmission; (c)-(d)). In addition, oscillations tend to have a large amplitude when mate inspection is not too costly (high c; (b), (d)), but their magnitude decreases with higher costs (low c; (a), (c)). Fixed parameters as in figure 7.1, with $\epsilon_H = \epsilon_P = 0.02$ and $c = 1$

and unstable PIPs. Taken together, these results can be used to illustrate why the dynamics outlined above emerge. For example, if mate inspection has yet to evolve and virulence is currently low, any small increase in γ will be selected against, but large jumps are evolutionarily stable (bistability; figure 7.5a). If γ stays close to 0, the transmission rate (hence virulence) will increase (figures 7.5d-e), so the region

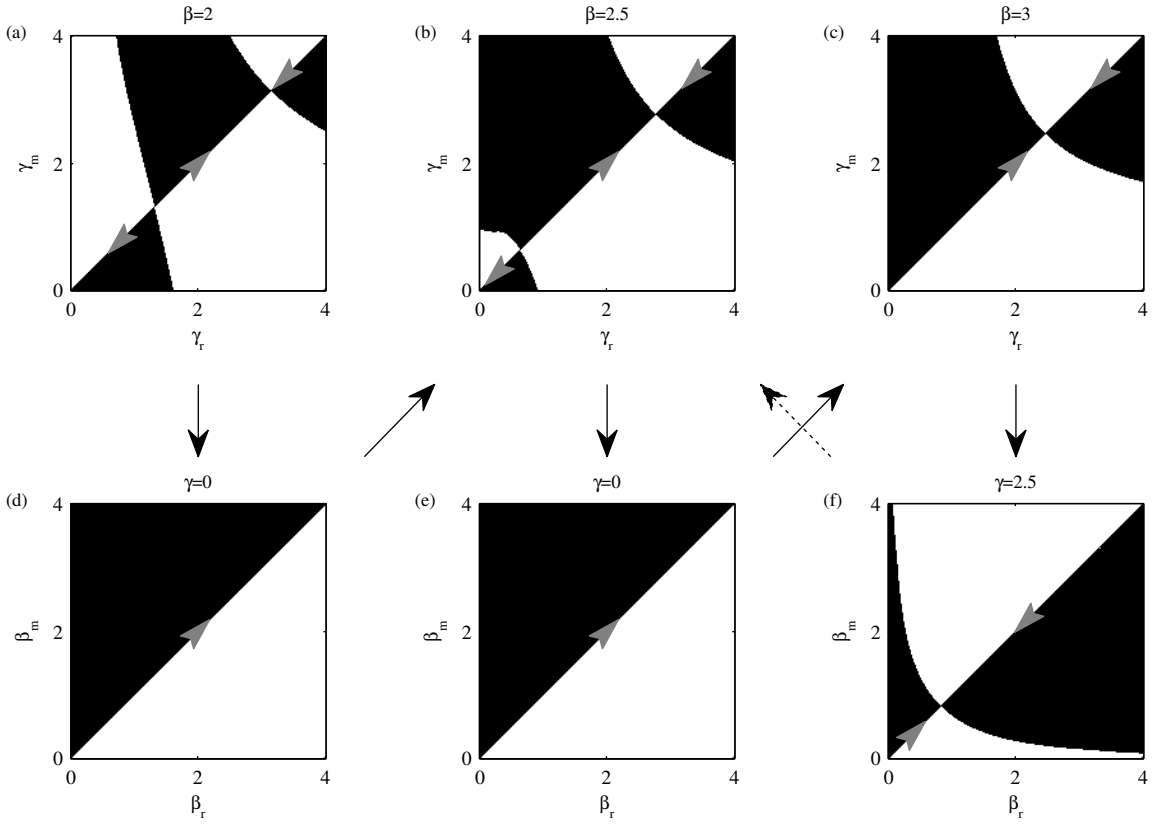

Figure 7.5: Pairwise invasion plots (PIPs), showing the coevolution of the strength of mate inspection (γ; (a)-(c)) and the transmission rate (β; (d)-(f)). White and black regions indicate that the resident (subscript r) and mutant (subscript m) dominate, respectively; hence, traits will evolve in the direction of the grey arrowheads. Each host PIP is generated by holding the trait value for the parasite population constant (and vice versa; constant trait values indicated above each PIP). Coevolution can lead to qualitative shifts in the direction of selection due to feedback, as illustrated by the sequence of black arrows. For example, suppose mate inspection has yet to evolve and virulence is currently low (a). Any small increase in γ will be selected against, but large jumps are evolutionarily stable. If γ stays close to 0, the transmission rate (hence virulence) will increase (d). As virulence increases, the region under which mate inspection is selected against shrinks (b), but the transmission rate will continue to increase (e). When the transmission rate passes a critical threshold, γ begins to increase (c), which restricts virulence (f). At this stage, populations either tend towards a stable equilibrium with $\beta, \gamma > 0$, or the system exhibits fluctuating selection. In the latter case, reductions in the transmission rate reintroduce selection against mate inspection ((b); dotted arrow), allowing the cycle to repeat. Fixed parameters as in figure 7.1, with $c = 0.9$.

under which mate inspection is selected against shrinks (figure 7.5b). Provided mate inspection is not too costly, this trait begins to increase when virulence passes a critical threshold (figure 7.5c), which prevents runaway evolution of the parasite (figure 7.5f). At this stage, populations will either tend towards a CoESS with $\beta, \gamma > 0$, or the system will exhibit fluctuating selection. In the latter case, reductions in virulence reintroduce selection against mate inspection (figure 7.5b; dotted arrow), allowing the cycle to repeat.

7.4 Discussion

While many studies have explored the effects of sexual contact patterns on epidemiology and parasite evolution (Hyman and Li, 1997; Eames and Keeling, 2002, 2004; Lloyd-Smith et al., 2004; Ashby and Gupta, 2013), or the consequences of PMSS for the evolution of mate choice (Hamilton and Zuk, 1982; Sheldon, 1993; Able, 1996; Thrall et al., 1997, 2000; Knell, 1999; Boots and Knell, 2002; Kokko et al., 2002; Howard and Lively, 2003; Nuismer et al., 2008; Ashby and Gupta, 2013), the current study is the first that examines both hosts and STIs in a coevolutionary context. The models presented herein demonstrate that ecological feedbacks are crucial in shaping selection in both populations, in line with other host-parasite models with non-STIs (Gandon et al., 2002; Best et al., 2009; Prado et al., 2009; Best et al., 2010a; Débarre et al., 2012; Ashby and Gupta, 2014; Ashby et al., 2014a,b). This is because the spread of traits that reduce the risk of infection will inevitably affect disease incidence at the population level, increasing selection for parasites that can avoid host defences, which in turn influences host evolution. Knell (1999) argued that this process will reduce selection for disease-avoidance traits, so previous models may have overstated the importance of parasitism in sexual selection. However, a lack of disease-avoidance traits is likely to induce selection for greater virulence, which again

modulates host evolution, and so on. In my model, a stronger preference for healthy mates can lead to reductions in virulence and consequently lessens the need to inspect new sexual partners, potentially allowing virulence to recover. My results show that this feedback can produce a range of coevolutionary dynamics, including fluctuating selection, stable equilibria and runaway virulence, depending on costs associated with mate inspection and the relationship between transmissibility and STI virulence.

Sustained fluctuating selection has not previously been observed in the context of mate choice and STIs, although Boots and Knell (2002) did observe damped fluctuations in promiscuity in response to a sterilising STI in a model that lacked parasite evolution. Graves and Duvall (1995) reasoned that the type of dynamics described herein could arise for hosts and STIs due to an ecological feedback, but these dynamics have yet to be demonstrated until now. In a related study, Prado et al. (2009) recently showed that selection for sociality and virulence in directly-transmitted parasites can fluctuate due to a similar process. Analysis of the PFM revealed that fluctuating selection is maintained due to bistability in certain regions of parameter space, but bistability can also allow mate inspection to remain present at fixed levels once it has emerged (figures 7.3c, 7.5a-b). This is because increasing levels of virulence can lead to the evolution of mate inspection, resulting in selection for milder STIs, but the bistability of the system will not necessarily lead to a subsequent reduction in mate inspection (e.g. figure 7.3c). In other words, hosts may retain disease-avoidance traits even when the forces that drove their initial evolution have been removed. These dynamics highlight a potentially significant problem with the use of comparative studies to determine the importance of parasitism in sexual selection, as two populations could experience equally virulent STIs with only one exhibiting mate choice, which would suggest that parasitism is unimportant, when in fact the opposite might be true.

Although the primary focus of the present study has been to understand host-STI

coevolutionary dynamics, my results also have implications for the evolution of virulence among sterilising infections (including non-STIs), which are typically predicted to become full castrators in the absence of vertical transmission (Lipsitch et al., 1996), unless dispersal is limited (O'Keefe and Antonovics, 2002; O'Keefe, 2005) or hosts are able to evolve tolerance mechanisms (Best et al., 2010b). I have demonstrated that behavioural adaptations can also prevent full castrators from evolving, which may help to explain why this disease outcome is relatively rare. Limits on the evolution of castration have potentially important implications for other aspects of coevolution, especially fluctuations in gene frequencies that govern specificity. For example, Ashby and Gupta (2014) demonstrated that castration is necessary for such coevolutionary cycling to persist in continuous-time models, although the authors focused on non-STIs (chapter 5).

The validity of my model hinges on individuals being able to detect STIs through visual, olfactory or behavioural cues and showing preference for healthy mates. Some STIs are known to produce observable damage (e.g. genital warts, abnormal discharge) and there is some evidence that visual mate inspection occurs among birds in the form of pre-copulatory displays and cloaca-pecking (Davies, 1983; Moller, 1987; Nakamura, 1990; Sheldon, 1993). A number of studies have shown that animals are capable of detecting the condition of other individuals (e.g. (Webster et al., 2003)), sometimes through olfactory cues alone (Penn and Potts, 1998; Kavaliers et al., 2004), with female humans seemingly preferring the sweat of males that are not infected with the STI gonorrhoea (Moshkin et al., 2012), but the underlying mechanisms involved are not well-understood and the extent to which these cues correlate with virulence or transmissibility is unknown. Parasitism can also reduce competitive ability due to direct damage or costs associated with mounting an immune response (e.g. (Jacot et al., 2004)), so infected individuals could still experience lower mating success even in the absence of visual or olfactory cues.

However, several studies have found little or no support for condition-dependent mate choice. For example, Webberley et al. (2002) found that the two-spot ladybird, *Adalia bipunctata*, showed no preference for mates that were not infected with a sexually transmitted ectoparasite (*Coccipolipus hippodamiae*). Similarly, the sexually transmitted nematode, *Mehdinema alii*, does not influence mate choice by female crickets (*Gryllodes sigillatus*) (Luong and Kaya, 2005). Furthermore, Nunn (2003) found no evidence of genital inspection as a means of STI avoidance among primates. Possible reasons for a lack of STI-dependent mate choice include sensory constraints, high costs of being choosy (e.g. aggressive behaviour or opportunity costs), host adaptations for dishonest signalling and counteradaptations by STIs to increase promiscuity or attractiveness, or to avoid detection (Knell and Webberley, 2004; Dass et al., 2011). For example, a viral STI that sterilises female crickets (*Gryllus texensis*) interferes with the production of immune-related proteins, which appears to be an adaptation by the parasite to prevent a reduction in host sexual activity (Adamo et al., 2014).

Many STIs are capable of remaining asymptomatic for long periods of time with no obvious effects on their hosts, but are still highly transmissible (e.g. chlamydia, HIV), potentially rendering mate inspection ineffective. Even if mate inspection is futile, we might still expect STIs to shape mating strategies in a number of ways. For example, females may choose lower quality mates as they are less likely to be infected than more popular males (Thrall et al., 2000; Ashby and Gupta, 2013). Similarly, high disease incidence may select against promiscuity (Boots and Knell, 2002). Nevertheless, in light of my results, one might speculate that some STIs have evolved to become asymptomatic as even a very weak preference for healthy mates can significantly reduce disease incidence (figure 7.1). For example, syphilis was highly virulent and produced obvious symptoms when it first appeared in Europe, but rapidly evolved to become milder and less conspicuous, possibly due to mate

inspection by humans (Knell, 2004). Unlike other parasites that can be transmitted through the environment or through social contact, STIs have far fewer opportunities to find a new host so face much greater selection in response to disease-avoidance traits, which may explain why STIs are more likely to be asymptomatic than non-STIs (Lockhart et al., 1996).

Much of the literature on PMSS has been motivated by the need to explain the evolution of extravagant male ornaments (Hamilton and Zuk, 1982; Loehle, 1995; Able, 1996; Loehle, 1997; Knell, 1999; Howard and Lively, 2003), but many species still exhibit mate choice in the absence of showy traits. As such, I have not incorporated showy traits in my models, but the results do suggest that STIs could facilitate their evolution if they signal health more accurately and are not too costly. However, it is rather doubtful that showy traits such as bright plumage will be significantly affected if STIs damage reproductive organs, as is commonly the case (Lockhart et al., 1996; Knell and Webberley, 2004). Still, some STIs do increase host mortality or reduce competitive ability, so it is possible that these non-sterilisers are more likely to select for showy traits. Here, I have focused on STIs that do not affect host mortality, but I expect my models would produce broadly similar patterns if this were not the case, as many of the principles that regulate the coevolutionary dynamics of hosts and sterilising STIs would still be applicable (e.g. preference for healthy mates will select against virulence, low virulence will select against mate inspection, etc). It remains to be seen whether host-STI coevolution will select for showy traits under these conditions.

The models presented herein did not incorporate behavioural or physiological differences between males and females, so the population could effectively be treated as hermaphroditic for the sake of simplicity. However, differences in mating preferences and risk of infection often exist between the sexes, which can have a significant influence on epidemiological and evolutionary dynamics (Thrall et al., 2000; Ashby and

Gupta, 2013). I have also assumed that mate inspection and virulence are smooth functions, that STIs must become less virulent to avoid detection and that virulence is correlated with transmissibility, but it is likely that these relationships will be more complex in reality. In particular, thresholds will exist below which STIs are undetectable, and greater transmissibility may not always lead to higher virulence or more obvious signs of infection.

As mentioned above, showy traits or other behavioural adaptations (e.g. post-copulatory urination; Nunn, 2003) may have evolved to reduce the risk of contracting STIs, but have not been included in the present study as they would greatly increase the complexity of the models. Similarly, I have not included any behavioural manipulation of the host by the parasite, which could increase mating frequency to facilitate more opportunities for transmission (Knell and Webberley, 2004). Finally, it is likely that parasitism plays some role in mate choice, but other factors such as inbreeding avoidance, dominance hierarchies and shared parental care may be equally important. Despite these obvious caveats, the dynamics observed in my models are likely to be quite general (e.g. fluctuating selection), as are the mechanisms that drive them (e.g. ecological feedbacks). I aim to address some of these issues in future work, especially how differences between the sexes influence coevolutionary dynamics.

Chapter 8

Effects of information exchange and infectious disease on host sociality

Ben Ashby[1] and Damien Farine[1]

Abstract

Sociality is known to facilitate both the spread of beneficial information (e.g. foraging or nesting sites, predation risk) and infectious disease, but little is known about how these processes combine to shape the evolution of different social strategies. Here, we use a stochastic individual-based model that captures fundamental properties of a well-studied empirical system of wild songbirds (*Parus major*) to investigate whether optimal social strategies are dependent on both the benefits of social information and the costs of infectious diseases. Our model exhibits multiple switches in the optimal social strategy as disease incidence increases, which can be explained by varying success in disease-avoidance. Specifically, individuals with high contact turnover rates are typically favoured when disease is either very rare or very common, whereas individuals with low contact turnover rates are favoured between these extremes. These findings are in good agreement with previous studies that have incorporated fixed rather than dynamic benefits of sociality, and have important implications for our

[1]Department of Zoology, University of Oxford, Oxford, UK

understanding of how sociality evolves in the presence of opposing selective pressures.

8.1 Introduction

Many animal populations exhibit strong variation in sociality, with some individuals preferring fewer, more stable associations (e.g. pair bonds) while others experience a greater number of interactions, which may be more transient (e.g. roaming individuals) (Watters and Sih, 2005; Pike et al., 2008; Croft et al., 2009; Tanner and Jackson, 2012; Aplin et al., 2013; Weber et al., 2013). Optimal behavioural strategies are likely to depend on a large number of factors, as greater sociality is often associated with many benefits (e.g. increased mating opportunities or information exchange) and costs (e.g. greater competition for resources or higher risk of disease) (Krause and Ruxton, 2002). Moreover, what constitutes an optimal strategy under one set of conditions may be a poor choice under different circumstances. For example, many populations are composed of isolated individuals that only become gregarious during mating seasons (e.g. lekking behaviour) or when food is scarce (e.g. seasonal flocking), which suggests that relationships between the costs and benefits of sociality are dynamic. Crucially, contact patterns are known to influence selection for various host and parasite traits (e.g. altruism, Jansen and van Baalen, 2006; Débarre et al., 2012; resistance, Best et al., 2011; virulence; Boots and Sasaki, 1999; Ashby and Gupta, 2013), so it is important to understand how changes in environment shape social behaviour.

Sociality is particularly important for animals that rely on information from other members of the population, often regarding foraging or nesting sites, or predation risk (Doligez et al., 2002; Danchin et al., 2004; Dall et al., 2005; Valone, 2007). For example, blue tits (*Cyanistes caeruleus*) use information about local breeding success when choosing whether to emigrate from the population or not (Parejo et al., 2007),

and in social animals, individuals can benefit from the collective vigilance of the group to provide information about predator attacks (Beauchamp, 2008). The transmission of information can occur through active signalling (Elgar, 1986) or via inadvertent cues (Pöysä, 1992; Galef and Giraldeau, 2001; Reader et al., 2003; Laidre, 2010), with highly social individuals likely to receive information sooner than poorly connected members of the population (Aplin et al., 2012). Although some individuals may try to maximise their ability to receive information from other members of the population through greater sociality, such behaviour is also likely to increase the risk of contracting and spreading infectious diseases. The risk of infection between social contacts will inevitably depend on multiple factors, including the method of transmission and the duration, frequency and proximity of contact, but in general, highly-connected members of a population and individuals with rapid contact turnover rates usually experience greater exposure to disease than those that form fewer contacts or have more stable associations (Woolhouse et al., 1997; Lloyd-Smith et al., 2005; Eames and Keeling, 2006; Ashby and Gupta, 2013). Hence, pathogens may constrain the social behaviour of their hosts, which has important ecological and evolutionary consequences for both populations (Prado et al., 2009). In contrast, without any turnover in contacts, individuals may not be able to access adaptive innovations or new information generated by others outside their social group. An outstanding question, which we address in the present study, is whether individuals should attempt to maintain stable associations with their current set of contacts, or seek out new contacts. Moreover, are contrasting social strategies able to coexist?

Parasitism has long been thought to influence sociality and has received much attention in the context of mate choice and sexual selection (Freeland, 1976; Hamilton and Zuk, 1982; Loehle, 1995, 1997; Boots and Knell, 2002; Kokko et al., 2002; Ashby and Gupta, 2013; chapter 7 herein), but relatively little is known about how individuals balance the need for social information with the threat of infectious dis-

eases. Previous studies that have investigated the spread of social information and infectious diseases have explored situations where these two processes are intrinsically linked, with a particular focus on behavioural changes in human populations during epidemics (see Funk et al., 2010 for a review of the recent literature). In this context, awareness of social information allows members of the population to take preventative measures, such as vaccination or reduced sociality to mitigate disease risk. Although it is plausible that some animals use social information about disease to adjust their behaviour, information is likely to be related to specific individuals (e.g. exclusion from social groups; Freeland, 1976) rather than to awareness of disease prevalence at a community or population level. Instead, social information in animal populations is probably dominated by gathering knowledge of foraging or nesting sites and of predation risk (Doligez et al., 2002; Danchin et al., 2004; Dall et al., 2005; Valone, 2007).

The spread of information and infectious disease are both dependent on the patterns of interaction between individuals, i.e. the structure of the social network, of the population. Yet, given that the two processes are likely to occur over contrasting timescales, these may exert varying selective pressures on social behavioural strategies. For example, information about predation or resource availability is expected to be of high value that rapidly decays over time (between seconds and days), whereas the risk of infectious disease may vary more slowly and over longer periods of time (e.g. weeks, months or years). These disparate timescales may influence selection in different ways and could lead to the evolution of multiple social strategies, some of which prioritise access to short-term information, whereas others minimise disease risk or strike a balance between the two. The success of a particular social strategy will depend on several factors, including how other individuals in the population choose to socialise, the fidelity of information transmission, the timescale over which information is valuable and the risk or severity of infectious disease. As such, explor-

ing how environmental changes (e.g. food availability, pathogen transmissibility or virulence) affect the prevalence of different social strategies is crucial for improving our understanding of how sociality evolves.

Here, we examine how social information and infectious disease affect the prevalence of two contrasting social strategies. Individuals in our model are classed as either 'reactive' (low contact turnover rate) or 'proactive' (high contact turnover rate), but are otherwise identical. These strategies are based on observations of wild songbirds (*Parus major*) (Aplin et al., 2013). We show that the optimal social strategy is tightly linked to a number of factors, most notably the prevalence of infectious disease and the relative benefits of social information.

8.2 Methods

8.2.1 Model description

We model the transmission of ephemeral social information (e.g. foraging information or predation risk) and a directly-transmitted infectious disease using a stochastic individual-based model (IBM) where each member of the population has at most n social contacts at any given time. Two members of the population that are social contacts can transmit and receive information, which leads to a temporary reduction in mortality, and infectious disease, from which hosts are unable to recover and leads to a higher mortality rate. As such, the contact network is identical for both information exchange and disease transmission. Each individual has a natural mortality probability of μ per time step, but infected individuals suffer an increased mortality probability of $\alpha + \mu$, where α represents the virulence of the pathogen. Awareness of survival-related information temporarily reduces the probability of mortality by μ (i.e. to α or 0 for infected and uninfected individuals, respectively). The value of information disappears with probability d per time step (e.g. food sites disappear or

predators are no longer a threat), at which point the probability of mortality per time step returns to $\alpha + \mu$ or μ for infected and uninfected members of the population, respectively.

Individuals conform to one of two heritable social strategies, hereafter referred to as 'reactive' (R) and 'proactive' (P). Reactive individuals attempt to maintain their current set of social contacts, whereas proactive individuals are prone to searching for new contacts to replace old ones. Hence, the population can be represented by a dynamic network with two types of nodes that differ in their contact turnover rates. Links between social contacts are broken upon death or if proactive individuals decide to seek a new contact. Thus, reactive-reactive links $[RR]$ only end upon death, proactive-reactive links $[PR]$ end upon death or if the proactive individual seeks a new contact and proactive-proactive links $[PP]$ end upon death or if either individual seeks a new contact. Following this process, all individuals that have less than n contacts are randomly assigned new contacts until no further links are possible. This means that reactive and proactive individuals have the same average number of concurrent contacts over the course of their lifetimes and only differ in their contact turnover rates. We do not incorporate preferential attachment or community structure into our model as these are typically driven by a range of non-social traits (Farine, 2014).

We update the population synchronously, with the following events occurring sequentially in each time step:

Independent discovery of information. Each individual that is currently unaware of the survival-related information discovers it independently with probability a, and subsequently has their probability of mortality reduced by μ.

Social information transmission. Individuals that are currently unaware of the survival-related information learn about it through social transmission with probability $1 - (1 - b)^c$, where b is the probability of

social information transmission per contact and c is the total number of contacts that individuals have that are currently aware of the information. Upon learning the information, individuals have their probability of mortality reduced by μ.

Mortality. Members of the population are randomly killed according to their probability of mortality (0, α, μ, or $\alpha + \mu$, depending on infection status and awareness of survival-related information). Links with other individuals are immediately terminated.

Births. The number of new offspring produced is drawn from a Poisson distribution with mean $rN(1-N/K)$, where r is the maximum per-capita birth rate, and N and K are the current and maximum population sizes, respectively. Offspring are assigned to be either reactive or proactive probabilistically, based on the relative abundance of the two social strategies in the population. The probability that a newborn individual will follow the proactive strategy is equal to:

$$\frac{N_P(1-2\epsilon)}{N_P+N_R}+\epsilon \tag{8.1}$$

where N_i is the current abundance of individuals following strategy i and ϵ is the mutation rate between strategies.

Link dissolution. Proactive individuals break links with each contact with probability p. Reactive individuals do not initiate contact dissolution.

Contact formation. All individuals that have less than n contacts are randomly assigned additional contacts until no further links are possible. Only one link is allowed between contacts.

Infection. Susceptible individuals become infectious with probability $1-(1-\beta)^I$, where β is the probability of pathogen transmission per contact and I is the total number of contacts that are infectious. Susceptible individuals are also challenged by an external force of infection (probability κ of becoming infectious, independent of social contacts). Newly infected individuals have their probability of mortality increased by α.

Information reset. Current survival-related information disappears with probability d. All individuals have their probability of non-disease-associated mortality reset to μ and are eligible to discover a new piece of information during the next time step.

8.2.2 Analysis

We focus our attention on the effects of the following parameters on the success of reactive and proactive social strategies: social information transmission (b); social information reset (d); infectious disease transmission (β); and virulence (α). We conduct 100 simulations for each parameter combination and record the frequency of each strategy and the prevalence of infection (number infected divided by population size) after 10,000 time steps. Simulations where one strategy has gone extinct before 10,000 time steps are terminated early. A strategy is deemed to be dominant if it accounts for at least 90% of individuals in the population. If neither strategy is dominant, then they are classed as coexisting. Each simulation is initiated with $K/2$ susceptible individuals for each strategy.

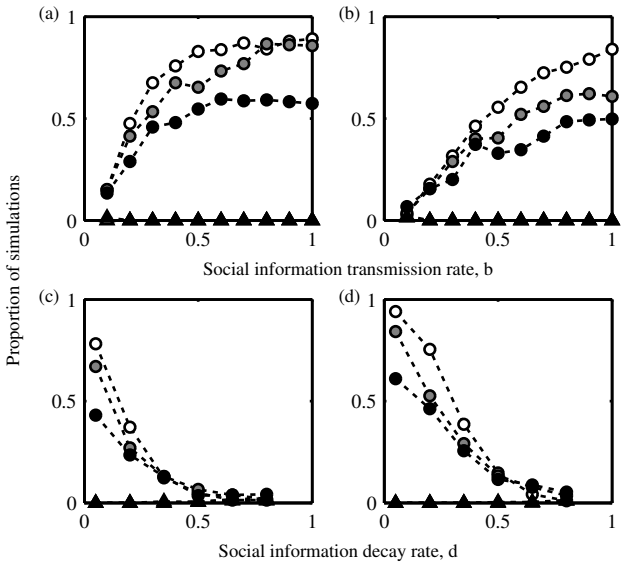

Figure 8.1: Host evolution in the absence of an infectious disease. Graphs show the proportion of simulations where reactive (triangles) and proactive (circles) individuals are at over 90% prevalence, with $n = 2$ (white), n=3 (grey) and $n = 4$ (black). (a) $d = 0.1$, (b) $d = 0.25$, (c) $b = 0.25$, (d) $b = 0.5$. Fixed parameters: $\alpha = 0$, $\beta = 0$, $\epsilon = 0.005$, $\kappa = 0$, $\mu = 0.05$, $a = 0.02$, $K = 500$, $p = 0.5$ and $r = 1$.

8.3 Results

8.3.1 No disease

We began by analysing the effects of social information transmission (b) and reset (d) on the maintenance of reactive and proactive strategies in the absence of disease. We found that the proactive strategy tends to dominate much more often than the reactive strategy, as it better facilitates the learning of survival-related information. Higher probabilities of social information transmission (b) and lower probabilities of information decay (d) were associated with a greater advantage for the proactive strategy (figure 8.1), as these factors allowed proactive individuals to learn information quicker and for the benefits to last longer. In addition, a lower maximum number

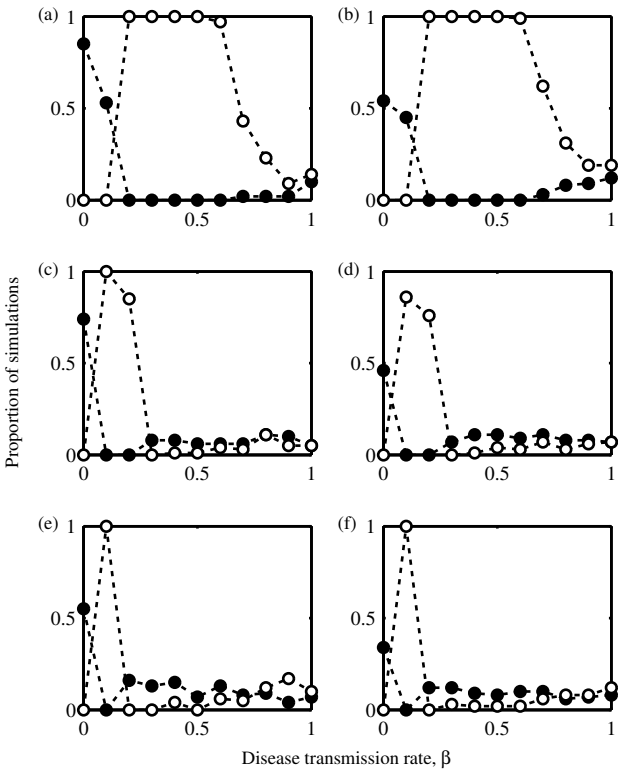

Figure 8.2: Host evolution in the presence of an infectious disease (low virulence). Graphs show the proportion of simulations where reactive (white) and proactive (black) individuals are at over 90% prevalence. The maximum number of contacts varies between rows, with (a, b) $n = 2$, (c, d) $n = 3$ and (e, f) $n = 4$. The decay rate of information varies between columns, with (a, c, e) $d = 0.1$ and (b, d, f) $d = 0.25$. Fixed parameters as specified in figure 8.1, with $\alpha = 0.2$ and $\kappa = 0.001$ (for $\beta > 0$).

of contacts (n) generally increased the difference in success between the two strategies, as this was associated with an increase in the average path length, so reactive individuals were less likely to learn information.

8.3.2 With disease

The introduction of an infectious agent led to drastic changes in how well the two social strategies fared (figures 8.2-8.4), which can be explained by the overall prevalence (and severity) of disease (figure 8.5) and the relative benefits of social information.

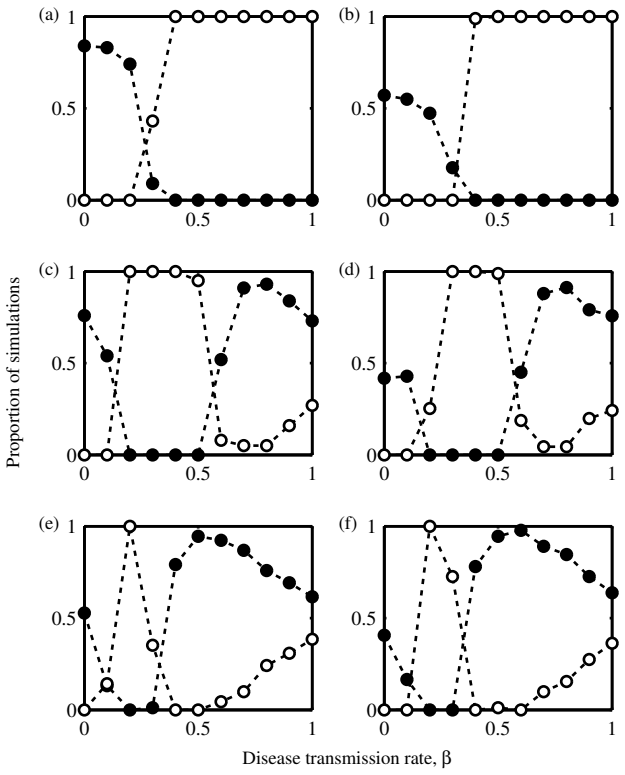

Figure 8.3: Host evolution in the presence of an infectious disease (moderate virulence). Graphs show the proportion of simulations where reactive (white) and proactive (black) individuals are at over 90% prevalence. The maximum number of contacts varies between rows, with (a, b) $n = 2$, (c, d) $n = 3$ and (e, f) $n = 4$. The decay rate of information varies between columns, with (a, c, e) $d = 0.1$ and (b, d, f) $d = 0.25$. Fixed parameters as specified in figure 8.1, with $\alpha = 0.4$ and $\kappa = 0.001$ (for $\beta > 0$).

When disease transmission was unlikely (small β), the risk of infection was generally low so proactive personalities tended to dominate for the same reasons described in the previous section (i.e. they were better placed to receive information). For intermediate transmission rates, disease was more common and the risk of infection was greater. Although reactive individuals were less likely to learn survival-related information due to a lower contact turnover rate, this behaviour also limited their exposure to disease, so they tended to dominate. Somewhat counter-intuitively, however, the proactive strategy was often favoured once again when transmission rates

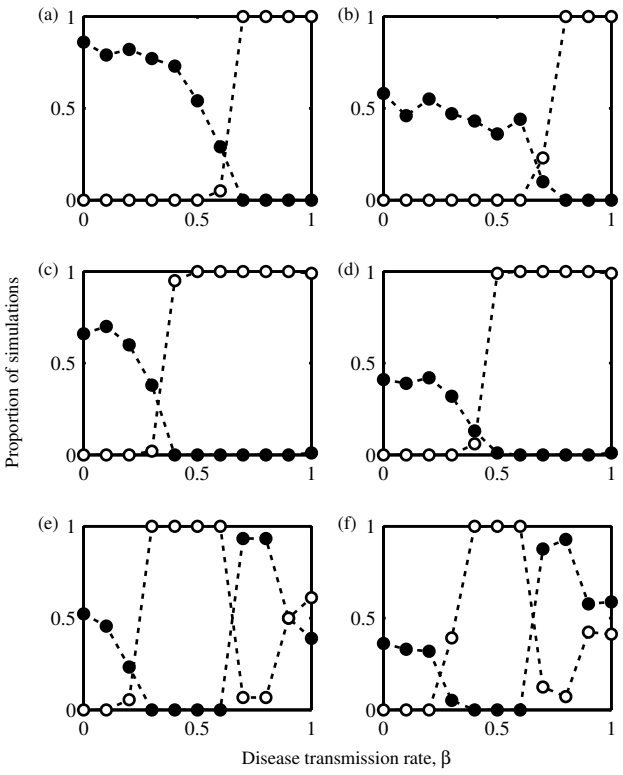

Figure 8.4: Host evolution in the presence of an infectious disease (high virulence). Graphs show the proportion of simulations where reactive (white) and proactive (black) individuals are at over 90% prevalence. The maximum number of contacts varies between rows, with (a, b) $n = 2$, (c, d) $n = 3$ and (e, f) $n = 4$. The decay rate of information varies between columns, with (a, c, e) $d = 0.1$ and (b, d, f) $d = 0.25$. Fixed parameters as specified in figure 8.1, with $\alpha = 0.6$ and $\kappa = 0.001$ (for $\beta > 0$).

were high. This is because a large proportion of the population was usually infected when β was large (figure 8.5), so all individuals were at high risk of infection, regardless of social strategy. In this region of parameter space, proactive individuals were able to learn ephemeral information through social contacts more quickly than reactive individuals, which was sufficient to offset a slightly higher risk of infection. Finally, when transmission rates were very high almost everyone in the population was infected, which led to a significant reduction in population size and increased the likelihood of stochasticity driving one strategy extinct. Hence, while the proactive

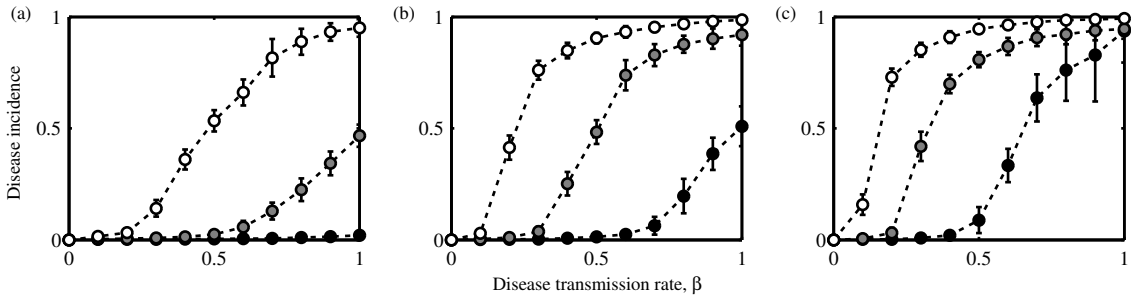

Figure 8.5: Average disease incidence (±1 standard deviation) for (a) $n = 2$, (b) $n = 3$ and (c) $n = 4$. Circles correspond to mean prevalence for different levels of virulence: $\alpha = 0.2$ (white), $\alpha = 0.4$ (grey) and $\alpha = 0.6$ (black). Fixed parameters as specified in figure 8.1, with $\kappa = 0.001$ (for $\beta > 0$).

strategy still tended to be optimal, the proportion of simulations where it dominated often decreased for very high β due to stochasticity (e.g. figures 8.3-8.4, e-f).

The maximum number of contacts (n) had a considerable impact on the optimal social strategy. In most cases, larger values of n greatly increased disease prevalence (figure 8.5) and in turn reduced the conditions under which the reactive strategy was favoured. This finding contrasts with the result outlined in the previous section when no disease was present. In the absence of disease, a greater number of contacts benefited the reactive strategy by reducing the average path length, making survival-related information more easily accessible (i.e. the performance gap between reactive and proactive strategies was lessened). However, a greater number of contacts also led to higher disease incidence in the population, so all individuals experienced a greater risk of infection. This allowed proactive individuals to dominate over a broader range of conditions as n increased, because they were better able to obtain survival-related information, which was sufficient to offset a slightly higher risk of infection.

Virulence (α) had two contrasting effects on epidemiological and evolutionary dynamics. At an individual level, greater virulence was more costly, but at a population level it reduced disease incidence due to a shorter infectious period on average. Hence,

the reactive strategy tended to become more prominent as virulence increased, because disease incidence was maintained at moderate levels over a greater region of parameter space (figure 8.5). Low virulence was generally associated with weak selection so tended to produce small regions of dominance and large regions of coexistence (figure 8.2), whereas pathogens that were moderately (figure 8.3) or highly (figure 8.4) virulent exerted much stronger selection on the host, so coexistence was rarer.

8.4 Discussion

We explored a model of information and disease transmission to determine their effects on the evolution of host sociality. Despite the simplicity of our model, we found that ecological parameters associated with information use and disease spread could profoundly influence the social structure of animal populations by selecting for and against different social strategies. Previous models have explored how social information and infectious disease influence the contact structure of a population (see Funk et al., 2010 and references therein), but we are unaware of any studies where these two processes have been treated as selective pressures that operate independently. Our model demonstrates that optimal social strategies are dependent on both the accessibility of social information and the risk of exposure to infectious disease. For the most part, our findings are intuitive; rapid contact turnover rates are favoured if there is a low risk of disease and slow contact turnover rates are favoured if there is a moderately high risk of disease. However, we also observed that rapid contact turnover rates may be favoured when transmission rates are high, in accordance with previous findings (Bonds et al., 2005). This scenario is somewhat counter-intuitive, but is analogous to a tragedy of the commons, as proactive members of the population increase overall disease incidence to the point where reactive members are at a disadvantage. Bonds et al. (2005) likened these dynamics to optimal host investment

in an immune response, which is predicted to peak when the risk of infection is neither too low, nor too high (van Baalen, 1998). We also found large regions of coexistence between social strategies, especially when virulence was low (figures 8.2c-f). Overall, our findings highlight how subtle changes in the environment can influence the social structure of animal populations, ranging from highly stable groups to randomly-mixed populations. For example, changes in habitat quality or availability will alter the relative value of social information, potentially leading to changes in sociality, and ultimately, pathogen life-history traits such as virulence. Similarly, we should expect the introduction or evolution of pathogens that are more transmissible or virulent to lead to shifts in host social structure (Bonds et al., 2005; Prado et al., 2009), or even extinction if alternative social strategies cannot be adopted quickly enough.

Many studies have explored the role of infectious diseases in the evolution of host sociality, but have typically been limited to mating behaviour (Freeland, 1976; Hamilton and Zuk, 1982; Loehle, 1995, 1997; Boots and Knell, 2002; Kokko et al., 2002; Ashby and Gupta, 2013; chapter 7 herein). Meanwhile, the influence of disease on more general social behaviour has tended to focus on host plasticity (avoidance of infected individuals) in response to epidemics rather than evolutionary changes in sociality (Funk et al., 2010). We are aware of only two other studies, by Bonds et al. (2005) and Prado et al. (2009), where the evolution of sociality has been explored in the context of infectious diseases. Our work is closely linked to these studies, but differs in a number of important ways.

First, we have examined a different aspect of sociality (contact turnover rate) to Bonds et al. (2005) and Prado et al. (2009), who defined sociality to be the number of concurrent contacts per individual. The latter approach is a sensible way of defining sociality, but misses a key aspect of social behaviour, as individuals may differ in the rate at which they associate with new individuals. In our model, proactive individuals are more adept at learning social information due to their higher contact turnover

rate, which allows them to come into contact with a greater number of individuals over the course of their lifetimes, even with the same average number of concurrent contacts as reactive members of the population. Contact turnover rates are especially important when considering empirical networks, as isolated snapshots will not be able to distinguish between transient and more stable associations (Aplin et al., 2013).

Second, the aforementioned studies assumed that sociality conferred fixed benefits in the form of improved survival in the absence of disease. Here, we have taken an alternative approach, whereby the advantages of sociality are ephemeral and depend entirely on the availability of social information. Hence, social contacts are more beneficial if they allow valuable information to be exchanged; once an individual has passed on information, they may no longer be useful as a contact, at least for a short period of time. Again, this is an important consideration, as the benefits of sociality are unlikely to be fixed and could vary seasonally or with other factors such as population size (Krause and Ruxton, 2002).

Finally, both studies considered coevolutionary scenarios, where feedbacks between changes in host sociality and pathogen virulence led to a stable equilibrium (Bonds et al., 2005) or cycling in both traits (Prado et al., 2009). We did not include pathogen evolution for the sake of simplicity, but there is good reason to believe that our model would respond in a similar manner to the one proposed by Bonds et al. (2005) if we invoked a trade-off between pathogen transmissibility and virulence. The two models produced comparable evolutionary outcomes for the host, with sociality (albeit defined differently) minimised for intermediate levels of disease. Still, some form of cycling may be possible due to our different definition of sociality or due to the explicit contact structure used in our model (similar to Prado et al., 2009), which was not captured by the mean-field approach of Bonds et al. (2005). We hope to explore coevolutionary scenarios involving social information in future work.

We used random mixing rather than imposing specific contact structures on our

populations, as we wished to make as few assumptions as possible to maintain the generality of our model. However, extra layers of complexity in network structure, such as clustering, preferential attachment and different distributions in the number of contacts per individual have been observed in many populations (Croft et al., 2005; Lusseau et al., 2006; Guttridge et al., 2011; Farine et al., 2012; Farine, 2014) and are known to influence both epidemiology (Eames and Keeling, 2002; House, 2011; Hock and Fefferman, 2012; Ashby and Gupta, 2013; chapter 7 herein) and evolution (Read and Keeling, 2003; Eames and Keeling, 2006; Ashby and Gupta, 2013). We might expect these or other forms of network heterogeneity to evolve in response to changes in the value of social information and the prevalence of infectious diseases, which will in turn affect transmission through the network. Furthermore, as social information and infectious diseases typically exist over different timescales, increased network heterogeneity may have profoundly different impacts on these processes.

We did not attempt to capture the dynamics of a specific system, as we wanted our findings to remain general, but we did base several aspects of our model on populations of wild songbirds, where individuals conform to proactive and reactive social strategies that govern contact turnover rates, similar to those modelled in the present study (Aplin et al., 2013). In addition, these populations form fission-fusion groups during winter months that correlate with the discovery of ephemeral food sites (Aplin et al., 2012). The fact that flocks of great tits are only formed during winter months when food is scarce suggests that sociality incurs a cost, potentially in the form of increased exposure to infectious diseases. Few studies have explicitly linked infectious diseases to social networks in animal populations, but several authors have recently found strong correlations between the two. For example, Vanderwaal et al. (2014) were able to map subtypes of the bacterium *Escherichia coli* to a social network of wild giraffes (*Giraffe camelopardalis*), which provided a better explanation for transmission than space-use patterns. Similarly, the social network positions of badgers (*meles meles*)

have been shown to correlate with test results for bovine tuberculosis (*Mycobacterium bovis*) (Weber et al., 2013). Together, these findings suggest that the structure of social networks in animal populations is likely to be crucial for the spread of both information and infectious disease.

Chapter 9

Summary and future directions

The antagonistic relationship that exists between hosts and parasites can provide conditions for rapid evolutionary and coevolutionary dynamics that result in the emergence of novel traits, such as resistance or virulence, and can lead to patterns of diversity across time and space. A number of factors are known to have significant effects on host and parasite evolution, but genetic specificity (i.e. genotype-by-genotype interactions) and population mixing patterns are thought to be especially important, as they usually have a strong influence on ecological dynamics and selection (Parker, 1994; Boots and Sasaki, 1999; Haraguchi and Sasaki, 2000; Agrawal and Lively, 2002; Otto and Nuismer, 2004; Eames and Keeling, 2006; Best et al., 2011). Together, the studies that constitute this thesis have explored how these genetic and environmental factors influence the evolution of various traits among hosts and parasites, including resistance and infectivity ranges, sexual reproduction, mate choice, virulence and sociality. Here, I summarise and discuss the main findings of the preceding chapters and suggest directions for future research.

9.1 Genetic specificity of host-parasite interactions

The role of genetic specificity in host-parasite coevolution was primarily explored in chapters 2 and 4. Chapter 2 focused on how the evolution of broader infectivity ranges among parasites is affected by genetic factors, namely epistasis and asymmetries in how hosts and parasites accumulate mutations that confer broader ranges. Gradual changes in host phenotype were shown to provide the best conditions for infectivity

range expansion when epistasis was strongly positive (i.e. parasites must accumulate multiple mutations to perform well on resistant hosts), in line with experiments involving bacteria and phages (Paterson et al., 2010; Hall et al., 2011a; Meyer et al., 2012). Conversely, sudden changes in host phenotype resulted in a low probability of parasites fixing mutations that conferred broader ranges under this form of epistasis. However, if the interactions between infectivity mutations were characterised by weaker forms of epistasis, then sudden rather than gradual changes in host phenotype were shown to provide better conditions for parasite range expansion due to a trade-off between selection and mutation supply. Hence, chapter 2 shows that there can be difficulties with generalising empirical observations from one system to others, as they may be governed by different genetic interactions that lead to qualitative changes in evolutionary outcomes. In addition, the models presented in chapter 2 demonstrate that the effects of stochasticity need to be considered in host-parasite models where multiple mutations accumulate over time, as extinctions of rare genotypes can prevent alleles from reaching fixation.

Chapter 4 demonstrated that a novel framework for genetic specificity could explain a wider set of coevolutionary dynamics than previous models. Most existing models of host-parasite specificity that allow broader resistance and infectivity ranges to evolve are based on either the gene-for-gene (GFG) or inverse gene-for-gene (IGFG) frameworks, but a number of empirical studies indicate that these frameworks do not fully capture the coevolutionary dynamics of real populations. For example, bacteria and phages typically exhibit directional selection as part of a coevolutionary arms race (Buckling and Rainey, 2002a; Mizoguchi et al., 2003; Hall et al., 2011b; Scanlan et al., 2011), but GFG models with explicit genetics predict that fluctuations in range will occur (Sasaki, 2000; Fenton and Brockhurst, 2007; Fenton et al., 2009, 2012). A key feature of the GFG and IGFG models is their assumption that only one population can trigger an arms race (GFG: hosts, IGFG: parasites), which means

that the other population is reactive (GFG: parasites, IGFG: hosts). In chapter 4, an alternative model (referred to as a 'symmetric gene-for-gene' (SGFG) framework) was presented, where arms races can be triggered by both hosts and parasites. Analysis of the SGFG framework revealed that it can produce a variety of coevolutionary dynamics that are more consistent with empirical observations of directional selection for broader ranges. Moreover, the inclusion of ecological feedbacks was shown to fundamentally change coevolutionary outcomes, which reinforces the need for more realistic models of host-parasite relationships.

However, experiments have shown that fluctuating selection between genotypes with similar ranges can also occur, either following an arms race (Hall et al., 2011b) or in its absence (Gomez and Buckling, 2011), which suggests that a second component may be involved in specificity. As such, a second stage was added to the infection process that was governed by matching alleles (MA) infection genetics. The extended (SGFG-MA) model was shown to produce fluctuating selection between genotypes of identical ranges, with the extent of range expansion depending on the strength of fitness costs. These dynamics closely match the experimental results of Hall et al. (2011b) and Gomez and Buckling (2011).

9.2 The role of contact patterns in host and parasite evolution

The remaining chapters in this thesis focused on the effects of contact patterns on host and parasite (co)evolution, ranging from simple models of spatial structure and sexual contact, to explicit models of social and sexual contact networks. Together, these chapters demonstrate the strong influence that mixing patterns can have on host-parasite relationships, with implications for the maintenance of sex according to the Red Queen Hypothesis (RQH), mating behaviour and the evolution of virulence.

Chapter 3 examined how spatial structure interacts with fitness costs associated with range expansion. Data from individual-based simulations of well-mixed and spatially structured environments indicated that the latter is more favourable to range expansion when this trait is costly. This is because (i) locally-optimal genotypes may differ from what is optimal at a global level, and (ii) clustering can limit the extent to which susceptible genotypes (with higher growth rates) re-emerge following the evolution of generalist hosts and parasites. As such, resistance often spread in spatially structured populations even when it was very costly, whereas well-mixed populations always experienced selection for the optimal genotype, which was highly dependent on fitness costs. This study has particular relevance to microbial communities, as bacteria often produce biofilms or live in highly structured environments (e.g. soil; Vos et al., 2009). Crucially, it demonstrates how spatial structure can have a strong impact on host and parasite evolution, in line with other theoretical and empirical work (Boots and Sasaki, 1999; Haraguchi and Sasaki, 2000; Brockhurst et al., 2003; Read and Keeling, 2003; Brockhurst et al., 2006; Lion and Boots, 2010; Best et al., 2011; Koskella et al., 2011).

Chapter 5 explored what characteristics of host-parasite interactions are most likely to produce sustained fluctuations in gene frequencies (coevolutionary cycling) and examined the implications for the maintenance of sex according to the RQH. Notably, this study differs from most other models of the RQH by using a continuous-time system with ecological feedbacks and diploid hosts, which represents a more realistic approach to modelling host-parasite coevolution. In the first part of this study, parasitic castration was shown to be necessary for cycling to persist indefinitely. In the second part, the effects of castration on competition between sexual and asexual populations were explored. In the absence of full castration, sexually-reproducing individuals did not experience an advantage over asexuals, as coevolutionary cycling was not maintained indefinitely. If parasites castrated their hosts, cycling persisted

and sex was shown to be beneficial provided multiple mating was sufficiently common. This is because recombination and segregation allow sexual populations to produce offspring that are genetically diverse and are more likely to avoid contemporaneous parasites, but the presence of a castrator reduces the availability of fertile mates and thus imposes a cost on sex; multiple mating reduces this cost. Previous models of parasitic castrators have based the reproductive output of the host on the average fertility of the population (Lively, 2010b), but chapter 5 shows that this approach may overstate the benefits of sex.

The remaining chapters (6-8) explored how sexual and social contact patterns evolve (or coevolve) with parasite virulence. Sexually transmitted infections (STIs) and mating behaviour were explored in chapters 6 and 7, and non-STIs and social contact patterns were explored in chapter 8. Chapter 6 used bipartite networks to represent asymmetrical (polygynous or polyandrous) mating systems through which an STI could spread. The results showed that male-biased mating systems can produce different epidemiological dynamics to female-biased mating systems, if decisions to move between mating groups are based on prior reproductive failure (movement was assumed to be more likely if one or both partners were infected). Moreover, mating patterns were shown to be crucial for the evolution of virulence. These results arise because females are usually better placed to assess their own reproductive success than males, so are more likely to correctly identify if their current mate is infertile and hence make a more accurate decision whether to 'stay or stray'. For the sake of simplicity, the host population was prevented from coevolving with the parasite in this study, but changes in STI virulence are likely to affect the selection for different mating strategies (as shown in chapter 7, discussed below). As such, further development of the model presented in chapter 6 is required to allow coevolution between hosts and parasites, which will be addressed in future work.

Chapter 7 did allow for the coevolution of host mating strategies and STI viru-

lence, but this was in the context of 'mate inspection' (a preference for healthy mates), rather than asymmetrical mating systems. This chapter was unique as it utilised both stochastic individual-based and deterministic pair formation models (IBMs and PFMs, respectively) to understand the coevolution of hosts and STIs. Coevolutionary behaviour was determined with the PFM using an adaptive dynamics approach, which produced a variety of dynamics that were later confirmed by simulations of the IBM. These ranged from coevolutionarily stable strategies (CoESSs) that could be approached monotonically or non-monotonically, runaway virulence, and cyclical dynamics that arose due to fluctuating selection. Interestingly, a bistability in the system meant that hosts were often able to retain a preference for healthy mates even when STIs evolved to be avirulent. These results have important implications for understanding the role that STIs play in sexual selection, as others have argued that disease-avoidance behaviour will select for avirulent STIs, which will subsequently remove sexual selection in the host (Knell, 1999). Here, I have shown that mate choice can be maintained even when the original selection pressure is subsequently removed.

Finally, chapter 8 explored the evolution of different social strategies when hosts are able to transmit infectious diseases as well as information that increases fitness. These opposing selection pressures result in multiple switches in the optimal social strategy as disease incidence increases, which can be explained by varying success in disease-avoidance. Individuals with high contact turnover rates are typically favoured when disease is either very rare or very common, whereas individuals with low contact turnover rates are favoured between these extremes. This study differs from previous work in a number of ways. In particular, models that have explored the evolution of sociality and infectious diseases have treated the benefits of sociality as being fixed (Bonds et al., 2005; Prado et al., 2009), whereas here, social contacts only confer an increase in fitness if beneficial information is exchanged. Existing models of information exchange have instead concentrated on the spread of information that is

directly related to disease, which can result in adaptive rewiring of contacts to avoid infection (Funk et al., 2010). In chapter 8, information is related to factors such as foraging, nesting or predation avoidance rather than the prevalence of disease, so the two processes are modelled independently. Moreover, they are modelled in an evolutionary rather than purely epidemiological context, unlike previous studies (Funk et al., 2010). As in chapter 6, coevolutionary scenarios were not explored, so it remains to be seen whether CoESSs or cyclical dynamics emerge (see Bonds et al., 2005; Prado et al., 2009).

9.3 Future directions

The models presented herein can be extended in many ways to answer further questions about host and parasite evolution, but three areas deserve particular attention. First, only a few models of genetic specificity have been developed by theoreticians and their dynamics have largely been analysed in the absence of ecological feedbacks. While these models roughly correspond to some genotype-by-genotype interactions under certain conditions, they are rarely explicitly evaluated in the context of empirical observations. Clearly, models that are only applicable to one system are of limited use, but chapter 4 demonstrates that there is considerable scope for developing general models of specificity that are still consistent with specific empirical observations. Future models of specificity will also need to carefully consider the role of ecological feedbacks, which have been shown to heavily influence coevolutionary dynamics. A better understanding of genetic specificity and its influence on coevolutionary dynamics is especially important for the use of biocontrol in industry and medicine (Tait et al., 2002; Levin and Bull, 2004). Second, many authors have suggested that parasites are likely to play an important role in the evolution of mating strategies and secondary sex traits (see chapter 7), but few have considered this topic in a coevo-

lutionary context. A greater appreciation for coevolution is needed in models that explore mating behaviour in response to infectious diseases. Finally, mathematical methods that approximate contact structure need to be used more widely as they provide a rigorous alternative to stochastic individual-based modelling. Here, only chapter 7 included a deterministic approximation for contact structure, which allowed for a better understanding of the model's behaviour. The results of chapters 3, 6 and 8 could potentially be extended through the use of similar methods, but their complexity may make this task less than straightforward. Still, future models that incorporate contact structure should aim to use deterministic approximations that are analytically more tractable.

9.4 Conclusion

Overall, this thesis has shown that genetic specificity and contact patterns are central aspects of host-parasite relationships that play vital roles in shaping evolutionary and coevolutionary dynamics. The topics that have been covered are of relevance to some of the most prominent and active areas of research in evolutionary biology (e.g. infection genetics, the maintenance of sex and factors that influence mate choice) and have important implications for future study in these areas. In particular, the models presented herein provide fresh impetus for future research on genetic specificity and parasite-mediated sexual selection. Throughout, a central theme has been the importance of making realistic assumptions about population dynamics, as ecological feedbacks are widely recognised as fundamental components of epidemiological models. Their frequent omission from evolutionary models is therefore somewhat of a mystery, and addressing this issue remains a key challenge for evolutionary biologists. The studies that constitute this thesis contribute a small part towards our understanding of host-parasite relationships, but many more questions remain.

Appendices

Appendix A

Supplementary material for chapter 2

A.1 Additional results

In the main text (chapter 2), we only present results for parasite populations with three loci ($n = 3$), contact patterns based on frequency-dependence (FD) and decelerating fitness costs ($\phi_H = \phi_P = 0.5$) affecting only the per-capita birth rate (r_i) and transmission coefficient (β_j). Here, we present the results for other numbers of loci (figure A.1), density-dependent contact patterns (figure A.2), linear and accelerating fitness costs (figure A.3) and different combinations of cost functions (figure A.4). These results are qualitatively similar to those presented in figures 2.3 and 2.4.

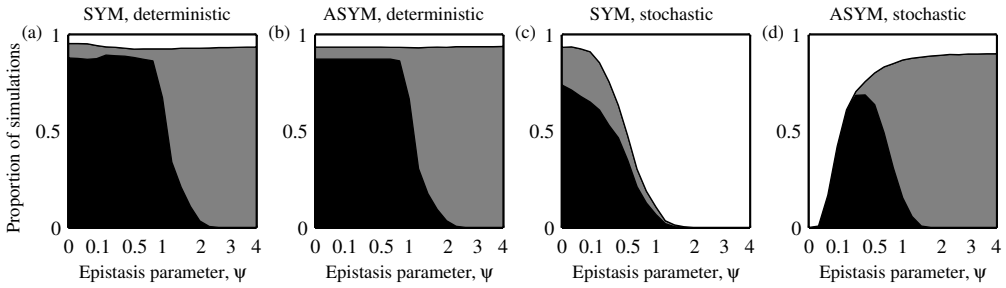

Figure A.1: Proportion of simulations where peak infectivity range (E; equation 2.4) was greater than 0.9 (black), less than 0.1 (white) or between these values (grey), for $n = 4$. Plots (a-b) and (c-d) show results for deterministic and stochastic simulations, respectively. Plots (a, c) and (b, d) show results for gradual (SYM) and sudden (ASYM) changes in host phenotype, respectively. The parameter ψ controls the type and strength of epistasis between infectivity alleles, ranging from strong positive ($\psi << 1$), through weak positive ($\psi < 1$), none ($\psi = 1$) and finally, negative ($\psi > 1$) epistasis. These results are broadly similar to those presented in the main text for $n = 3$ (figures 2.3 and 2.4)

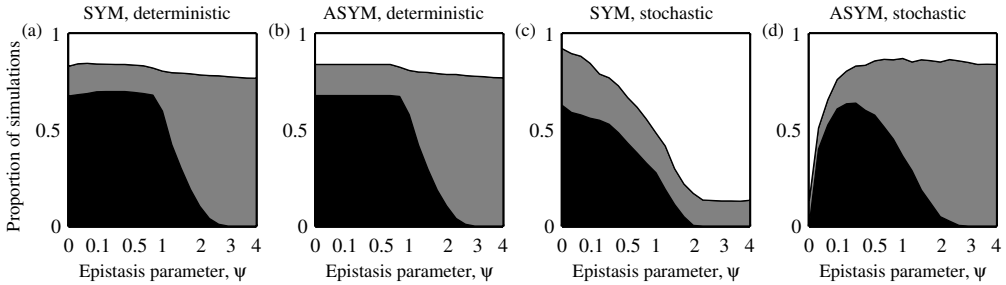

Figure A.2: Proportion of simulations where peak infectivity range (E; equation 2.4) was greater than 0.9 (black), less than 0.1 (white) or between these values (grey), for density-dependent transmission (DD). Plots (a-b) and (c-d) show results for deterministic and stochastic simulations, respectively. Plots (a, c) and (b, d) show results for gradual (SYM) and sudden (ASYM) changes in host phenotype, respectively. The parameter ψ controls the type and strength of epistasis between infectivity alleles, ranging from strong positive ($\psi << 1$), through weak positive ($\psi < 1$), none ($\psi = 1$) and finally, negative ($\psi > 1$) epistasis. These results are broadly similar to those presented in the main text for frequency-dependent transmission (figures 2.3 and 2.4)

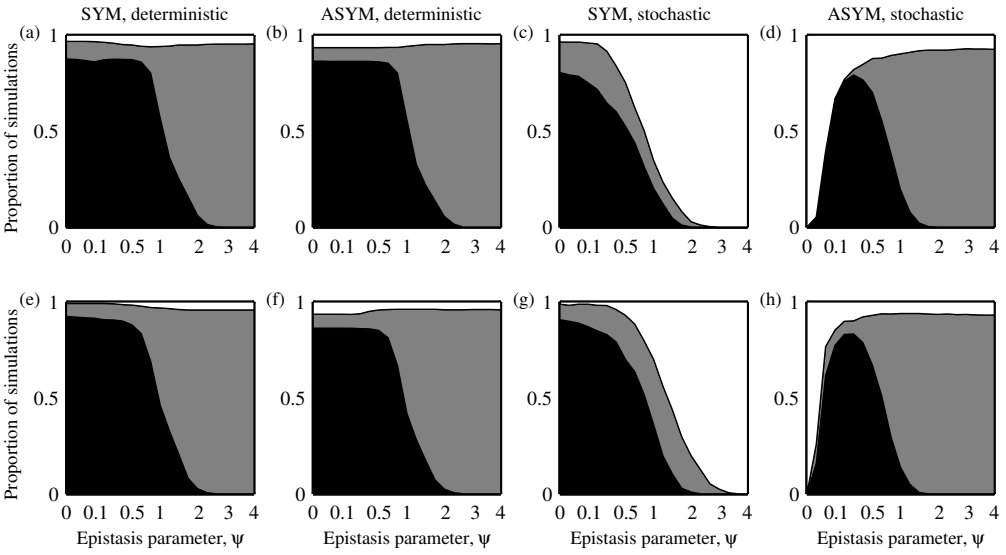

Figure A.3: Proportion of simulations where peak infectivity range (E; equation 2.4) was greater than 0.9 (black), less than 0.1 (white) or between these values (grey), for (a-d) linear ($\phi_H = \phi_P = 1$) and (e-h) accelerating ($\phi_H = \phi_P = 2$) fitness costs. The first and last two columns show results for deterministic and stochastic simulations, respectively. The first and third, and second and fourth columns show results for gradual (SYM) and sudden (ASYM) changes in host phenotype, respectively. The parameter ψ controls the type and strength of epistasis between infectivity alleles, ranging from strong positive ($\psi << 1$), through weak positive ($\psi < 1$), none ($\psi = 1$) and finally, negative ($\psi > 1$) epistasis. These results are broadly similar to those presented in the main text for decelerating fitness costs (figures 2.3 and 2.4)

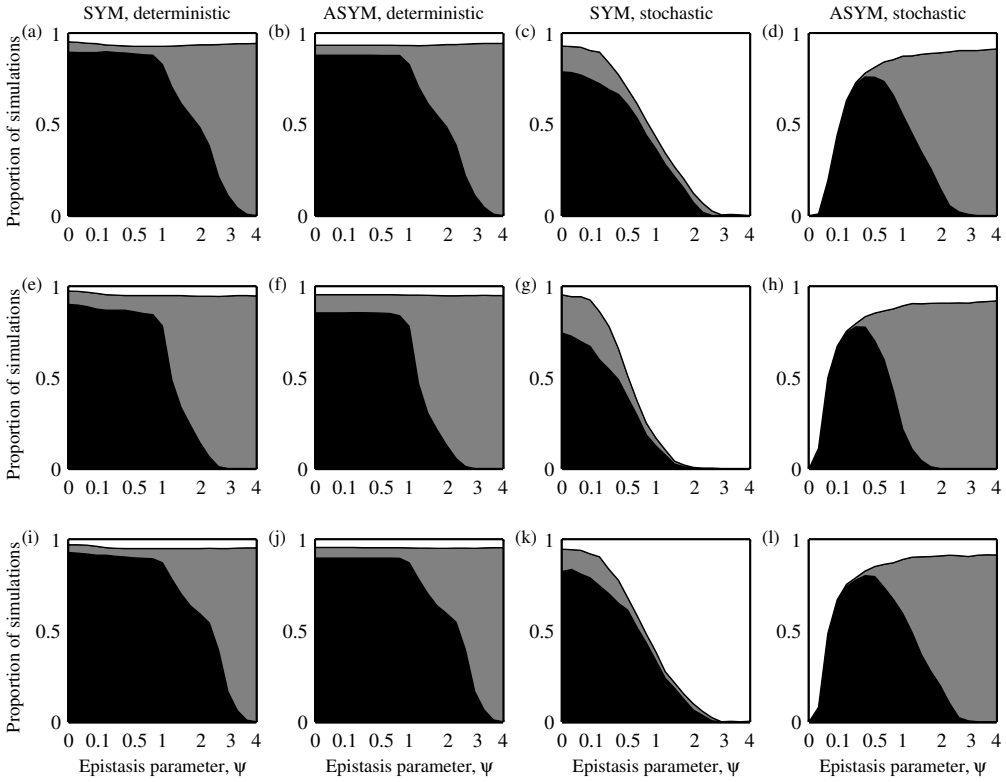

Figure A.4: Proportion of simulations where peak infectivity range (E; equation 2.4) was greater than 0.9 (black), less than 0.1 (white) or between these values (grey), for different combinations of fitness costs: (a-d) natural mortality (μ_i) and transmission (β_j) rates; (e-h) birth (r_i) and disease-associated mortality (α_j) rates; (i-l) natural (μ_i) and disease-associated (α_j) mortality rates. The first and last two columns show results for deterministic and stochastic simulations, respectively. The first and third, and second and fourth columns show results for gradual (SYM) and sudden (ASYM) changes in host phenotype, respectively. The parameter ψ controls the type and strength of epistasis between infectivity alleles, ranging from strong positive ($\psi << 1$), through weak positive ($\psi < 1$), none ($\psi = 1$) and finally, negative ($\psi > 1$) epistasis. These results are broadly similar to those presented in the main text for fitness costs in the form of reduced birth (r_i) and transmission (β_j) rates (figures 2.3 and 2.4).

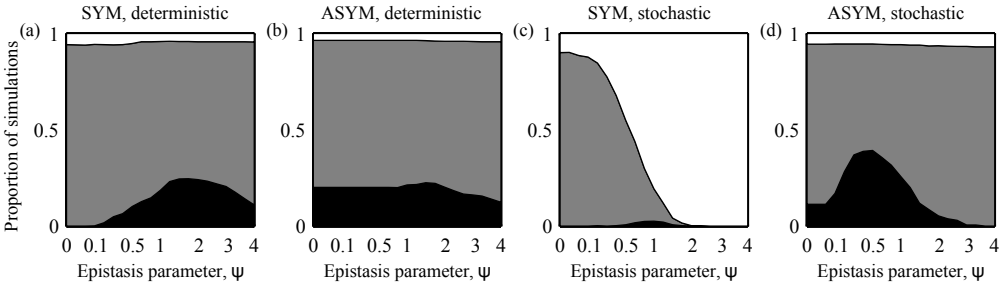

Figure A.5: Proportion of simulations where peak resistance range for hosts was greater than 0.9 (black), less than 0.1 (white) or between these values (grey). The measure used here for hosts is analagous to the one described for parasites in equation 2.4, with resistance substituted for infectivity. Plots (a-b) and (c-d) show results for deterministic and stochastic simulations, respectively. Plots (a, c) and (b, d) show results for gradual (SYM) and sudden (ASYM) changes in host phenotype, respectively. The parameter ψ controls the type and strength of epistasis between infectivity alleles, ranging from strong positive ($\psi << 1$), through weak positive ($\psi < 1$), none ($\psi = 1$) and finally, negative ($\psi > 1$) epistasis.

Appendix B

Supplementary material for chapter 3

B.1 Parameter space

The parameters sampled for the primary model were drawn from the uniform random distributions shown in table B.1.

B.2 Additional results

In the main text we only present data for $n = 3$, but we found our results to be qualitatively similar for different numbers of loci. Figures 3.4-3.5 show trends for resistance and infectivity ranges, respectively; here, we show corresponding figures for $n = 2$ (figures B.1, B.3-B.4) and $n = 4$ (figures B.2, B.5-B.6). Tables B.2-B.4 show the full results of our statistical analysis of the data using analysis of covariance (ANCOVA).

	minimum value	maximum value
α^*	$.05 \times \delta$	$.5 \times \delta$
β	20	50
δ	.25	1
ϵ_H	.001	.005
ϵ_P	.004	.02
η_P	.1	.5
σ	.2	.5
$\tau^\dagger$	1	7
D	.0625	.5
$T^\dagger$	1	7

Table B.1: Parameter distributions used to analyse the primary model in chapter 3.*Value chosen so that the ratio of α to δ lies between .05 and .5. †Integer values only (inclusive of limits).

	environment	η_H	environment $\times$ η_H
Resistance range, R^*			
$\mu = .001$	319	2,644	227
$\mu = .01$	129	3,610	112
$\mu = .05$	32	12,243	4.3
Infectivity range, I^*			
$\mu = .001$	302	2,993	188
$\mu = .01$	89	3,982	69
$\mu = .05$	11	4,747	34

Table B.2: ANCOVA results (F-ratios) for the primary model in chapter 3 with 2 loci. Resistance and infectivity ranges were $\log_{10}$-transformed prior to analysis. Results were highly significant with $p < .0001$. d.f. $= 3,436$ for $\mu = .001$; d.f. $= 3,540$ for $\mu = .01$; d.f. $= 2,742$ for $\mu = .05$

	environment	η_H	environment $\times$ η_H
Resistance range, R^*			
$\mu = .001$	297	2,885	194
$\mu = .01$	98	3,850	114
$\mu = .05$	39	11,294	11
Infectivity range, I^*			
$\mu = .001$	286	2,843	157
$\mu = .01$	62	3,508	58
$\mu = .05$	7.4	3,194	37

Table B.3: ANCOVA results (F-ratios) for the primary model in chapter 3 with 3 loci. Resistance and infectivity ranges were $\log_{10}$-transformed prior to analysis. Results were highly significant with $p < .0001$. d.f. $= 3, 234$ for $\mu = .001$; d.f. $= 3, 538$ for $\mu = .01$; d.f. $= 2, 688$ for $\mu = .05$

	environment	η_H	environment $\times$ η_H
Resistance range, R^*			
$\mu = .001$	130	1,324	70
$\mu = .01$	35	1,743	57
$\mu = .05$	19	4,416	$8.2^{\dagger}$
Infectivity range, I^*			
$\mu = .001$	100	1,299	38
$\mu = .01$	23	1,414	24
$\mu = .05$	$.76^{\dagger\dagger}$	909	$13.8^{\dagger}$

Table B.4: ANCOVA results (F-ratios) for the primary model in chapter 3 with 4 loci. Resistance and infectivity ranges were $\log_{10}$-transformed prior to analysis. Most results were highly significant with $p < .0001$ in almost all cases ($^{\dagger}p < .01$, $^{\dagger\dagger}p = .38$). d.f. $= 1, 356$ for $\mu = .001$; d.f. $= 1, 550$ for $\mu = .01$; d.f. $= 1, 152$ for $\mu = .05$

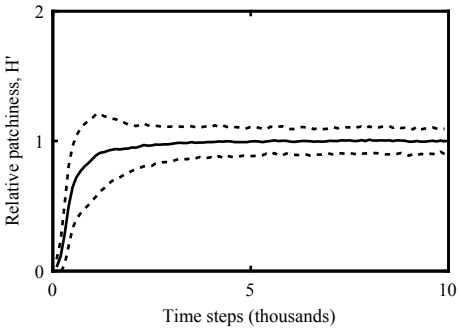

Figure B.1: Relative patchiness of spatially structured populations as a function of time (mean ± standard deviation) for $n = 2$. Patchiness is defined by the Shannon index (as described in the chapter 3), which measures the richness and evenness of patch sizes at a given time point; relative patchiness is equal to this value divided by the average over the final 5,000 generations, giving an indication as to how the spatial structure of the population tends to change during the course of a simulation. The first 1,000 generations are characterized by a rapid increase in patchiness, which then plateaus. The standard deviation does not increase over the final 5,000 generations, which implies that patchiness remains roughly constant within each simulation over this period.

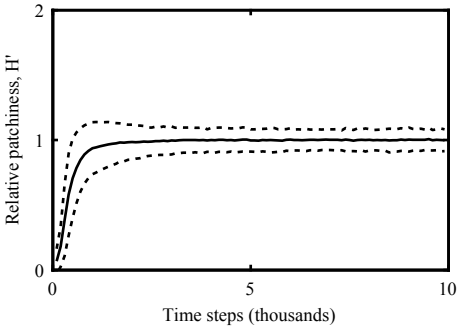

Figure B.2: Relative patchiness of spatially structured populations as a function of time (mean ± standard deviation) for $n = 4$. See figure B.1 for full description.

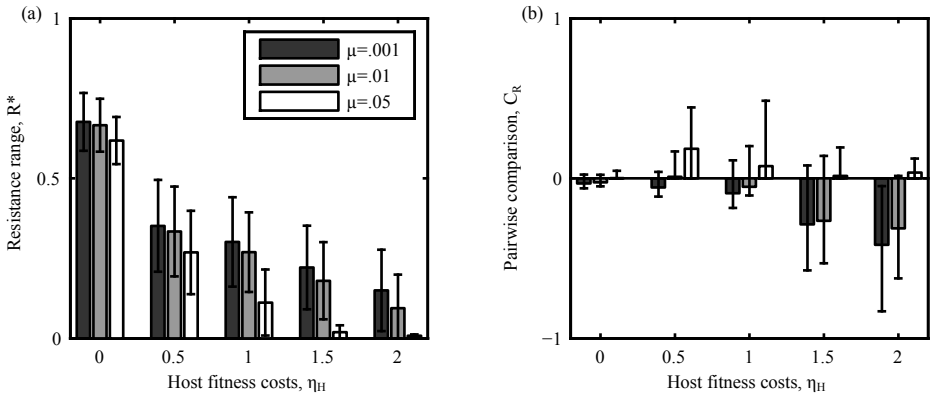

Figure B.3: The effects of fitness costs, spatial structure, and the natural mortality rate on the resistance range of the host population for $n = 2$. (a) The average resistance range (taken over both environments), R^*, tends to decrease with the host fitness cost parameter η_H and the natural mortality rate μ. (b) A pairwise comparison, C_R, of resistance range expansion in the two environments. If $C_R > 0$, then the mixed environment tends to produce broader resistance ranges than the spatially structured environment and vice versa if $C_R < 0$. Spatially structured populations generally exhibit broader resistance ranges than well-mixed populations as the associated fitness costs increase. However, high natural mortality rates can negate this effect, as resistance becomes less beneficial in both environments.

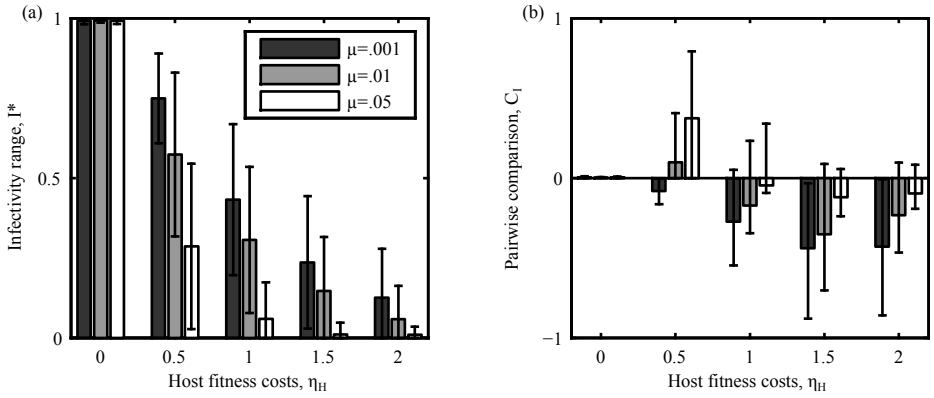

Figure B.4: The effects of fitness costs, spatial structure, and the natural mortality rate on the infectivity range of the parasite population for $n = 2$. (a) The average infectivity range (taken over both environments), I^*, tends to decrease with the host fitness cost parameter η_H, and the natural mortality rate μ. (b) A pairwise comparison, CI, of infectivity range expansion in the two environments. If $C_I > 0$, then the mixed environment tends to produce broader infectivity ranges than the spatially structured environment and vice versa if $C_I < 0$. Spatially structured populations generally exhibit broader infectivity ranges than well-mixed populations as the fitness costs for hosts increase. However, high natural mortality rates can negate this effect, as selection for infectivity decreases in both environments.

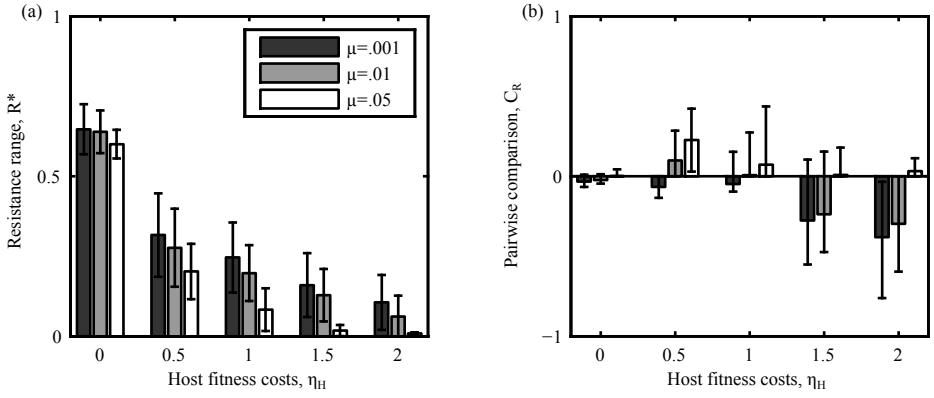

Figure B.5: The effects of fitness costs, spatial structure, and the natural mortality rate on the resistance range of the host population for $n = 4$. See figure B.3 for full description.

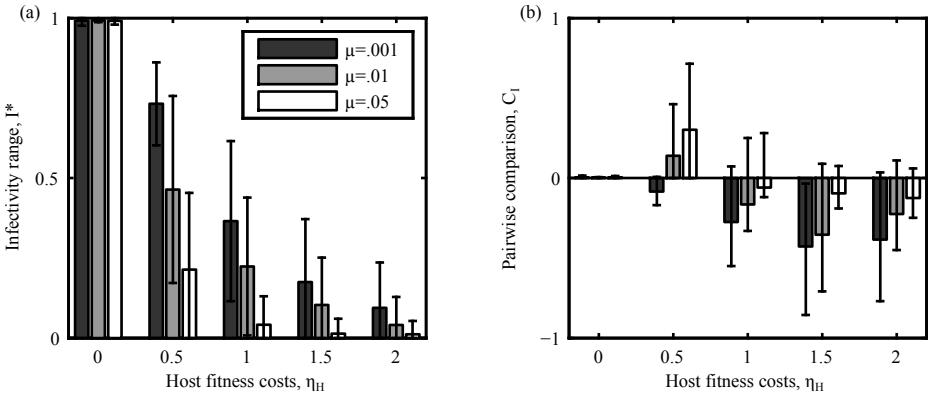

Figure B.6: The effects of fitness costs, spatial structure and the natural mortality rate on the infectivity range of the parasite population for $n = 4$. See figure B.4 for full description.

B.3 Probabilistic cellular automata (PCA)

B.3.1 Introduction

In chapter 3, we introduced a model of antagonistic coevolution in which hosts could become infected by free-living parasites in the environment. Our key finding was that spatial structure strongly mitigates the effects of fitness costs, with broader ranges more likely to emerge in structured than in well-mixed environments. We also found that as the natural mortality rate increased, spatially structured populations were no longer able to bear the high costs of resistance, leading to similar coevolutionary outcomes in the two environments. Here, we wish to test the generality of these results in a simplified model, using probabilistic cellular automata (PCA).

B.3.2 Model description

B.3.2.1 Genetic specificity

As described in chapter 3.

B.3.2.2 Simulation rules

The simulation rules for the PCA follow those of the full model in chapter 3, with the following exceptions. First, the parasite decay and diffusion processes are no longer necessary as free-living parasites are prohibited. Second, the processes for disease-induced mortality and infection are modified as outlined below:

Parasite-induced host mortality and new infections. Infected hosts are killed after a fixed number of time steps, τ. All immediately adjacent hosts are then challenged either by the same parasite strain or by a mutant. Mutations occur with probability ϵ_P at each locus, with the restriction that the genotype of a new strain can only differ from the previous generation by one bit. The probability that parasite genotype j is able to infect host genotype i is given by:

$$\rho_\alpha(i,j) = \lambda Q_{ij} c_P(j) \tag{B.1}$$

where λ is the probability of successful transmission, Q_{ij} is the strength of interaction between host and parasite and $c_P(j)$ captures fitness costs associated with broader infectivity ranges (Q_{ij} and $c_P(j)$ as defined in chapter 3). Susceptible hosts can be challenged by up to four strains in a single time step and the probability that at least one strain causes an infection is given by:

$$z_1(i) = 1 - \prod_k (1 - \rho_\alpha(i,k)) \tag{B.2}$$

If a uniform random number, $RAND_1 \in (0,1)$, satisfies $RAND_1 < z_1(i)$, then one of these strains is chosen at random to infect the host. The probability of strain j causing the infection is then equal to: $\rho_\alpha(i,j)/\sum_k \rho_\alpha(i,k)$.

	minimum value	maximum value
ϵ_H	.001	.005
ϵ_P	.004	.02
η_P	.1	.5
λ	.8	1
σ	.2	.5
$\tau^\dagger$	1	7
$T^\dagger$	1	7

Table B.5: Parameter distributions used to analyse the PCA model. †Integer values only (inclusive of limits).

We assume that coinfection does not occur.

These processes are the same for both spatially structured and well-mixed versions of the model. All other processes are identical to those described for the primary model in the main text. The initial grid consists of uninfected hosts at every site except one, which contains an infected host; both populations start without any resistance or infectivity alleles (i.e. $\sum_k h_k^i = \sum_k p_k^j = 0$).

B.3.2.3 Analysis

As described in chapter 3, with parameters drawn from the uniform random distributions in table B.5.

B.3.3 Results

We conducted 500 simulations for each combination of the parameters n, σ and μ; 46% of our simulations resulted in host-parasite coexistence in both environments for at least 10,000 time steps. Coexistence was more difficult to achieve in the PCA than in the primary model, as parasites were not able to persist in the environment.

However, certain regions of the parameter space were more suited to coexistence: for example, limiting η_P to values less than 0.15 increased the probability of coexistence to 69%.

B.3.3.1 Epidemiological dynamics

The epidemiological dynamics for the PCA were broadly similar to those observed for the primary model. In summary, well-mixed environments were prone to large amplitude epidemic cycles, whereas spatially structured populations typically demonstrated steady levels of infection (figure B.7). The patchiness of spatially structured populations was more variable than in the primary model, but still settled down to a stable distribution within 5,000 time steps (figures B.8-B.10).

B.3.3.2 Resistance and infectivity

Our measures of resistance and infectivity ranges followed similar patterns to those of the primary model: higher fitness costs for hosts (η_H) were associated with lower resistance ranges in both environments, but the decline was less severe in spatially structured populations (figures B.11-B.16, tables B.6-B.8).

The difference between the two environments was most notable for low values of μ. Our results were robust to changes in the measurement window: e.g. for $n = 3$ and $\mu = .001$ the median differences in resistance and infectivity ranges between time steps 5,001-7,500 and 7,500-10,000 were 3% and 2.4% respectively, indicating that both populations had reached a quasi-steady state prior to the beginning of the measurement window. We observed minor fluctuations about the mean values of R^* and I^* during this period (as shown in figure B.7), but the amplitudes of these oscillations were generally slightly lower than in the primary model.

	environment	η_H	environment $\times$ η_H
Resistance range, R^*			
$\mu = .001$	167	2,427	60
$\mu = .01$	106	5,213	$13^{\dagger}$
$\mu = .05$	235	6,471	94
Infectivity range, I^*			
$\mu = .001$	192	3,738	88
$\mu = .01$	71	4,636	22
$\mu = .05$	$.61^{\dagger\dagger}$	267	28

Table B.6: ANCOVA results (F-ratios) for the PCA model with 2 loci. Resistance and infectivity ranges were $\log_{10}$-transformed prior to analysis. Most results were highly significant with $p < .0001$ in almost all cases ($^{\dagger}p < .001$, $^{\dagger\dagger}p = .43$). d.f. $= 1,522$ for $\mu = .001$; d.f. $= 2,476$ for $\mu = .01$; d.f. $= 2,154$ for $\mu = .05$

	environment	η_H	environment $\times$ η_H
Resistance range, R^*			
$\mu = .001$	185	1,709	41
$\mu = .01$	90	4,607	$13^{\dagger}$
$\mu = .05$	296	5,685	129
Infectivity range, I^*			
$\mu = .001$	174	2,308	43
$\mu = .01$	73	2,076	$8.6^{\dagger}$
$\mu = .05$	$8.7^{\dagger}$	60	$15^{\dagger}$

Table B.7: ANCOVA results (F-ratios) for the PCA model with 3 loci. Resistance and infectivity ranges were $\log_{10}$-transformed prior to analysis. Results were highly significant with $p < .0001$ in almost all cases ($^{\dagger}p < .01$). d.f. $= 1,038$ for $\mu = .001$; d.f. $= 2,276$ for $\mu = .01$; d.f. $= 2,082$ for $\mu = .05$

	environment	η_H	environment $\times$ η_H
Resistance range, R^*			
$\mu = .001$	190	876	23
$\mu = .01$	89	3,417	15
$\mu = .05$	330	5,351	136
Infectivity range, I^*			
$\mu = .001$	156	1,110	$13^{\dagger}$
$\mu = .01$	87	893	$3.1^{\dagger\dagger}$
$\mu = .05$	18	29	$6.6^{\dagger}$

Table B.8: ANCOVA results (F-ratios) for the PCA with 4 loci. Resistance and infectivity ranges were $\log_{10}$-transformed prior to analysis. Most results were highly significant with $p < .0001$ in almost all cases ($^{\dagger}p < .01$, $^{\dagger\dagger}p = .08$. d.f. $= 726$ for $\mu = .001$; d.f. $= 2,074$ for $\mu = .01$; d.f. $= 2,054$ for $\mu = .05$

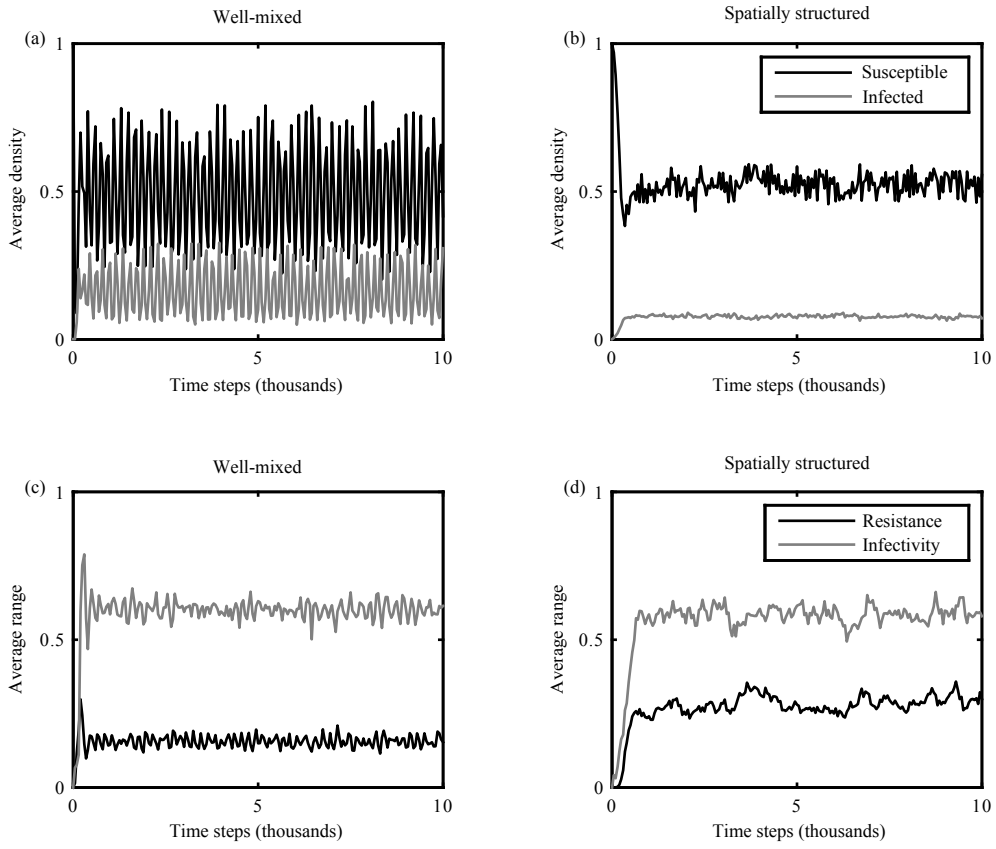

Figure B.7: Example coevolutionary dynamics from well-mixed (left column) and spatially structured (right column) populations in the PCA model. (a)-(b) Average density of susceptible (black) and infected (grey) members of the population. (c)-(d) Average resistance (black) and infectivity (grey) ranges. Well-mixed populations are prone to rapid fluctuating dynamics, whereas spatially structured populations generally produce more stable dynamics. Parameters: $\epsilon_H = 0.001$; $\epsilon_P = 0.01$; $\eta_H = 1$; $\eta_P = 0.1$; $\lambda = 0.95$; $\mu = 0.001$; $\sigma = 0.25$; $\tau = 3$; $N = 100$; $n = 3$; $T = 3$.

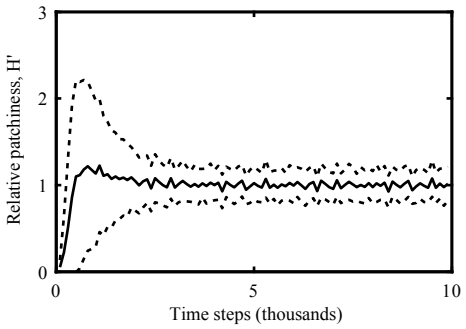

Figure B.8: Relative patchiness of spatially-structured populations as a function of time (mean ± standard deviation) for $n = 2$ in the PCA model. Patchiness is defined by the Shannon index (as described in chapter 3), which measures the richness and evenness of patch sizes at a given time point; relative patchiness is equal to this value divided by the average over the final 5,000 generations, giving an indication as to how the spatial structure of the population tends to change during the course of a simulation. The first 1,000 generations are characterized by a rapid increase in patchiness, which then plateaus. The standard deviation does not increase over the final 5,000 generations, which implies that patchiness remains roughly constant within each simulation over this period.

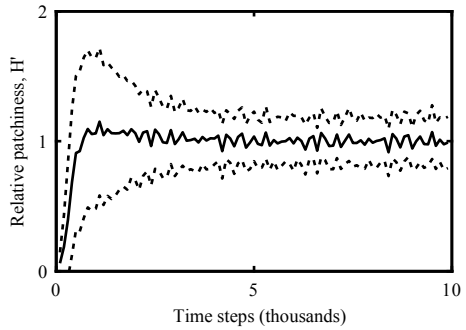

Figure B.9: Relative patchiness of spatially-structured populations as a function of time (mean ± standard deviation) for $n = 3$ in the PCA model. See figure B.8 for full description.

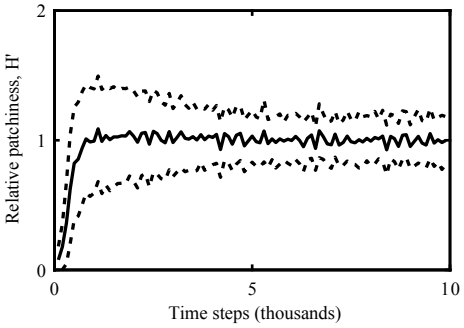

Figure B.10: Relative patchiness of spatially-structured populations as a function of time (mean ± standard deviation) for $n = 4$ in the PCA model. See figure B.8 for full description.

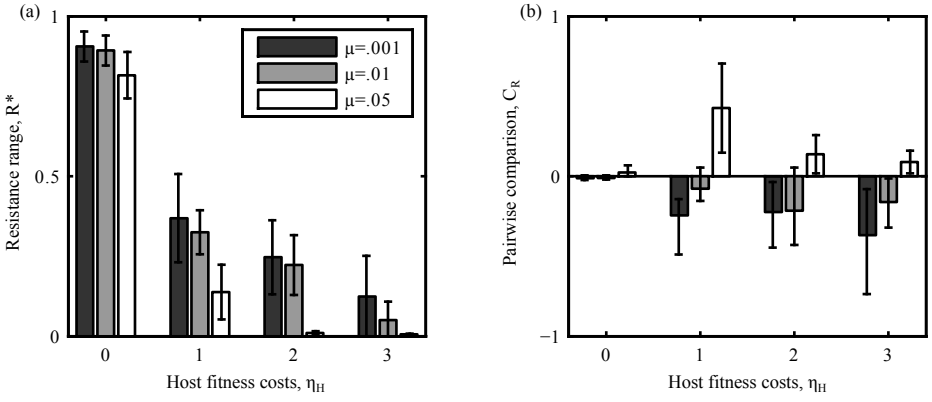

Figure B.11: The effects of fitness costs, spatial structure and the natural mortality rate on the resistance range of the host population for $n = 2$ in the PCA model. (a) The average resistance range (taken over both environments), R^*, tends to decrease with the host fitness cost parameter, η_H, and the natural mortality rate, μ. (b) A pairwise comparison, C_R, of resistance range expansion in the two environments. If $C_R > 0$, then the mixed environment tends to produce broader resistance ranges than the spatially structured environment and vice versa if $C_R < 0$. The results are qualitatively similar to the primary model.

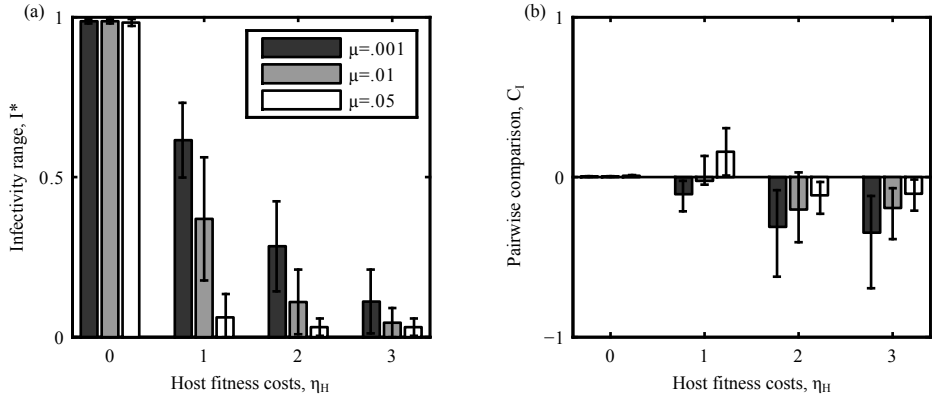

Figure B.12: The effects of fitness costs, spatial structure and the natural mortality rate on the infectivity range of the parasite population for $n = 2$ in the PCA model. (a) The average infectivity range (taken over both environments), I^*, tends to decrease with the host fitness cost parameter, η_H, and the natural mortality rate, μ. (b) A pairwise comparison, C_I, of infectivity range expansion in the two environments. If $C_I > 0$, then the mixed environment tends to produce broader infectivity ranges than the spatially-structured environment and vice versa if $C_I < 0$. The results are qualitatively similar to the primary model.

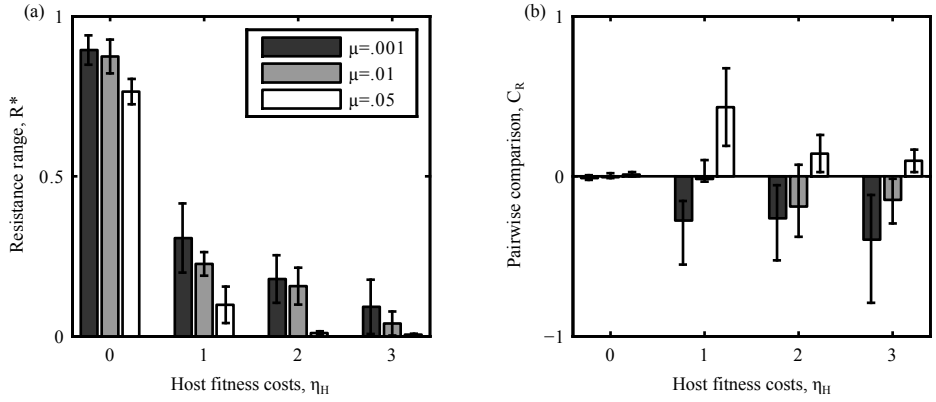

Figure B.13: The effects of fitness costs, spatial structure and the natural mortality rate on the resistance range of the host population for $n = 3$ in the PCA model. See figure B.11 for full description.

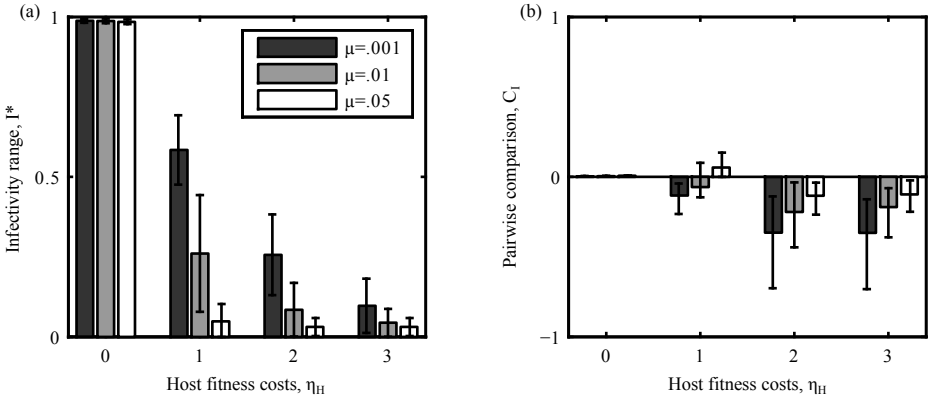

Figure B.14: The effects of fitness costs, spatial structure and the natural mortality rate on the infectivity range of the parasite population for $n = 3$ in the PCA model. See figure B.12 for full description.

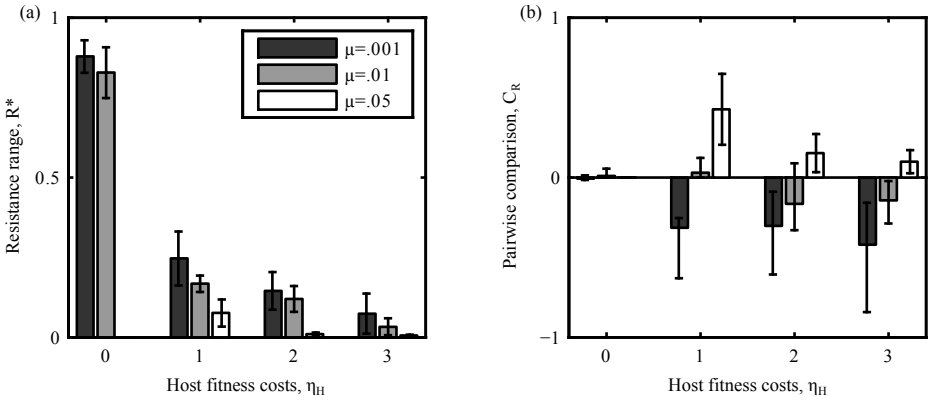

Figure B.15: The effects of fitness costs, spatial structure and the natural mortality rate on the resistance range of the host population for $n = 4$ in the PCA model. See figure B.11 for full description.

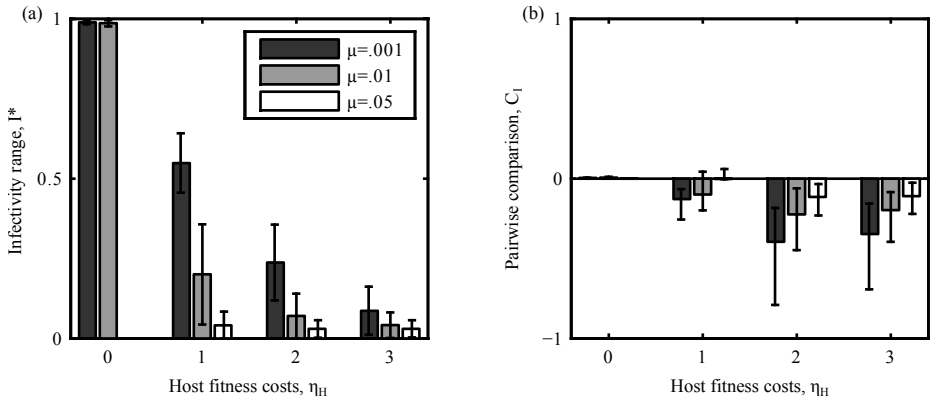

Figure B.16: The effects of fitness costs, spatial structure and the natural mortality rate on the infectivity range of the parasite population for $n = 4$ in the PCA model. See figure B.12 for full description.

Appendix C

Supplementary material for chapter 6

C.1 AIDS model

Figure 6.2 shows how the degree of assortative mixing in a population can influence epidemiology. The model used to generate this figure is based on that of Anderson et al. (1986), where a homosexual population is split into a number of distinct sexual activity classes. It is fully described by the following set of ordinary differential equations:

$$\left.\begin{aligned} \frac{dS_i}{dt} &= \nu N_i - \frac{\sum_j \beta c_i c_j \mathbf{p}_{ij} S_i I_j}{\sum_i N_i} - \mu S_i \\ \frac{dI_i}{dt} &= \frac{\sum_j \beta c_i c_j \mathbf{p}_{ij} S_i I_j}{\sum_i N_i} - (\gamma + \mu) I_i \\ \frac{dA_i}{dt} &= \gamma d I_i - (m + \mu) A_i \end{aligned}\right\} \tag{C.1}$$

where S_i and I_i represent susceptible and infected individuals, A_i corresponds to those individuals who have developed AIDS and N_i is the total number of individuals in sexual activity class i. The epidemiological parameters are given by β (the pairwise transmission rate), $1/\gamma$ (the average infectious period), d (the proportion of infections that lead to AIDS) and m (the AIDS-associated mortality rate). The parameter ν (μ) describes the rate at which individuals join (leave) the homosexual population.

The mixing matrix, $\mathbf{p}_{ij}$, describes the pattern of sexual contacts in the population. The degree of assortative mixing in the population, Q, can be measured using equa-

tion 6.6. Highly assortative mixing tends to lead to rapid growth during the early stages of an epidemic and can produce multiple peaks in disease incidence. Highly disassortative mixing is usually characterised by slower initial growth, but a higher peak in the incidence of AIDS.

The dynamics in figure 6.2 were generated using a population split into high (5 partners per year) and low (1 partner per year) sexual activity classes, with 80% of the initial population in the low activity class. Parameters used in figure 6.2 (all rates in years-1): $\beta = 0.375$; $\gamma = 0.1$; $d = 0.3$; $m = 1$; $\mu = \nu = 0.02$. The mixing matrix used in this example is based on that of Gupta et al. (1989):

$$\mathbf{p}_{ij} = \begin{pmatrix} \delta & 1-\delta \\ \frac{(1-\delta)c_1P_1}{c_2P_2} & 1-\frac{(1-\delta)c_1P_1}{c_2P_2} \end{pmatrix} \tag{C.2}$$

where P_i is the proportion of the population with partner acquisition rate c_i and δ scales the degree of assortative mixing in the population (Gupta et al., 1989). In figure 6.2 $\delta = 0.999$ for the highly assortative example and $\delta = 0$ for the highly disassortative example.

C.2 Additional figures

Figures C.1-C.2 show how the degree of polygamy affects the evolution of virulence for sub-linear ($s = 0.9$; figure C.1) and linear ($s = 1$; figure C.2) trade-offs between the transmission (β_p) and recovery (σ_p) rates. These figures correspond to figure 6.8, which features a superlinear ($s = 1.1$) trade-off.

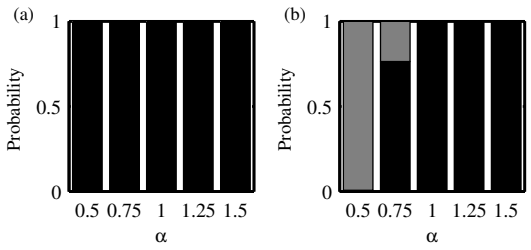

Figure C.1: Probabilities of different outcomes when two strains compete, with varying degrees of polygamy (α) for (a) polygynous and (b) polyandrous mating systems, and a sub-linear trade-off between transmission and recovery rates. Black bars correspond to strain 1 (less virulent) dominating, white bars to strain 2 (more virulent) dominating and grey bars to coexistence. A strain is defined to be dominating if it accounts for at least 95% of infections during the final 20% of a simulation. Parameters: $N_m = N_f = 500$; $r = 1.5$; $s = 0.9$; $\beta_1 \approx 0.07$; $\beta_1 \approx 0.56$; $\gamma = 0.5$; $\kappa = 10^{-4}$; $\mu = 0.01$; $\sigma_1 = 0.05$; $\sigma_2 = 0.4$.

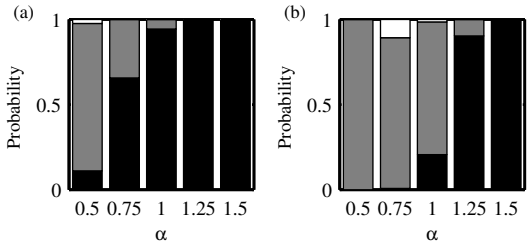

Figure C.2: Probabilities of different outcomes when two strains compete, with varying degrees of polygamy (α) for (a) polygynous and (b) polyandrous mating systems, and a linear trade-off between transmission and recovery rates. Description and parameters as in figure C.1, with $s = 1$.

Appendix D

Supplementary material for chapter 7

D.1 Full pair formation model

In chapter 7, I introduced a pair formation model (PFM) for monomorphic host and parasite populations (equation 7.1). The PFM can be generalised for polymorphic populations, where hosts differ in their strength of mate inspection (γ_i) and parasites differ in their transmission rates (β_k) and virulence ($1 - f_{\beta_k}$). In this expanded form, susceptible individuals are grouped into classes S_{γ_i} and those that are infected are grouped into classes $I_{\gamma_i\beta_k}$, where subscripts correspond to different host and parasite trait values. The number of unpaired members of the population is then given by $N_U = \sum_u \left(S_{\gamma_u} + \sum_v I_{\gamma_u\beta_v}\right)$ and the total population size is:

$$N = N_U + 2\left(\sum \left[S_{\gamma_i} S_{\gamma_j}\right] + \sum \left[S_{\gamma_i} I_{\gamma_j\beta_k}\right] + \sum \left[I_{\gamma_i\beta_k} I_{\gamma_j\beta_m}\right]\right) \tag{D.1}$$

where sums are taken over all possible combinations of subscripts. Setting $G[X]$ to be the sum of all pairs containing individuals in class X we obtain the following set of equations:

$$
\left.
\begin{aligned}
\frac{dS_{\gamma_i}}{dt} &= b_{\gamma_i} + (d+\mu)\left(2[S_{\gamma_i}S_{\gamma_i}] + G[S_{\gamma_i}]\right) \\
&\quad - \frac{pc^{\gamma_i}S_{\gamma_i}}{N_U}\left(\sum_j c^{\gamma_j}\left(S_{\gamma_j} + \sum_m f_{\beta_m}^{\gamma_i} I_{\gamma_j\beta_m}\right)\right) - \mu S_{\gamma_i} \\
\\
\frac{dI_{\gamma_i\beta_k}}{dt} &= (d+\mu)\left(2[I_{\gamma_i\beta_k}I_{\gamma_i\beta_k}] + G[I_{\gamma_i\beta_k}]\right) \\
&\quad - \frac{pc^{\gamma_i}I_{\gamma_i\beta_k}}{N_U}\left(\sum_j (cf_{\beta_k})^{\gamma_j}\left(S_{\gamma_j} + \sum_m f_{\beta_m}^{\gamma_i} I_{\gamma_j\beta_m}\right)\right) \\
&\quad - \mu I_{\gamma_i\beta_k} \\
\\
\frac{d[S_{\gamma_i}S_{\gamma_j}]}{dt} &= \frac{pc^{\gamma_i+\gamma_j}S_{\gamma_i}S_{\gamma_j}}{(1+\delta_{ij})N_U} - (d+2\mu)[S_{\gamma_i}S_{\gamma_j}] \\
\\
\frac{d[S_{\gamma_i}I_{\gamma_j\beta_k}]}{dt} &= \frac{pc^{\gamma_i+\gamma_j}f_{\beta_k}^{\gamma_i}S_{\gamma_i}I_{\gamma_j\beta_k}}{N_U} - (d+2\mu+\beta_k)[S_{\gamma_i}I_{\gamma_j\beta_k}] \\
\\
\frac{d[I_{\gamma_i\beta_k}I_{\gamma_j\beta_m}]}{dt} &= \frac{pc^{\gamma_i+\gamma_j}f_{\beta_m}^{\gamma_i}f_{\beta_k}^{\gamma_j}I_{\gamma_i\beta_k}I_{\gamma_j\beta_m}}{(1+\delta_{ij}\delta_{km})N_U} \\
&\quad - (d+2\mu)[I_{\gamma_i\beta_k}I_{\gamma_j\beta_m}] + \delta_{km}\beta_k\left([S_{\gamma_i}I_{\gamma_j\beta_k}] + [S_{\gamma_j}I_{\gamma_i\beta_k}]\right)
\end{aligned}
\right\} \quad \text{(D.2)}
$$

where $\delta_{ij} = 1$ if $i = j$ and is 0 otherwise, and b_{γ_i} is the birth rate for new offspring with mate inspection parameter γ_i given by:

$$
\begin{aligned}
b_{\gamma_i} = r\left(1 - \frac{N}{K}\right)\Big(&\sum (1+\delta_{ij})[S_{\gamma_i}S_{\gamma_j}] + \sum f_{\beta_k}(1+\delta_{ij})[S_{\gamma_i}I_{\gamma_j\beta_k}] \\
&+ \sum f_{\beta_k}f_{\beta_m}(1+\delta_{ij})[I_{\gamma_i\beta_k}I_{\gamma_j\beta_m}]\Big)
\end{aligned}
\quad \text{(D.3)}
$$

I use this generalised form of the PFM to explore the (co)evolution of virulence and mate inspection using an adaptive dynamics approach (Hofbauer and Sigmund, 1990;

Dieckmann and Law, 1996). Adaptive dynamics assumes that resident host and parasite populations reach equilibrium relatively quickly and that mutants, which exhibit minor differences in phenotype from residents, then attempt to invade. This amounts to a separation of ecological and evolutionary timescales so that evolutionarily and coevolutionarily stable strategies (ESS/CoESS) can be determined. I use numerical methods as the complexity of the PFM makes it intractable to algebraic analysis (figures 7.3, 7.5).

D.2 Derivation of R_0

The basic reproductive ratio, R_0, is the expected number of secondary infections caused by an a single infectious individual in an otherwise susceptible population (Anderson and May, 1991). This quantity can be derived for the current model by taking the product of (i) the expected number of sexual partnerships that an infectious individual will make in their lifetime, and (ii) the probability that transmission will occur before the partnership is terminated. I assume that the host population is monomorphic (constant γ) and that there are always a large number of unpaired individuals, so that the probability of pairing with an individual twice is negligible. In an otherwise susceptible population, the probability that an infectious individual will form at least one sexual partnership is given by:

$$q_1(f,\gamma) = \frac{pc^{2\gamma}f^{\gamma}}{pc^{2\gamma}f^{\gamma}+\mu} \tag{D.4}$$

Further sexual contacts cannot be formed until the partnership ends and the infectious individual becomes single once again, which happens with probability:

$$q_2 = \frac{d+\mu}{d+2\mu} \tag{D.5}$$

The probability of forming another sexual partnership is given by $q_1(f, \gamma)$ again, and the process can repeat indefinitely. The expected number of pairs that the infectious individual forms in their lifetime can be calculated by considering the process of pair dissolution and subsequent pair formation as a Bernoulli trial. The probability of dying before the current pair dissolves and a new one is formed is equal to $1 - q_1(f, \gamma)q_2$, so the expected number of pairs formed by the infectious individual is given by the product of forming an initial pair and the mean of a geometric distribution with probability $1 - q_1(f, \gamma)q_2$:

$$E(f, \gamma) = \frac{q_1(f, \gamma)}{1 - q_1(f, \gamma)q_2} \tag{D.6}$$

If an infectious individual is in a sexual partnership, the probability that transmission occurs before the pair ends is given by:

$$q_3(\beta) = \frac{\beta}{\beta + d + 2\mu} \tag{D.7}$$

R_0 is then equal to the product of the quantities in equations D.6 and D.7:

$$\begin{aligned} R_0 &= E(f, \gamma)q_3(\beta) \\ &= \frac{pc^{2\gamma}f^{\gamma}\beta(d + 2\mu)}{\mu(pc^{2\gamma}f^{\gamma} + d + 2\mu)(\beta + d + 2\mu)} \end{aligned} \tag{D.8}$$

Publications

doi:10.1111/evo.12479

EFFECTS OF EPISTASIS ON INFECTIVITY RANGE DURING HOST-PARASITE COEVOLUTION

Ben Ashby,[1,2] Sunetra Gupta,[1] and Angus Buckling[3]

[1]*Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, United Kingdom*

[2]*E-mail: ben.ashby@zoo.ox.ac.uk*

[3]*Biosciences, University of Exeter, Penryn Campus, TR10 9FE, United Kingdom*

Received February 13, 2014
Accepted June 11, 2014

Understanding how parasites adapt to changes in host resistance is crucial to evolutionary epidemiology. Experimental studies have demonstrated that parasites are more capable of adapting to gradual, rather than sudden changes in host phenotype, as the latter may require multiple mutations that are unlikely to arise simultaneously. A key, but as yet unexplored factor is precisely how interactions between mutations (epistasis) affect parasite evolution. Here, we investigate this phenomenon in the context of infectivity range, where parasites may experience selection to infect broader sets of genotypes. When epistasis is strongly positive, we find that parasites are unlikely to evolve broader infectivity ranges if hosts exhibit sudden, rather than gradual changes in phenotype, in close agreement with empirical observations. This is due to a low probability of fixing multiple mutations that individually confer no immediate advantage. When epistasis is weaker, parasites are more likely to evolve broader infectivity ranges if hosts make sudden changes in phenotype, which can be explained by a balance between mutation supply and selection. Thus, we demonstrate that both the rate of phenotypic change in hosts and the form of epistasis between mutations in parasites are crucial in shaping the evolution of infectivity range.

KEY WORDS: **Epistasis, generalism, host-parasite coevolution, infectivity range, multiple mutations, resistance.**

Antagonistic coevolution between hosts and parasites can lead to directional selection for more effective defense and counter-defense mechanisms (Thrall and Burdon 2003; Labrie et al. 2010; Schulte et al. 2010; Brown and Tellier 2011). In many cases, these dynamics (often referred to as "coevolutionary arms races") are characterized by reciprocal expansions in the range of genotypes that the host can resist and the parasite can infect, which means that populations tend to fare better than their ancestors when confronted with contemporary antagonists (Buckling and Rainey 2002; Mizoguchi et al. 2003; Thrall and Burdon 2003; Brown and Tellier 2011; Scanlan et al. 2011). Understanding precisely why some parasites develop broader infectivity ranges than others has important implications for our ability to predict how parasites will evolve in response to shifting patterns of host resistance or other environmental changes, with particular relevance for the use of biocontrol in industry and medicine (Tait et al. 2002; Levin and Bull 2004). While variation in infectivity range is typically explained by selection (e.g., fitness costs; Fenton and Brockhurst 2007; Ashby et al. 2014) or fundamental genetic constraints (e.g., parasites may be forced to specialize on one group of hosts or another; Dybdahl and Lively 1998; Decaestecker et al. 2007; Koskella and Lively 2007), a lack of broad infectivity ranges may also result from the need to fix multiple, rather than single, mutations (Benmayor et al. 2009; Paterson et al. 2010; Hall et al. 2011; Meyer et al. 2012). Here, we investigate how key parameters (epistasis and the rate of phenotypic change in the host) affect the fixation of multiple mutations, and hence infectivity range, during coevolution.

Parasites frequently require multiple amino acid substitutions to infect a novel host, and the likelihood of several beneficial mutations occurring simultaneously or in quick succession is usually slim (Benmayor et al. 2009; Hall et al. 2011; Scanlan et al. 2011; Gurarani et al. 2012; Meyer et al. 2012; Russell et al. 2012). In some cases, however, subsets of mutations may confer an immediate fitness advantage on contemporaneous hosts, increasing the probability that a complete set will eventually

become fixed (Meyer et al. 2012). Empirical observations using bacteria and viruses suggest that these conditions are most likely to be realized when parasites are exposed to genetically diverse host populations, such that they experience gradual, rather than sudden changes in phenotype during coevolution, as individual mutations may increase performance on subsets of the host population (Hall et al. 2011; Meyer et al. 2012). Crucially, a wide range of genetic and ecological processes, such as recombination and gene flow, could alter the rate of phenotypic change in the host population (Sasaki 2000; Gandon 2002; Gandon and Nuismer 2009), and hence the likelihood of broad infectivity ranges evolving.

Empirical studies that demonstrate the importance of coevolution for the emergence of broad infectivity ranges have used host-parasite interactions that are governed by strong positive epistasis between infectivity mutations (Paterson et al. 2010; Hall et al. 2011; Meyer et al. 2012). This means that parasites with an incomplete set of mutations will fare no better (or even worse) on the novel host than parasites with none. However, both quantitatively and qualitatively different forms of epistasis governing infectivity have been identified, including weak positive, negative, and no epistasis (Lenski 1984; Wilfert and Schmid-Hempel 2008), and it is currently unclear how this will impact on infectivity range during coevolution. Here, we demonstrate theoretically that different forms of epistasis have contrasting effects on the ability of parasites to expand their infectivity ranges when hosts exhibit gradual or sudden changes in phenotype during coevolution. Our results are in good agreement with empirical observations when epistasis is strongly positive (gradual changes in host phenotype promote broader infectivity ranges), but notably, we find that the opposite outcome is expected for weaker forms of epistasis (sudden changes in host phenotype promote broader infectivity ranges).

Methods

MODEL DESCRIPTION

We compare two types of genetic specificity that govern host-parasite interactions, both of which allow the evolution of parasites with broader infectivity ranges. The first is similar to the multilocus gene-for-gene framework proposed by Sasaki (2000), with interactions occurring at n biallelic loci in both host and parasite. Increasing the number of infectivity alleles improves infectivity to a wider range of host genotypes, and increasing the number of resistance alleles improves resistance to a wider range of parasite genotypes. We refer to this as a "symmetric" (SYM) interaction, because there is a one-to-one correspondence between resistance and infectivity alleles. We compare this scenario to an "asymmetric" (ASYM) form of genetic specificity, where interactions occur between a single locus in the host and n loci in the parasite (one-to-many). In this case, there are only two possible host genotypes (susceptible and resistant), and increasing the number of infectivity alleles can improve performance on the resistant host. Genotypes are represented by binary strings (hosts: $h_1^i \ldots h_n^i$ (SYM) or h_1^i (ASYM); parasites: $p_1^j \ldots p_n^j$; superscripts identify each genotype), where each locus corresponds to the presence (1) or absence (0) of a resistance (host) or infectivity (parasite) allele. Infectivity alleles interact with each other and with resistance alleles to modulate the overall strength of infectivity, Q, on a given host, such that:

$$Q_{ij} = \begin{cases} 1 - \left(1 - d_{ij}\right)^{\psi} & \text{if } d_{ij} < 1 \\ 1 & \text{otherwise} \end{cases} \quad (1)$$

where d_{ij} is the proportion of infectivity alleles that match or exceed either (i) the resistance allele at each corresponding locus (SYM: $d_{ij} = 1 - \frac{1}{n}\sum_{k=1}^{n} h_k^i(1 - p_k^j)$), or (ii) the sole resistance allele in the host (ASYM: $d_{ij} = 1 - \frac{h_1^i}{n}\sum_{k=1}^{n}(1 - p_k^j)$) (Fig. 1A–B). The parameter ψ modulates the type and strength of epistasis between infectivity alleles, such that $0 \leq \psi < 1$, $\psi > 1$ and $\psi = 1$ give positive, negative and no epistasis, respectively (Fig. 1C). Values of ψ further away from 1 give stronger forms of epistasis; in the special case of $\psi = 0$ infection is only possible if $p_k^j \geq h_k^i$ at all loci.

We base the epidemiological dynamics of our model on the SI framework, where hosts of genotype i are classed as either susceptible (S_i) or infected by parasite genotype j $\left(I_{ij}\right)$. Hosts are haploid and reproduce asexually with a maximum per-capita birth rate of $\bar{r}$, and experience a density-dependent per-capita mortality rate of at least $\bar{\mu}N$, where $\bar{\mu}$ measures the strength of competition for resources and N is the total population size. We set $\bar{\mu} = \bar{r}/K$, so that the host population tends towards a carrying capacity of K in the absence of disease. Initial populations are composed of K susceptible and $K/20$ infected hosts, with no resistance or infectivity alleles present. The host population mixes randomly and exhibits either frequency- (FD) or density-dependent (DD) contact patterns, so that a susceptible host of genotype i will be infected with parasite j at a rate of $\lambda_{ij} = \beta_j Q_{ij} \sum_k I_{kj}/N$ (FD) or $\lambda_{ij} = \beta_j Q_{ij} \sum_k I_{kj}$ (DD) per unit time, where β_j is the transmission coefficient of the parasite (base transmission coefficient: $\bar{\beta}$). Infected hosts are unable to recover and suffer an increased mortality rate, given by the parameter α_j (base disease-associated mortality rate: $\bar{\alpha}$) coinfection does not occur. New generations are subject to mutation rates of ε_H and ε_P at each locus for hosts and parasites, respectively, with the restriction that multiple mutations cannot arise simultaneously (i.e., the genotypes of parent and progeny never differ at more than one locus).

Broader resistance and infectivity ranges are often associated with a fitness cost (Chao et al. 1977; Webster and Woolhouse 1999; Bohannan et al. 2002; Poullain et al. 2008), which we incorporate into either the host per-capita birth rate (r_i) or the

SYM

A	P_{000}	P_{001}	P_{010}	P_{100}	P_{011}	P_{101}	P_{110}	P_{111}
H_{000}	1	1	1	1	1	1	1	1
H_{001}	2/3	1	2/3	2/3	1	1	2/3	1
H_{010}	2/3	2/3	1	2/3	1	2/3	1	1
H_{100}	2/3	2/3	2/3	1	2/3	1	1	1
H_{011}	1/3	2/3	2/3	1/3	1	2/3	2/3	1
H_{101}	1/3	2/3	1/3	2/3	2/3	1	2/3	1
H_{110}	1/3	1/3	2/3	2/3	2/3	2/3	1	1
H_{111}	0	1/3	1/3	1/3	2/3	2/3	2/3	1

ASYM

B	P_{000}	P_{001}	P_{010}	P_{100}	P_{011}	P_{101}	P_{110}	P_{111}
H_0	1	1	1	1	1	1	1	1
H_1	0	1/3	1/3	1/3	2/3	2/3	2/3	1

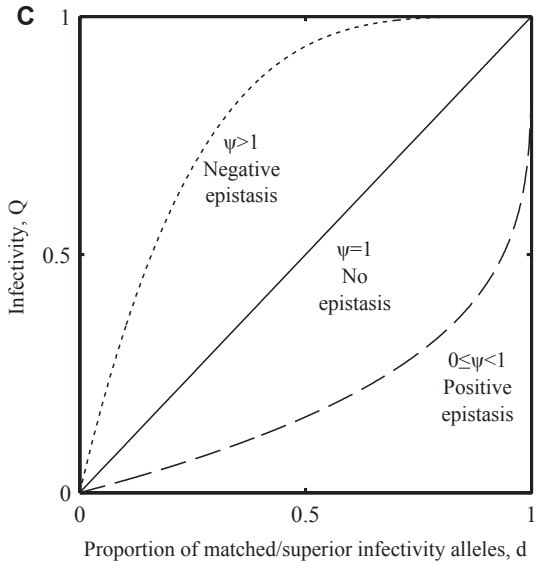

Figure 1. **Genetic interactions between hosts and parasites. The tables show the proportion of parasite loci, *d*, that match or are superior to (a) corresponding loci (symmetric [SYM] scenario) or (b) the sole locus (asymmetric [ASYM] scenario) in the host. Interactions between hosts (H) and parasites (P) are shown for $n = 3$, where subscripts correspond to the presence (1) or absence (0) of a resistance or infectivity allele at a given locus. (c) Infectivity (Q; eq. 1) as a function of *d* for different values of the epistasis parameter, ψ, which modifies the type and strength of epistasis between infectivity alleles: $0 \leq \psi < 1$, $\psi > 1$ and $\psi = 1$ give positive, negative, and no epistasis, respectively.**

coefficient of density-dependent mortality (μ_i), and either the disease-associated mortality rate (α_j) or transmission coefficient (β_j) for parasites. We limit simulations to one type of fitness cost per population, giving a total of four combinations. For example, if hosts experience a fitness cost in the form of a reduced birth rate, then the coefficient of density-dependent mortality remains constant for all genotypes ($\mu_i = \bar{\mu}$) and if parasites with broader infectivity ranges have a lower transmission coefficient, then the disease-associated mortality rate does not vary $(\alpha_j = \bar{\alpha})$. When fitness costs do affect a particular life-history trait, they do so based on the following equations:

$$r_i = \bar{r}\left(1 + \left(\frac{1}{c_H} - 1\right) q_i^{\varphi_H}\right) \tag{2a}$$

$$\mu_i = \bar{\mu}\left(1 + (c_H - 1)\, q_i^{\varphi_H}\right) \tag{2b}$$

$$\alpha_j = \bar{\alpha}\left(1 + (c_P - 1)\, q_j^{\varphi_P}\right) \tag{2c}$$

$$\beta_j = \bar{\beta}\left(1 + \left(\frac{1}{c_P} - 1\right) q_j^{\varphi_P}\right) \tag{2d}$$

where for hosts (parasites), q_i (q_j) gives the proportion of loci that contain a resistance (infectivity) allele, $c_H \geq 1$ ($c_P \geq 1$) is the maximum strength of the fitness cost and φ_H (φ_P) controls whether costs are accelerating ($\varphi_H, \varphi_P > 1$), decelerating ($0 < \varphi_H, \varphi_P < 1$) or linear ($\varphi_H, \varphi_P = 1$). Note that for $q_i = 1$, $r_i = \bar{r}/c_H$ or $\mu_i = \bar{\mu} c_H$, which means that the birth rate or coefficient of density-dependent mortality is c_H times lower/higher for individuals with a full complement of resistance alleles than the base values (similarly for $q_j = 1$: $\beta_j = \bar{\beta}/c_P$ and $\alpha_j = \bar{\alpha} c_P$). Although resistance and infectivity alleles can behave epistatically for both specificity (Q) and fitness costs (see e.g., Fenton and Brockhurst 2007), we shall only refer to epistasis in the context of the parasite's ability to infect a given host (i.e., in terms of the parameter ψ) to avoid confusion. Variations in ψ will be referred to as positive, negative or no epistasis, whereas variations in φ_H and φ_P will be referred to accelerating, decelerating or linear fitness costs.

The dynamics of our model (excluding mutations) are captured by the following set of coupled ordinary differential equations:

$$\frac{dS_i}{dt} = S_i\left(r_i - \mu_i N - \sum_j \lambda_{ij}\right) \tag{3a}$$

$$\frac{dI_{ij}}{dt} = \lambda_{ij} S_i - \left(\mu_i N + \alpha_j\right) I_{ij} \tag{3b}$$

We translate this deterministic framework into a stochastic model by using the τ-leap method proposed by Gillespie (2001), which uses a fixed step size, τ, and assumes the number of events occurring within a time step is Poisson distributed. The optimal genotype will always emerge in a deterministic framework with no extinction threshold, but we should expect parasites to struggle

to accumulate infectivity alleles when demographic stochasticity is included, especially if resistance spreads rapidly. Thus, by comparing the deterministic and stochastic models, we are able to establish if broader infectivity ranges do not evolve due to selection (i.e., broader infectivity ranges are not beneficial), or if mutations are struggling to reach fixation due to stochasticity.

ANALYSIS

We analyze the deterministic and stochastic versions of our model to evaluate how the previously described forms of genetic specificity (SYM and ASYM) and different types of epistasis (ψ) influence the evolution of broader infectivity ranges. In other words, we establish how these genetic factors affect the ability of parasites to accumulate infectivity alleles. At each time point, we measure the average proportion of parasite loci that contain an infectivity allele and define "peak infectivity range," E, to be the maximum of this value over the course of a simulation (20,000 time units). Thus, if $x_k(t)$ is the proportion of parasites that have a total of k infectivity alleles at time t, then:

$$E = \max_t \left\{ \frac{1}{n} \sum_{k=1}^{n} k x_k(t) \right\} \tag{4}$$

We measure the maximum value over the duration of each simulation as GFG frameworks can produce fluctuations in range (e.g., Sasaki 2000), but the focus of the present study is whether genetic factors affect the initial emergence of broader infectivity ranges and not whether they are evolutionarily stable. We wish to determine the general behavior of our model, but since we are not modeling a particular host-parasite system, the parameter space is somewhat arbitrary. To overcome this issue, we fix $\tau = 0.1$ and vary ψ, φ_H and φ_P incrementally to cover all qualitatively different forms of epistasis (positive, negative, or none) and fitness costs (accelerating, decelerating, or linear), and use a Latin hypercube sample (LHS) to draw the remaining parameters from the distributions in Table 1, the majority of which are varied over at least an order of magnitude, covering biologically plausible areas of parameter space (e.g., population sizes of 10^5–10^9 are appropriate for microbial communities). Note that the mutation rates and the coefficient of density-dependent mortality are not used in the construction of the LHS, but are instead fixed by the base per-capita birth rate and/or carrying capacity, which are part of the LHS design. The parameters c_H and c_P vary over relatively narrow ranges compared to the other parameters as previous studies have demonstrated that high fitness costs greatly limit infectivity range in well-mixed populations (Sasaki 2000; Ashby et al. 2014). The LHS contains 1000 parameter combinations, each of which is tested in both SYM and ASYM scenarios with all four possible combinations of fitness costs (eq. 2) for different values of ψ, φ_H, and φ_P. This method allows us to determine if one form of genetic specificity (SYM or ASYM) consistently allows broader infectivity ranges to evolve than the other, and if this relationship holds for all forms of epistasis (ψ) and fitness costs (φ_H and φ_P). We discard simulations where the parasite dies out in either the SYM or ASYM scenarios for a given set of parameters.

Table 1. Parameter distributions used for the Latin hypercube sample.

	Description	Range
$\bar{\alpha}$	Base disease-associated mortality rate	(0.01–0.1)
$\bar{\beta}$	Base transmission coefficient	FD: (0.1–1), DD: (10^{-9}–10^{-5})*
ε_H	Host mutation rate	$1/K$**
ε_P	Parasite mutation rate	$1/K$**
$\bar{\mu}$	Base coefficient of density-dependent mortality	$\bar{r}/K$**
c_H	Maximum strength of host fitness costs	(1.05–1.2)
c_P	Maximum strength of parasite fitness costs	(1.05–1.5)
K	Carrying capacity	(10^5–10^9)
$\bar{r}$	Base per-capita birth rate	(0.01–0.1)

*FD, frequency-dependent transmission; DD, density-dependent transmission.

**Values are fixed by the base per-capita birth rate and/or carrying capacity.

Results

For the sake of brevity, here we only present results for parasite populations with three loci ($n = 3$), contact patterns based on frequency-dependence (FD) and decelerating fitness costs ($\varphi_H = \varphi_P = 0.5$) affecting only the per-capita birth rate (r_i) and transmission coefficient (β_j). However, the results are qualitatively similar for other numbers of loci (Supplementary Fig. S1), density-dependent contact patterns (Supplementary Fig. S2), linear and accelerating fitness costs (Supplementary Fig. S3) and different combinations of cost functions (Supplementary Fig. S4).

ALGEBRAIC ANALYSIS

Using equation (3), we can derive the basic reproductive ratio, R_0^j, for a parasite in a naive, fully susceptible host population when transmission is frequency-dependent:

$$R_0^j = \frac{\beta_j}{\bar{\mu} N + \bar{\alpha}} \tag{5}$$

In the absence of resistant hosts, this quantity is maximized when the parasite has no infectivity alleles. We can also derive the effective basic reproductive ratio, R_{eff}^j, which is the average

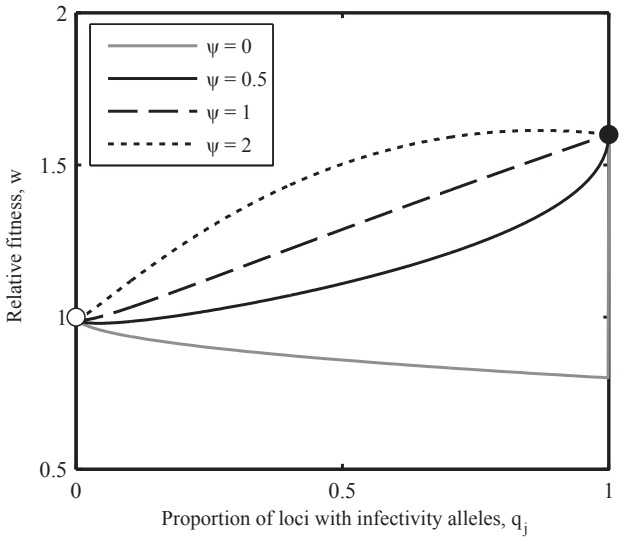

Figure 2. Relative fitness (w; eq. 7) of parasites in a fixed host population for different values of the epistasis parameter, ψ. The white circle indicates the fitness of the wild-type parasite (no infectivity alleles) and the black circle indicates the fitness of parasites with a full complement of infectivity alleles. When epistasis is strongly positive (gray), parasites with a complete set of infectivity alleles may have the highest fitness, but intermediate genotypes could perform worse than the wild type, so parasites may struggle to accumulate infectivity alleles (note the step-change when $q_j = 1$ for $\psi = 0$). When epistasis is weaker (solid black) or nonexistent (dashed), parasites with an incomplete set of infectivity alleles are likely to experience an immediate increase in fitness, which may allow infectivity ranges to expand. When epistasis is negative (dotted), an incomplete set of mutations may be optimal due to the presence of fitness costs, which outweigh the benefits of broader infectivity ranges. Parameters: $\varphi_P = 0.5$; $c_P = 1.25$.

number of secondary infections produced for any composition of hosts:

$$R^j_{eff} = \frac{\beta_j H_j}{\bar{\mu} N + \bar{\alpha}} \tag{6}$$

where $H_j = \sum_i Q_{ij} S_i / N$ is the proportion of the host population that the parasite is able to infect. For a given composition of hosts, parasite j should initially perform better than the wild type (no infectivity alleles, H_0) provided $H_0 = 0$ and $H_j > 0$, or, for $H_0 > 0$:

$$w_j = \frac{R^j_{eff}}{R^0_{eff}} = \frac{\left(1 + \left(\frac{1}{c_P} - 1\right) q_j^{\varphi_P}\right) H_j}{H_0} > 1 \tag{7}$$

Figure 2 shows several examples of how this fitness function varies with the number of infectivity alleles in the parasite for different values of the epistasis parameter, ψ. While Figure 2 represents an idealized scenario with the host population held constant (50% wild type, 50% maximal resistance), it does reveal some interesting patterns. In particular, parasites with an intermediate number of infectivity alleles may perform worse than the wild type if epistasis is positive. This suggests that broad infectivity ranges are unlikely to emerge if hosts make sudden jumps in phenotype. Clearly, the composition of the host population will alter selection among parasites and vice versa, which prevents further algebraic analysis of this system. In addition, real populations are subject to stochasticity, which can have a considerable influence on the accumulation of rare mutations. However, based on the above analysis, we can make two predictions to test numerically: (i) selection for parasites with broad infectivity ranges should peak for low values of ψ, as complete or near complete sets of mutations are required to overcome host resistance; (ii) parasites in the ASYM scenario will struggle to accumulate mutations when demographic stochasticity is included if epistasis is strongly positive, as intermediate genotypes may perform worse than the wild type.

NUMERICAL ANALYSIS

Strong positive epistasis selects for parasites with broad infectivity ranges

Numerical analysis of the deterministic model revealed that infectivity range peaks when epistasis is positive ($\psi < 1$; Fig. 3), but declines rapidly if epistasis is negative ($\psi > 1$). The lack of an extinction threshold in the deterministic model means that the optimal genotype is always able to emerge, so there are no qualitative differences between the SYM and ASYM scenarios.

Stochasticity constrains infectivity range when hosts exhibit sudden changes in phenotype

The pattern of parasite evolution in the stochastic SYM scenario was broadly similar to that described for the deterministic model: infectivity ranges peaked for strong positive epistasis ($\psi << 1$) and decreased with greater ψ (Fig. 4A). Yet, unlike the deterministic version of the SYM scenario, the stochastic version exhibited very little, if any, selection for broader infectivity ranges for negative epistasis ($\psi > 1$). This is due to reduced selection for resistance in the host (Supplementary Fig. S5C). When epistasis between infectivity mutations is negative, resistance is largely ineffective as hosts require multiple mutations to achieve a significant reduction in susceptibility. Hence, hosts do not tend to evolve broader resistance ranges when demographic stochasticity is included and consequently parasites do not need to evolve broader infectivity ranges under these conditions.

In contrast, the stochastic version of the ASYM scenario produced markedly different outcomes to both the deterministic model and the SYM scenario (Fig. 4B). Specifically, broad infectivity ranges were extremely rare for strong positive epistasis and instead peaked for weaker interactions (intermediate

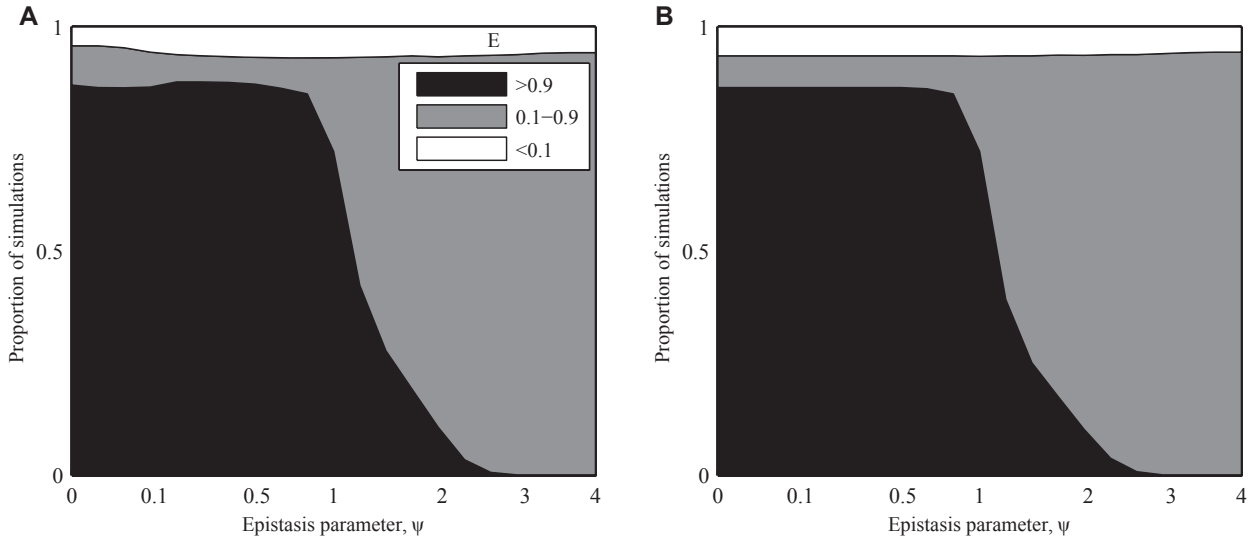

Figure 3. Proportion of deterministic simulations where peak infectivity range (E; eq. 4) was greater than 0.9 (black), less than 0.1 (white), or between these values (gray), for (A) gradual (SYM) and (B) sudden (ASYM) changes in host phenotype. The parameter ψ controls the type and strength of epistasis between infectivity alleles, ranging from strong positive ($\psi << 1$), through weak positive ($\psi < 1$), none ($\psi = 1$), and finally, negative ($\psi > 1$) epistasis. The fittest genotype always emerges in a deterministic framework with no extinction threshold, so the SYM and ASYM scenarios produce almost identical outputs.

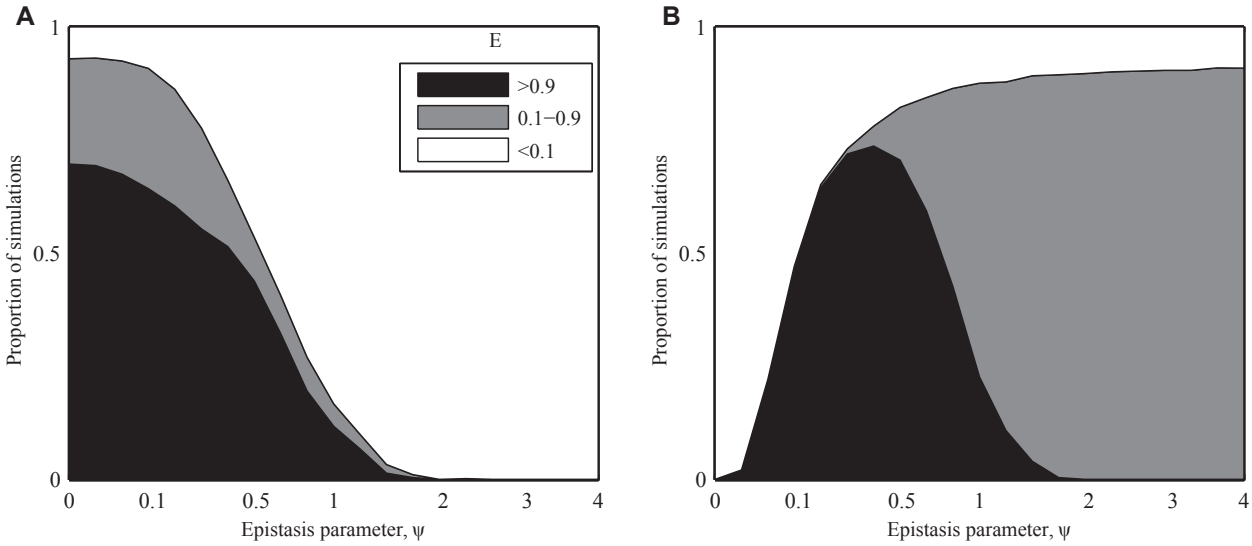

Figure 4. Proportion of stochastic simulations where peak infectivity range (E; eq. 4) was greater than 0.9 (black), less than 0.1 (white), or between these values (gray), for (a) gradual (SYM) and (b) sudden (ASYM) changes in host phenotype. The parameter ψ controls the type and strength of epistasis between infectivity alleles, ranging from strong positive ($\psi << 1$), through weak positive ($\psi < 1$), none ($\psi = 1$), and finally, negative ($\psi > 1$) epistasis. The patterns in the SYM scenario are broadly similar to those in the deterministic version of the model (Fig. 3A), but infectivity range is predicted to peak for weak positive epistasis in the ASYM scenario. This disparity can be explained by the low probability of fixing multiple mutations in the presence of strong positive epistasis.

values of ψ). This pattern was consistent for different types of fitness cost, variations in the rate at which fitness costs increased (i.e., accelerating, decelerating, or linear), different numbers of loci and density-dependent transmission (see Supplementary Material).

We extended the duration of 10% of our stochastic simulations to 200,000 time steps (a tenfold increase) to ensure that the differences between the two versions of the ASYM scenario were not attributable to faster evolution in the deterministic setting. However, longer simulations did not change the overall pattern of

our results and only led to relatively minor quantitative differences (average change in E: 0.03). Notably, parasites that experienced very strong positive epistasis ($\psi = 0$) were still unable to accumulate infectivity alleles, even over this longer time period.

The disparity between the deterministic and stochastic versions in the ASYM scenario can be explained by the low probability of fixing multiple mutations that were characterized by strong positive epistasis (i.e., parasites were trapped at a local fitness peak). Thus, although a complete set of infectivity alleles may have been optimal for low values of ψ (Figs. 2, 3B), parasites struggled to accumulate mutations that were not immediately beneficial. This situation did not occur in the SYM scenario, as individual mutations conferred an immediate increase in fitness due to the presence of genetically intermediate hosts (hosts exhibited gradual changes in phenotype). Figure 5 shows contrasting dynamics from the two scenarios. When hosts exhibit gradual changes in phenotype (SYM; Fig. 5A), the fitness of parasites with incomplete sets of infectivity alleles is greater than the wild type, which allows individual mutations to become fixed. When hosts exhibit sudden changes in phenotype (ASYM; Fig. 5B), parasites with the broadest ranges still have the highest fitness, but an incomplete set of infectivity alleles is costly, so mutations are unlikely to accumulate when demographic stochasticity is included.

Although broad infectivity ranges were much more common in the stochastic version of the SYM scenario than in the corresponding ASYM scenario for strong positive epistasis ($\psi << 1$), the converse was true for weak positive epistasis ($0.1 < \psi < 1$). This pattern can again be explained by less effective resistance in the host as epistasis between infectivity alleles weakens (Fig. S5C). Gradual changes in phenotype are less advantageous to the host as ψ increases, reducing the likelihood that resistance alleles will become fixed in a stochastic setting, which in turn reduces selection for parasites with broader infectivity ranges. Thus, for weak positive epistasis, overall levels of resistance will tend to be lower if hosts are restricted to gradual (SYM) rather than sudden (ASYM) changes in phenotype (SYM).

Discussion

Our study was inspired by recent experiments where gradual changes in host phenotype provided optimal conditions for parasite evolution, presumably by facilitating the accumulation of multiple mutations in quick succession, whereas sudden changes in host phenotype prevented mutations from being fixed (Benmayor et al. 2009; Paterson et al. 2010; Hall et al. 2011; Meyer et al. 2012). However, we hypothesized that the presence of strong positive epistasis between mutations was likely to be a key factor in these experiments and that alternative forms of epistasis could lead to different evolutionary outcomes. Using a theoretical approach, we explored how the rate of phenotypic change in the host (gradual vs. sudden) and the type and strength of epistasis shape the evolution of broader infectivity ranges. Our findings support empirical observations that gradual changes in host phenotype promote broad infectivity ranges, provided epistasis is strongly positive (Paterson et al. 2010; Hall et al. 2011; Meyer et al. 2012). Moreover, by comparing deterministic and stochastic models, we have shown why sudden changes in host phenotype can restrict parasite evolution (low probabilities of fixing individual mutations that confer no immediate increase in fitness). However, we have also shown that this prediction does not hold for weak positive epistasis: parasites are more likely to evolve broad infectivity ranges if hosts exhibit sudden changes in phenotype. These results demonstrate that the nature of epistasis can be crucial for shaping parasite evolution and that gradual changes in host phenotype may not always provide the optimal conditions for broad infectivity ranges to evolve.

When epistasis is strongly positive, parasites require a complete (or near-complete) set of mutations to overcome resistance. These are difficult to accumulate if the host makes sudden changes in phenotype (e.g., due to the loss of a key receptor) as individual mutations may carry an intrinsic fitness cost without conferring any benefits with regards to increased infectivity. If, however, hosts experience gradual changes in phenotype (e.g., reduced expression of a key receptor), then parasites may have access to hosts that are phenotypically intermediate between ancestral and future populations. Individual mutations may then confer an immediate fitness advantage, dramatically increasing the probability that multiple mutations will become fixed. Hence the symmetric (SYM) scenario in our model, which featured intermediate hosts, was much more favorable to the emergence of broad infectivity ranges under strong positive epistasis than the asymmetric (ASYM) scenario, which did not allow intermediate hosts to evolve. Conversely, when epistasis is negative there is a diminishing benefit associated with the acquisition of multiple infectivity alleles regardless of the potential presence of intermediate hosts. In other words, parasites can infect a reasonably broad set of hosts with a single mutation (e.g., Jonah et al. 2003; Brault et al. 2004) and the advantages of a full complement of mutations are outweighed by associated fitness costs. Between these extremes (i.e., for weak positive epistasis), individual infectivity alleles may confer a slight increase in fitness, allowing parasites to accumulate successive mutations. These mutations are likely to be under strong selection when hosts exhibit sudden changes in phenotype, but will be less beneficial if hosts evolve more gradually. Hence, infectivity range peaks for weak positive epistasis in our ASYM scenario and is more common than in the SYM scenario under these conditions.

The results for the ASYM scenario can also be interpreted in terms of a balance between mutation supply and selection for

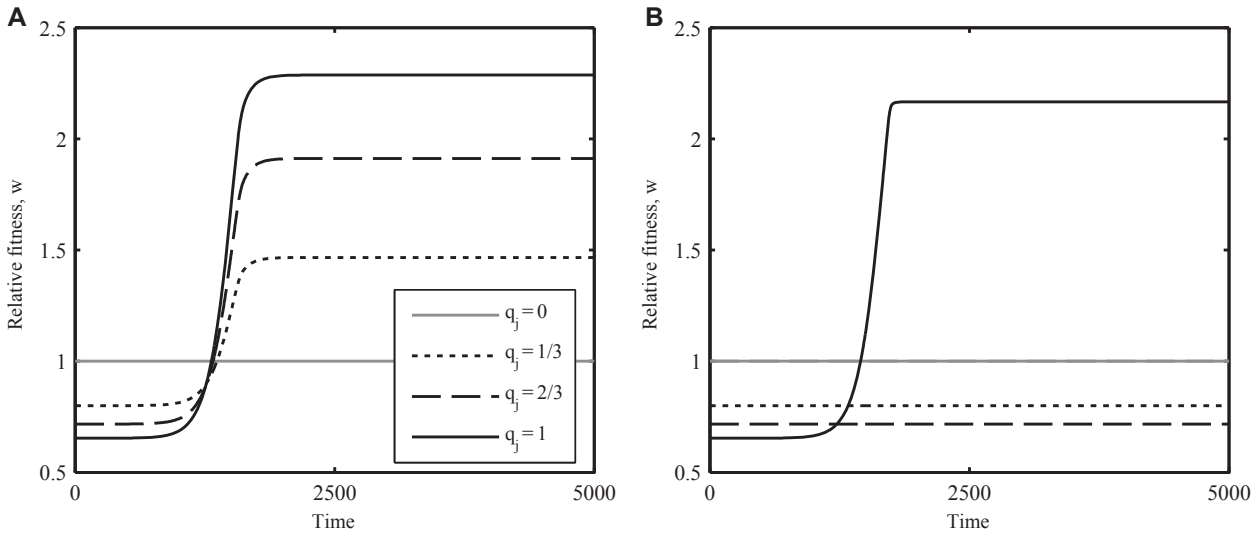

Figure 5. **Example dynamics showing the relative fitness (w; eq. 7) of parasites with different proportions of loci containing infectivity alleles ($q_j = 0$ (gray), 1/3 (dotted), 2/3 (dashed), and 1 (solid)) for (A) gradual (SYM) and (B) sudden (ASYM) changes in host phenotype (deterministic version). (A) When resistance begins to spread in a host population that exhibits gradual changes in phenotype (around 1000 time units) parasites with incomplete sets of infectivity alleles experience an immediate increase in fitness, which allows these mutations to be fixed, leading to the evolution of broad infectivity ranges. (B) If the host makes a sudden change in phenotype, then parasites with incomplete sets of infectivity alleles may not experience an immediate increase in fitness. Hence, infectivity alleles may not accumulate when demographic stochasticity is included, even though parasites with a full complement of infectivity alleles have the highest fitness. Parameters: $\alpha = 0.01$; $\beta = 0.25$; $\varepsilon_H = 10^{-6}$; $\varepsilon_P = 10^{-6}$; $\varphi_H = 0.5$; $\varphi_P = 0.5$; $\psi = 0$; $c_H = 1$; $c_P = 1.25$; $K = 10^6$; $n = 3$; $r = 0.05$.**

broader infectivity ranges. If parasites perform poorly on current hosts, then selection for broader infectivity ranges is strong, but the mutation supply is constrained due to a lack of suitable hosts. This problem is accentuated when epistasis between infectivity alleles is strongly positive, as multiple mutations that may be costly in isolation are required before parasites can infect large numbers of hosts. Hence, strong positive epistasis results in a low mutation supply and broader infectivity ranges are unlikely to evolve. Conversely, if parasites perform well on the current host population, then the mutation supply is much greater, but selection for broader infectivity ranges is inevitably weaker. When epistasis is negative, parasites with few infectivity alleles can perform reasonably well on hosts that have invested in resistance; accumulating further (costly) infectivity alleles is unlikely to be beneficial under these circumstances, so again, broader infectivity ranges do not evolve. Between these extremes, parasites can infect sufficient hosts to maintain a modest supply of mutations while experiencing fairly strong selection for broader infectivity ranges. Hence, broader infectivity ranges are most likely to evolve in the stochastic ASYM scenario for intermediate values of ψ, which corresponds to weak positive epistasis.

The discrepancies between the deterministic and stochastic versions of the ASYM scenario are striking, but were not attributable to different rates of evolution, as simulations that were allowed to run for much longer time periods produced very similar results (although over infinitely long time scales or with simultaneous mutations, the stochastic model should eventually produce pathogens with complete sets of infectivity alleles when epistasis is strongly positive). If hosts make sudden jumps in phenotype, then parasites that experience stochasticity and strong positive epistasis between infectivity alleles are likely to get stuck at a local fitness peak. In the deterministic model, parasites are able to explore the entire fitness landscape, which always allows the globally optimal phenotype to emerge. The contrasting outcomes in the deterministic and stochastic models highlight the need to compare different modeling approaches when studying host-parasite coevolution, as optimal phenotypes may not always emerge when demographic stochasticity is included (see also Ashby et al. 2014). It is important to note that our approach differs from most other theoretical studies of host-parasite coevolution, which typically omit ecological feedbacks and stochasticity, assume that population sizes are infinite and focus only on changes in gene frequencies (Sasaki 2000; Agrawal and Lively 2002, 2003; Fenton and Brockhurst 2007; Fenton et al. 2009, 2012). As a consequence, these studies do not explicitly model interactions between mutations, so it is not clear how ecological processes (and stochasticity) affect how they accumulate.

Our results were robust to changes in a number of modeling assumptions, which indicates that our findings are likely to be

quite general. Still, it is somewhat surprising that different forms of fitness costs (accelerating, decelerating, or linear) did not produce markedly different results, as different trade-off shapes are often associated with contrasting outcomes in studies of host and parasite evolution (Kisdi 2006; Best et al. 2010). This suggests that epistasis plays a dominant role in shaping host-parasite coevolution in our model.

The findings presented herein are consistent with recent empirical work showing that infectivity evolution tends to proceed faster in the presence of coevolving antagonists that exhibit gradual changes in phenotype (Poullain et al. 2008; Paterson et al. 2010; Schulte et al. 2010; Hall et al. 2011; Morran et al. 2011; Zhang and Buckling 2011). The host-parasite systems in these experiments appear to feature multiple reciprocal genetic adaptations, comparable to the symmetric (SYM) scenario in our model. Furthermore, some studies have focused specifically on the evolution of broad infectivity ranges, with parasites experiencing either coevolving or constant mixtures of host genotypes (Benmayor et al. 2009; Hall et al. 2011). The latter treatment is analogous to our asymmetric (ASYM) scenario, as parasites must adapt to a large phenotypic change in the host and do not have access to intermediate populations. Using the host bacterium *Pseudomonas fluorescens* and the lytic phage Φ2, Benmayor et al. (2009) manipulated the ratio of sensitive to resistant hosts and observed an apparent trade-off between mutation supply and selection for broader infectivity ranges, in much the same way as epistasis affected our ASYM scenario. Building on this work, Hall et al. (2011) demonstrated that phages evolved broader infectivity ranges when hosts were allowed to coevolve. This microbial system is known to exhibit strong positive epistasis between infectivity mutations (Scanlan et al. 2011), so these findings are in excellent agreement with the results of our stochastic simulations. In another recent study involving bacteria and viruses (Meyer et al. 2012), a lytic mutant of phage lambda required four mutations to infect resistant *Escherichia coli*. These mutations showed all-or-nothing (i.e., strongly positive) epistasis, but in contrast to studies on *P. fluorescens*, the acquisition of these mutations resulted from fitness benefits on a subpopulation of bacteria that had reverted to susceptibility. As such, broader infectivity ranges were again promoted by a gradual change in host phenotype, but this was facilitated by host polymorphism.

Several studies have explored the evolution of broader infectivity ranges, using either explicit genetics (as here; see also Sasaki 2000; Agrawal and Lively 2002, 2003; Fenton and Brockhurst 2007; Fenton et al. 2009, 2012) or by treating infectivity as a quantitative trait (e.g., Best et al. 2010). However, few theoretical studies have examined how different forms of epistasis influence parasite evolution. As an exception, Fenton and Brockhurst (2007) explored the role of accelerating, decelerating and linear fitness costs on coevolutionary dynamics in a GFG framework. Our study complements this work by focusing on the effects of epistasis on infectivity, while still allowing qualitatively different forms of fitness costs to exist. Similarly, studies focusing specifically on genetic factors that influence the evolution of infectivity ranges are rare. Poullain and Nuismer (2012) recently demonstrated that frameworks of infection genetics with overlapping ranges (e.g., gene-for-gene; Sasaki 2000) are better suited to allow adaptation to a novel host than if ranges are disjoint (e.g., matching alleles; Hamilton 1980), but the authors did not consider the effects of multilocus interactions, epistasis, or the rate of phenotypic change in the host on the evolution of infectivity ranges, as have been explored in the present study.

We have focused on how epistasis influences parasite evolution, as the emergence of broader infectivity ranges is of greater biological relevance to epidemiology and public health. Still, it is also important to understand how epistasis shapes host evolution, as this can elucidate patterns of selection among parasites (Supplementary Fig. S5). For example, hosts did not tend to become resistant in the stochastic SYM scenario when epistasis was negative, which explains why broader infectivity ranges did not evolve under these conditions. Host-parasite coevolution can lead to a variety of coevolutionary outcomes with respect to infectivity range, including competitive exclusion, stable polymorphism, and fluctuating selection (Sasaki 2000; Fenton and Brockhurst 2007). Indeed, our models produced fluctuations in range under certain conditions, but the primary purpose of our study has been to highlight conditions that promote or inhibit the accumulation of mutations that confer broader infectivity ranges, rather than whether they are evolutionarily stable. Hence, we chose to measure peak rather than final infectivity range, as the latter would have been heavily influenced by the choice of simulation length when fluctuations occurred. While evolutionary stability is often an important consideration, there are many circumstances where the initial emergence of a trait is more significant. For example, the chief concern in the context of emerging infectious diseases is whether a parasite will be able to cause an epidemic in a new population, rather than if it will survive over longer timescales. Compensatory mutations may also arise after a trait has initially spread, which could allow parasites to offset associated fitness costs. We hope to address questions surrounding the evolutionary stability of infectivity ranges under various genetic and ecological conditions in future work.

Understanding the effects of epistasis in real biological systems represents a significant challenge for empiricists and we are not aware of many nonmicrobial systems where the strength of interactions between infectivity mutations have been accurately measured (Hall and Ebert 2013). Finding a suitable host-parasite system to test the effects of weaker epistasis under different ecological conditions may prove to be difficult, but some of the predictions from our model should be relatively straightforward

to verify experimentally with current approaches that utilize coevolving and noncoevolving communities.

ACKNOWLEDGMENTS

We thank two anonymous reviewers for extremely helpful comments on the manuscript. The work was funded by NERC and BBSRC. SG acknowledges support from the European Research Council (DIVERSITY). SG and AB are Royal Society Wolfson Research Fellows. The authors would like to acknowledge the use of the Advanced Research Computing (ARC) in carrying out this work.

DATA ARCHIVING

The doi for our data is 10.5061/dryad.25kn0.

LITERATURE CITED

Agrawal, A. F., and C. M. Lively. 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evol. Ecol. Res. 4:79–90.

———. 2003. Modelling infection as a two-step process combining gene-for-gene and matching-allele genetics. Proc. R. Soc. B 270:323–334.

Ashby, B., S. Gupta, and A. Buckling. 2014. Spatial structure mitigates fitness costs in host-parasite coevolution. Am. Nat. 183:E64–E74.

Benmayor, R., D. J. Hodgson, G. G. Perron, and A. Buckling. 2009. Host mixing and disease emergence. Curr. Biol. 19:764–767.

Best, A., A. White, E. Kisdi, J. Antonovics, M. A. Brockhurst, and M. Boots. 2010. The evolution of host-parasite range. Am. Nat. 176:63–71.

Bohannan, B. J. M., B. Kerr, C. M. Jessup, J. B. Hughes, and G. Sandvik. 2002. Trade-offs and coexistence in microbial microcosms. Antonie Van Leeuwenhoek 81:107–115.

Brault, A. C., A. M. Powers, D. Ortiz, J. G. Estrada-Franco, R. Navarro-Lopez, and S. C. Weaver. 2004. Venezuelan equine encephalitis emergence: enhanced vector infection from a single amino acid substitution in the envelope glycoprotein. Proc. Natl. Acad. Sci. USA 101:11344–11349.

Brown, J. K. M., and A. Tellier. 2011. Plant-parasite coevolution: bridging the gap between genetics and ecology. Annu. Rev. Phytopathol. 49:345–367.

Buckling, A., and P. B. Rainey. 2002. Antagonistic coevolution between a bacterium and a bacteriophage. Proc. R. Soc. B 269:931–936.

Chao, L., B. R. Levin, and F. M. Stewart. 1977. A complex community in a simple habitat: an experimental study with bacteria and phage. Ecology 58:369–378.

Decaestecker, E., S. Gaba, J. A. M. Raeymaekers, R. Stoks, L. Van Kerckhoven, D. Ebert, and L. De Meester. 2007. Host-parasite "Red Queen" dynamics archived in pond sediment. Nature 450:870–873.

Dybdahl, M. F., and C. M. Lively. 1998. Host-parasite coevolution: evidence for rare advantage and time-lagged selection in a natural population. Evolution 52:1057–1066.

Fenton, A., J. Antonovics, and M. A. Brockhurst. 2009. Inverse-gene-for-gene infection genetics and coevolutionary dynamics. Am. Nat. 174:E230–E242.

Fenton, A., J. Antonovics, and M. A. Brockhurst. 2012. Two-step infection processes can lead to coevolution between functionally independent infection and resistance pathways. Evolution 66:2030–2041.

Fenton, A., and M. A. Brockhurst. 2007. Epistatic interactions alter dynamics of multilocus gene-for-gene coevolution. PLoS One 2:e1156.

Gandon, S. 2002. Local adaptation and the geometry of host—parasite coevolution. Ecol. Lett. 5:246–256.

Gandon, S., and S. L. Nuismer. 2009. Interactions between genetic drift, gene flow, and selection mosaics drive parasite local adaptation. Am. Nat. 173:212–224.

Gillespie, D. T. 2001. Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. 115:1716–1733.

Gururani, M. A., J. Venkatesh, C. P. Upadhyaya, A. Nookaraju, S. K. Pandey, and S. W. Park. 2012. Plant disease resistance genes: current status and future directions. Physiol. Mol. Plant Pathol. 78:51–65.

Hall, A. R., P. D. Scanlan, and A. Buckling. 2011. Bacteria-phage coevolution and the emergence of generalist pathogens. Am. Nat. 177:44–53.

Hall, M. D., and D. Ebert. 2013. The genetics of infectious disease susceptibility: has the evidence for epistasis been overestimated? BMC Biol. 11:79.

Hamilton, W. D. 1980. Sex versus non-sex versus parasite. Oikos 35:282–290.

Jonah, G., A. Rainey, A. Natonson, L. F. Maxfield, and J. M. Coffin. 2003. Mechanisms of avian retroviral host range extension. J. Virol. 77: 6709–6719.

Kisdi, E. 2006. Trade-off geometries and the adaptive dynamics of two coevolving species. Evol. Ecol. Res. 8:959–973.

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

Labrie, S. J., J. E. Samson, and S. Moineau. 2010. Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. 8:317–327.

Lenski, R. E. 1984. Two-step resistance by *Escherichia coli* B to bacteriophage T2. Genetics 107:1–7.

Levin, B. R., and J. J. Bull. 2004. Population and evolutionary dynamics of phage therapy. Nat. Rev. Microbiol. 2:166–173.

Meyer, J. R., D. T. Dobias, J. S. Weitz, J. E. Barrick, R. T. Quick, and R. E. Lenski. 2012. Repeatability and contingency in the evolution of a key innovation in phage lambda. Science 335:428–432.

Mizoguchi, K., M. Morita, C. R. Fischer, M. Yoichi, Y. Tanji, and H. Unno. 2003. Coevolution of bacteriophage PP01 and *Escherichia coli* O157: H7 in continuous culture. Appl. Environ. Microbiol. 69: 170–176.

Morran, L. T., O. G. Schmidt, I. A. Gelarden, R. C. Parrish, and C. M. Lively. 2011. Running with the red queen: host-parasite coevolution selects for biparental sex. Science 333:216–218.

Paterson, S., T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M. Quail, F. Smith, D. Walker, B. Libberton, et al. 2010. Antagonistic coevolution accelerates molecular evolution. Nature 464: 275–278.

Poullain, V., S. Gandon, M. A. Brockhurst, and A. Buckling. 2008. The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage. Evolution 62:1–11.

Poullain, V., and S. L. Nuismer. 2012. Infection genetics and the likelihood of host shifts in coevolving host-parasite interactions. Am. Nat. 180:618–628.

Russell, C. A., J. M. Fonville, A. E. Brown, D. F. Burke, D. L. Smith, S. L. James, S. Herfst, S. van Boheemen, M. Linster, E. J. Schrauwen, et al. 2012. The potential for respiratory droplet–transmissible A/H5N1 influenza virus to evolve in a mammalian host. Science 336: 1541–1547.

Sasaki, A. 2000. Host-parasite coevolution in a multilocus gene-for-gene system. Proc. R. Soc. B 267:2183–2188.

Scanlan, P. D., A. R. Hall, L. D. C. Lopez-Pascua, and A. Buckling. 2011. Genetic basis of infectivity evolution in a bacteriophage. Mol. Ecol. 25:1–9.

Schulte, R. D., C. Makus, B. Hasert, N. K. Michiels, and H. Schulenburg. 2010. Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. Proc. Natl. Acad. Sci. USA 107:7359–7364.

Tait, K., L. C. Skillman, and I. W. Sutherland. 2002. The efficacy of bacteriophage as a method of biofilm eradication. Biofouling 18:305–311.

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

Webster, J. P., and M. E. J. Woolhouse. 1999. Cost of resistance: relationship between reduced fertility and increased resistance in a snail-schistosome host-parasite system. Proc. R. Soc. B 266:391–396.

Wilfert, L., and P. Schmid-Hempel. 2008. The genetic architecture of susceptibility to parasites. BMC Evol. Biol. 8:187.

Zhang, Q.-G., and A. Buckling. 2011. Antagonistic coevolution limits population persistence of a virus in a thermally deteriorating environment. Ecol. Lett. 14:282–288.

Associate Editor: E. Kisdi

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. Proportion of simulations where peak infectivity range (E; eq. 4) was greater than 0.9 (black), less than 0.1 (white), or between these values (gray) for $n = 4$.

Figure S2. Proportion of simulations where peak infectivity range (E; eq. 4) was greater than 0.9 (black), less than 0.1 (white), or between these values (gray) for density-dependent transmission (DD).

Figure S3. Proportion of simulations where peak infectivity range (E; eq. 4) was greater than 0.9 (black), less than 0.1 (white), or between these values (gray) for (a–d) linear ($\varphi_H=\varphi_P=1$) and (e–h) accelerating ($\varphi_H=\varphi_P=2$) fitness costs.

Figure S4. Proportion of simulations where peak infectivity range (E; eq. 4) was greater than 0.1 (black), less than 0.1 (white), or between these values (gray), for different combinations of fitness costs: (a–d) natural mortality (μ_i) and transmission (β_j) rates; (e–h) birth (r_i) and disease-associated mortality (α_j) rates; (i–l) natural (μ_i) and disease-associated (α_j) mortality rates.

Figure S5. Proportion of simulations where peak resistance range for hosts was greater than 0.9 (black), less than 0.1 (white) or between these values (grey). The measure used here for hosts is analagous to the one described for parasites in equation 4, with resistance substituted for infectivity.

vol. 183, no. 3 the american naturalist march 2014

E-Article

Spatial Structure Mitigates Fitness Costs in Host-Parasite Coevolution

Ben Ashby,[1,*] Sunetra Gupta,[1] and Angus Buckling[2]

1. Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom; 2. Biosciences, University of Exeter, Cornwall Campus, Penryn TR10 9EZ, United Kingdom

Submitted February 22, 2013; Accepted August 21, 2013; Electronically published January 14, 2014

Online enhancements: appendixes, code files.

abstract: The extent of population mixing is known to influence the coevolutionary outcomes of many host and parasite traits, including the evolution of generalism (the ability to resist or infect a broad range of genotypes). While the segregation of populations into interconnected demes has been shown to influence the evolution of generalism, the role of local interactions between individuals is unclear. Here, we combine an individual-based model of microbial communities with a well-established framework of genetic specificity that matches empirical observations of bacterium-phage interactions. We find the evolution of generalism in well-mixed populations to be highly sensitive to the severity of associated fitness costs, but the constraining effect of costs on the evolution of generalism is lessened in spatially structured populations. The contrasting outcomes between the two environments can be explained by different scales of competition (i.e., global vs. local). These findings suggest that local interactions may have important effects on the evolution of generalism in host-parasite interactions, particularly in the presence of high fitness costs.

Keywords: host-parasite coevolution, spatial structure, resistance and infectivity range, fitness costs.

Introduction

Antagonistic coevolution between hosts and parasites is often associated with the emergence of generalism, where populations develop the ability to resist or infect a broad range of genotypes. This means that contemporary populations may be well adapted to ancestral lineages but perform poorly against future populations (Buckling and Rainey 2002*a*; Mizoguchi et al. 2003; Scanlan et al. 2011). The fundamental principles of these "coevolutionary arms races" are captured by the gene-for-gene (GFG) framework, in which hosts can avoid infection by accumulating resistance alleles at multiple loci but parasites can counter these adaptations by gaining infectivity alleles at matching loci (Flor 1956; Sasaki 2000). Hence there is a gene-for-gene correspondence between resistance and infectivity alleles. (Note that the literature often refers to these parasite adaptations as "virulence" alleles, but to avoid confusion with disease severity we will refer to them as "infectivity" alleles instead). Under the GFG framework, parasites must match or exceed the host's resistance alleles at each locus to have a high probability of causing an infection, which naturally leads to the evolution of generalism in the form of broader resistance and infectivity ranges. These dynamics have been observed in a variety of real host-parasite relationships, including bacterium-phage (Bohannan and Lenski 2000; Buckling and Rainey 2002*a*; Mizoguchi et al. 2003; Brockhurst et al. 2006; Forde et al. 2008; Scanlan et al. 2011), plant-pathogen (Flor 1956; Thompson and Burdon 1992; Thrall and Burdon 2003) and nematode-bacterium systems (Schulte et al. 2010). Recent studies of bacterium-phage coevolution have found that infectivity range is correlated with the number of amino acid changes in tail fibers relative to the ancestral genotype (Scanlan et al. 2011), providing further support for the GFG framework. However, coevolutionary arms races are unlikely to be maintained indefinitely as fitness costs associated with generalism (usually in the form of lower growth/infectivity rates) can reduce selection for broad ranges (Chao et al. 1977; Webster and Woolhouse 1999; Sasaki 2000; Bohannan et al. 2002; Lopez-Pascua and Buckling 2008; Poullain et al. 2008). Sasaki (2000) predicted that fitness costs will lead to fluctuations between specialism (narrow range) and generalism (broad range), but empirical observations suggest that fitness costs may instead lead to fluctuating selection among genotypes with similar ranges (Hall et al. 2011).

Fitness costs clearly have considerable influence on the extent of range expansion among hosts and parasites, but other factors are known to have an equally profound im-

* Corresponding author; e-mail: ben.ashby@zoo.ox.ac.uk.

Am. Nat. 2014. Vol. 183, pp. E64–E74.
DOI: 10.1086/674826

pact on coevolutionary dynamics. In particular, it is well established that spatial structure affects both epidemiological dynamics and the scale of competition within a population (Thrall and Burdon 2003; Forde et al. 2004, 2007; Morgan et al. 2007) and can allow polymorphism to be maintained even in the absence of fitness costs (Damgaard 1999). In a spatially structured environment the optimal genotype for a particular location will depend on local selection pressures, which may differ between locations and from what would be considered the globally optimal genotype in a well-mixed population (Thompson 1994). Experiments with the bacterium *Pseudomonas fluorescens* and the lytic phage Φ2 suggest that while limited population mixing will slow down the rate of coevolution (Brockhurst et al. 2003), it may also provide more stable conditions for coexistence (Brockhurst et al. 2006). Most empirical studies exploring the effects of spatial structure on range expansion have focused on scenarios where the population is split into interconnected demes, mainly to address questions associated with local adaptation (Burdon and Thrall 1999; Thrall and Burdon 2002, 2003; Forde et al. 2004, 2007; Morgan et al. 2007). Similarly, theoretical studies have generally been limited to metapopulation analyses (Frank 1993; Gandon et al. 1996, 2008; Damgaard 1999), which incorporate a certain degree of spatial structure but do not capture local interactions between individuals within subpopulations, which are known to be critical in many epidemiological scenarios (Rand et al. 1995; Rhodes and Anderson 1996; Keeling et al. 2001; Eames and Keeling 2002). Individual-based models are able to capture local interactions and have been used to study a diverse set of biological phenomena including the evolution of life histories and virulence (Boots and Sasaki 1999; Haraguchi and Sasaki 2000; Read and Keeling 2003; Heilmann et al. 2010), altruism (Jansen and van Baalen 2006), and various other aspects of coevolution (Hartvigsen and Levin 1997; Kerr et al. 2006; Mitchell et al. 2006; Best et al. 2011; Haerter et al. 2011; Zaman et al. 2011; Heilmann et al. 2012). However, the role of local interactions on range expansion has yet to be determined. Here, we attempt to address this gap in the literature by adapting an individual-based model of bacteria and phages first proposed by Heilmann et al. (2010). Although the model was originally used to explore the evolution of virulence in spatially structured populations, it can be readily adapted to serve our focus on range expansion by incorporating the multilocus GFG framework of Sasaki (2000). The model implements spatial structure by situating hosts and parasites on a two-dimensional grid, which is of particular relevance to bacteria (Kerr et al. 2006; Hellweger and Bucci 2009) as colonies often live attached to surfaces in biofilms (Matz et al. 2005; Faruque et al. 2006), providing potential spatial refuges to infection by phages (Levin and Bull 2004; Gallet et al. 2009).

Our primary focus in this study is to explore how the impact of fitness costs associated with range expansion is affected by the degree of population mixing. We show that global competition in well-mixed populations leads to rapid selective sweeps, preventing range expansion at high fitness costs. In spatially structured environments however, we find that local competition and spatial clustering can maintain selection for broader ranges even when fitness costs are high.

Methods: Model Description

Genetic Specificity

The genetic specificity of our model is based on the multilocus GFG framework proposed by Sasaki (2000). Host and parasite genotypes are represented by binary strings of length n ($h_1^i \ \ldots \ h_n^i$ for host genotype i and $p_1^j \ \ldots \ p_n^j$ for parasite genotype j), where each locus corresponds to the presence (1) or absence (0) of a resistance or infectivity allele. For example, a string of 000 represents a highly susceptible host (or a specialist parasite), whereas a string of 111 represents a highly resistant host (or a generalist parasite). We follow Sasaki (2000) by assuming a resistance allele at a particular locus is only effective against parasites that do not have a corresponding infectivity allele at that location and each effective resistance allele reduces the probability of infection by a factor of σ. The parameter σ represents the strength of resistance conferred by each locus: when $\sigma \approx 0$, the acquisition of a single resistance allele will lead to a strong reduction in susceptibility, but when $\sigma \approx 1$, each allele has only a mild effect. We define Q_{ij} to be the infectivity of parasite j on host i, such that

$$Q_{ij} = \sigma^{d_{ij}}, \ d_{ij} = \sum_{k=1}^{n} h_k^i(1 - p_k^j), \tag{1}$$

where d_{ij} is the sum of effective resistance alleles.

Simulation Rules

We adapt the bacterium-phage model proposed by Heilmann et al. (2010) to incorporate the GFG framework outlined above, thus allowing the evolution of varying degrees of generalism. We conduct simulations on a square grid of side length $N = 100$, where boundary effects are removed by wrapping the grid around the surface of a torus, so that all grid sites have exactly four orthogonal neighbors. A maximum of one host is allowed per grid site, so that each location is either empty or contains an infected or uninfected host; there are no restrictions on parasite density. The initial grid consists of uninfected

hosts at every site and 500 parasites at one location; both populations start without any resistance or infectivity alleles (i.e., $\sum_k h_k^i = \sum_k p_k^j = 0$). The grid is updated synchronously at the end of each time step. We implement two versions of our model (spatial and well-mixed) based on the following rules, which mostly follow those of Heilmann et al. (2010):

Host Replication. Spatial version. A host is only able to replicate if it satisfies the following criteria: (i) The host is uninfected, and (ii) at least T time steps have elapsed since the host's previous replication event (tracked by individual replication timers). If these criteria are satisfied, then replication proceeds with probability $Ec_H(i)$, where E is the proportion of empty grid sites adjacent to the host and $c_H(i)$ is a fitness cost associated with resistance, given by

$$c_H(i) = \exp(-\eta_H|i|), \tag{2}$$

with $|i| = \sum_k h_k^i$ equal to the total number of resistance alleles for host genotype i and η_H a scaling parameter for the strength of the fitness cost. Note that fitness costs were not included in the original model by Heilmann et al. (2010) due to the absence of host range expansion. Offspring are placed in a randomly chosen empty grid site adjacent to their parent; if multiple offspring attempt to occupy the same grid location, then one is chosen at random to survive and the others are removed from the population. The replication timers for successful parents and offspring are reset following this procedure. Mutations occur with probability ε_H at each locus, with the restriction that parents and offspring can only differ by one bit.

Well-mixed version. As per the spatial version, except that (i) the probability of replication is equal to $\overline{E}c_H(i)$, where $\overline{E}$ is the proportion of sites across the entire grid that are empty, and (ii) new offspring are placed at randomly chosen empty grid sites.

Infection. Both versions. We modify the overall probability of infection derived by Heilmann et al. (2010) to allow competition between multiple host and parasite genotypes. Given a probability of infection (α) and decay (δ) per free parasite, the probability that genotype j is able to infect host genotype i is given by

$$\rho_\alpha(i, j) = 1 - \exp\left\{-\alpha P(j) c_P(j) Q_{ij}\left[\frac{1 - \exp(-\delta)}{\delta}\right]\right\}, \tag{3}$$

where Q_{ij} is the strength of interaction between host and parasite and $P(j)$ is the local density of the parasite. Broader infectivity ranges are associated with fitness costs, which reduce the probability of infection, captured here by $c_P(j) = \exp(-\eta_P|j|)$, where η_P scales the strength of the fitness cost and $|j| = \sum_k p_k^j$ is the total number of infectivity alleles for the parasite. The probability that at least one parasite is able to infect the host is given by

$$z_1(i) = 1 - \prod_k (1 - \rho_\alpha(i, k)). \tag{4}$$

If a uniform random number, $\text{RAND}_1 \in (0, 1)$, satisfies $\text{RAND}_1 < z_1(i)$, then one parasite strain is chosen at random to infect the host. The probability of parasite j causing the infection is then equal to $\rho_\alpha(i, j)/\sum_k \rho_\alpha(i, k)$. We assume that coinfection does not occur.

Parasite Decay. Both versions. Free parasites decay with probability $1 - \exp(-\delta)$ and are immediately removed from the environment.

Parasite Diffusion. Both versions. We assume that parasites move between adjacent grid sites with probability proportional to the negative concentration gradient between those locations, with diffusion constant D. Parasites are restricted to one grid site movement per time step.

Parasite-Induced Host Mortality. Spatial version. Infected hosts are killed after a fixed number of time steps (latent period), τ, which results in the release of β new parasites into the environment. New parasites are placed at the same grid site as the newly deceased host. If a uniform random number, $\text{RAND}_2 \in (0, 1)$, satisfies $\text{RAND}_2 < n\varepsilon_P$, then one of the new parasites mutates at a random locus, gaining or losing an infectivity allele accordingly.

Well-mixed version. As per the spatial version, except parasites are distributed to randomly chosen grid sites.

Natural Host Mortality. Both versions. All hosts die with probability μ per time step. Parasites are not released into the environment if an infected host dies before the full latent period has expired.

Host Mixing. Well-mixed version only. Hosts are randomly assigned new grid sites at the end of each time step.

Analysis

We draw all parameters from uniform random distributions (table A1; tables A1–A4 available online), except for the strength of the fitness cost for hosts (η_H), the probability of natural death (μ) and the number of loci (n). A total of 500 simulations are conducted for each combination (s) of these parameters in both spatially structured and well-mixed environments. We run simulations for a maximum of 10,000 time steps and parameter combinations where hosts or parasites die out in either environment are discarded from further analysis. We allow a burn-

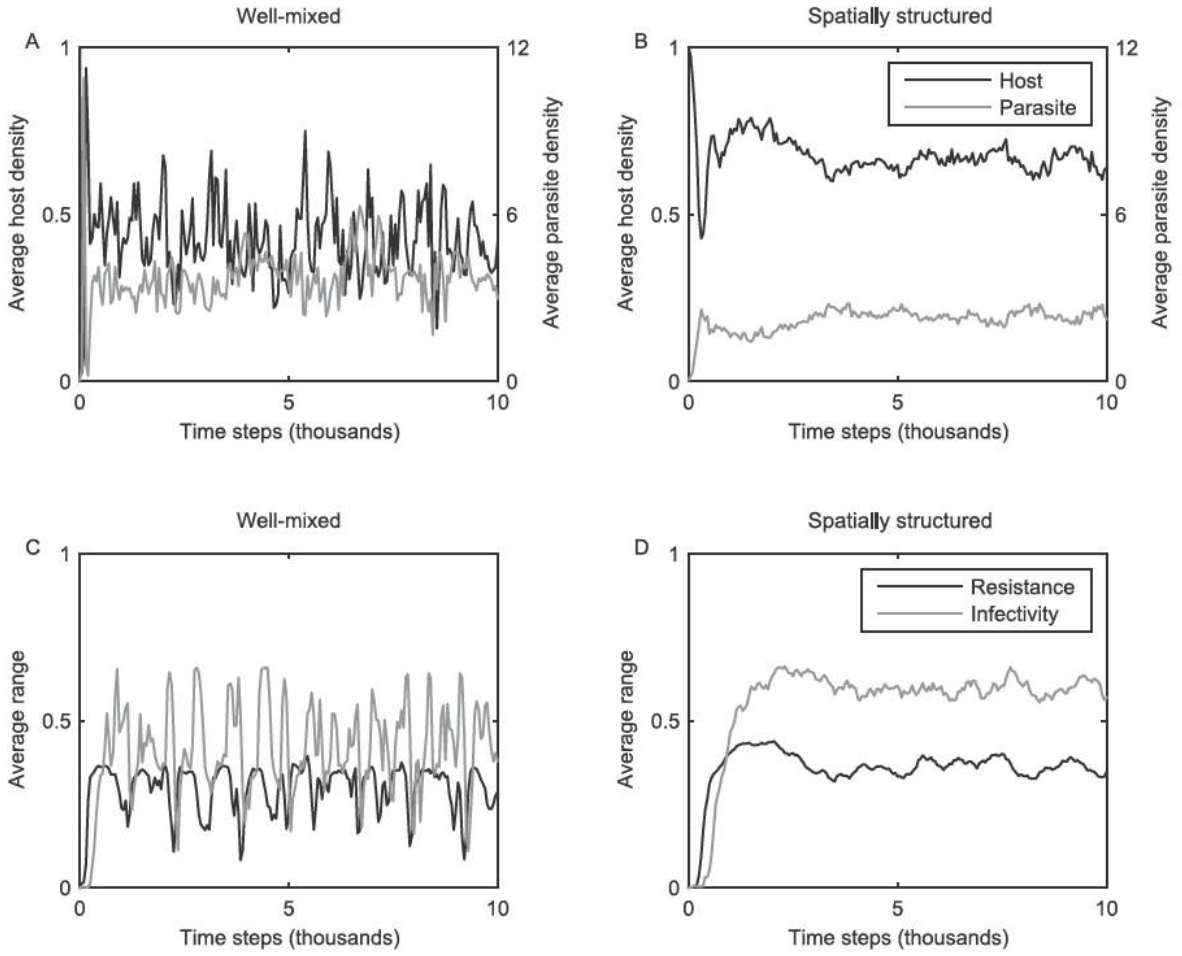

Figure 1: Example coevolutionary dynamics from well-mixed (*left panels*) and spatially structured (*right panels*) populations. *A*, *B*, Average density of host (black) and parasite (gray) populations. *C*, *D*, Average resistance (black) and infectivity (gray) ranges. Well-mixed populations are prone to rapid fluctuating dynamics, whereas spatially structured populations generally produce more stable dynamics. Parameters: $\alpha = 0.15$; $\beta = 50$; $\delta = 0.75$; $\varepsilon_H = 0.001$; $\varepsilon_P = 0.01$; $\eta_H = 1$; $\eta_P = 0.4$; $\mu = 0.001$; $\sigma = 0.25$; $\tau = 3$; $D = 0.5$; $N = 100$; $n = 3$; $T = 3$.

in period for the first half of each simulation before measuring the mean resistance, $R^*(s)$, and infectivity, $I^*(s)$, ranges over the final 5,000 time steps. Each measure varies between 0 and 1, where 0 indicates no range expansion by any members of the population and 1 corresponds to full range expansion by all members of the population. The resistance range of the host population at time step t is given by $R(s, t) = \sum_{i=1}^{n} ix(i, s, t)/n$, where $x(i, s, t)$ is the proportion of the host population that has i resistance alleles. Similarly, the infectivity range of the parasite population at time step t is given by $I(s, t) = \sum_{i=1}^{n} iy(i, s, t)/n$, where $y(i, s, t)$ is the proportion of the parasite population that has i infectivity alleles. We $\log_{10}$ transform the data and use ANCOVA to explore the extent to which spatial structure influences range expansion. In addition, we define the following two pairwise comparisons to visualize differences between the environments:

$$C_R(s) = \frac{R^*_{\text{mixed}}(s) - R^*_{\text{spatial}}(s)}{R^*_{\text{mixed}}(s) + R^*_{\text{spatial}}(s)}, \tag{5}$$

$$C_I(s) = \frac{I^*_{\text{mixed}}(s) - I^*_{\text{spatial}}(s)}{I^*_{\text{mixed}}(s) + I^*_{\text{spatial}}(s)}, \tag{6}$$

where the subscripts "mixed" and "spatial" correspond to the two environments. These measures vary between -1 and 1, indicating whether ranges are on average broader (>0) or narrower (<0) in well-mixed environments. Values close to the extremes correspond to a large disparity in evolutionary outcomes between the environments.

Finally, we measure the "patchiness" of spatially struc-

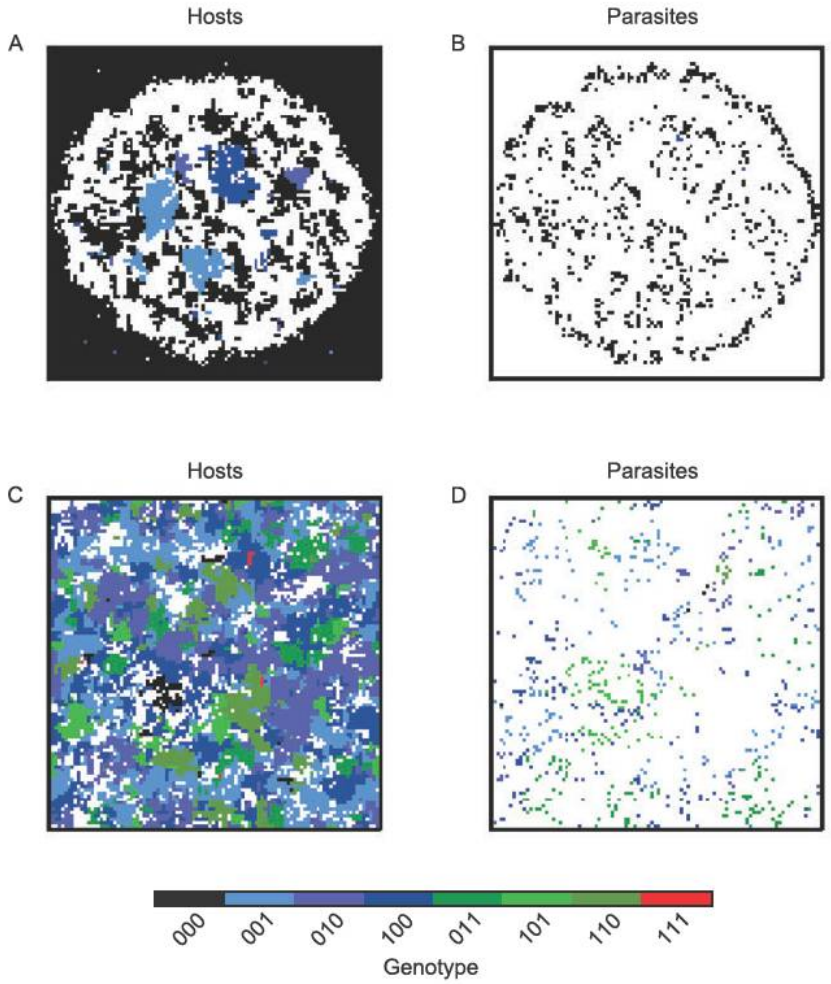

Figure 2: Snapshots of a simulation from the spatially structured population in figure 1 at two time points: *A*, *B*, 250 generations; *C*, *D*, 1,000 generations. Hosts (*left panels*) and intracellular parasites (*right panels*) are colored according to their genotype (i.e., number of resistance/infectivity alleles, where $n = 3$), with black = no alleles (000); blue = 1 allele (001, 01,0 or 100); green = 2 alleles (011, 101, or 110); and red = 3 alleles (111). White regions indicate no hosts/intracellular parasites. The Shannon index for the host population, $H(t)$, is equal to 2.19 (*A*) and 5.19 (*C*) at these two time points.

tured populations to assess how the distribution of host patch sizes varies with time. At a given time step, each patch, q, is defined by a unique set of orthogonally connected hosts sharing the same genotype. We measure the richness and evenness of these patches using the Shannon index, $H(s, t) = -\sum_q a(q, s, t) \ln a(q, s, t)$, where $a(q, s, t)$ is the number of hosts in patch q (Shannon 1948). This index increases as the number of patches grows and/or the distribution of patch sizes becomes more even. We define the "relative patchiness" of the population to be $H'(s, t) = H(s, t)/\overline{H}(s)$, where $\overline{H}(s)$ is the average value of $H(s, t)$ over the final 5,000 time steps of the simulation. Hence, if H' is consistently close to 1, then the patchiness of the population has reached a stable distribution.

Probabilistic Cellular Automata

We also explore a simplified version of our primary model using probabilistic cellular automata (PCA), which does not permit the existence of free-living parasites in the environment (Bak and Chao 1990). The model structure and analysis for the PCA are largely similar to the methods described above for our primary model; the full PCA simulation rules, MATLAB code, analysis, and results are detailed in appendix B (apps. A and B available online).[1]

[1] Code that appears in the *American Naturalist* is provided as a convenience to the readers. It has not necessarily been tested as part of the peer review.

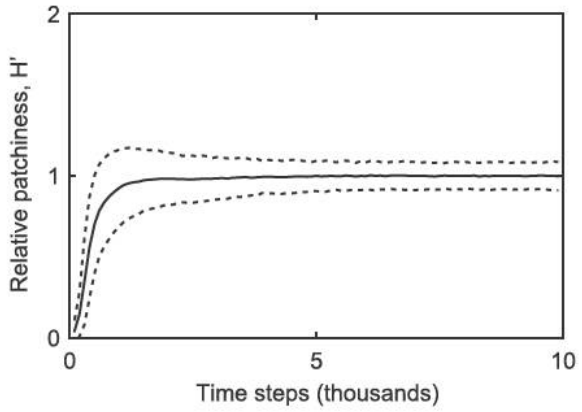

Figure 3: Relative patchiness of spatially structured populations as a function of time (mean ± standard deviation) for $n = 3$. Patchiness is defined by the Shannon index (as described in the main text), which measures the richness and evenness of patch sizes at a given time point; relative patchiness is equal to this value divided by the average over the final 5,000 generations, giving an indication as to how the spatial structure of the population tends to change during the course of a simulation. The first 1,000 generations are characterized by a rapid increase in patchiness, which then plateaus. The standard deviation does not increase over the final 5,000 generations, which implies that patchiness remains roughly constant within each simulation over this period.

Results

For the sake of brevity we present data for only $n = 3$ here, but similar results were obtained for other numbers of loci (see app. A).

Epidemiological Dynamics. We began our investigation by qualitatively comparing epidemiological dynamics in the presence and absence of spatial structure. Overall, 63% of simulations led to host-parasite coexistence for 10,000 generations in both environments. We observed that well-mixed environments were prone to large-amplitude epidemic cycles, whereas spatially structured populations typically demonstrated much steadier dynamics (fig. 1). The patchiness of spatially structured populations (H) typically increased sharply during the first 1,000 time steps of a simulation, before settling down to a stable distribution (figs. 2, 3). The snapshots in figure 2 demonstrate how these populations are typically structured into clusters of identical genotypes.

Resistance. We found that the evolution of broad resistance ranges was highly constrained by the severity of fitness costs (η_H) and to a lesser extent by the probability of natural mortality (μ; fig. 4*A*). When $\mu = 0.001$, R^* was on average 80% lower in the presence of high fitness costs ($\eta_H = 2$) compared to no fitness costs. This difference increased to 88% for $\mu = 0.01$ and to 98% for $\mu = 0.05$, as higher values of μ led to a faster turnover in the host population, reducing the benefits associated with broad resistance ranges.

While both spatially structured and well-mixed populations experienced a decrease in R^* with η_H, the magnitude of the constraint was generally much greater in the absence of spatial structure (figs. 4*B*, 6*A*; ANCOVA,

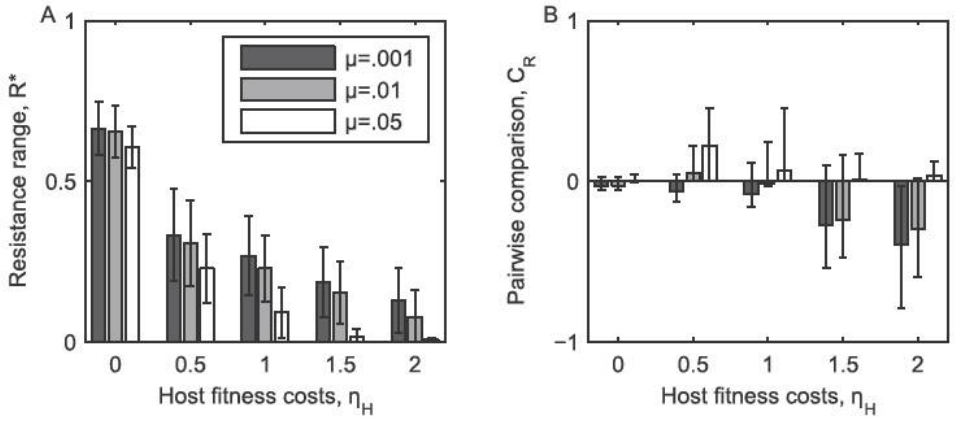

Figure 4: The effects of fitness costs, spatial structure, and the natural mortality rate on the resistance range of the host population for $n = 3$. *A*, The average resistance range (taken over both environments), R^*, tends to decrease with the host fitness cost parameter, η_H, and the natural mortality rate, μ. *B*, A pairwise comparison, C_R, of resistance range expansion in the two environments. If $C_R > 0$, then the mixed environment tends to produce broader resistance ranges than the spatially structured environment and vice versa if $C_R < 0$. Spatially structured populations generally exhibit broader resistance ranges than well-mixed populations as the associated fitness costs increase. However, high natural mortality rates can negate this effect, as resistance becomes less beneficial in both environments.

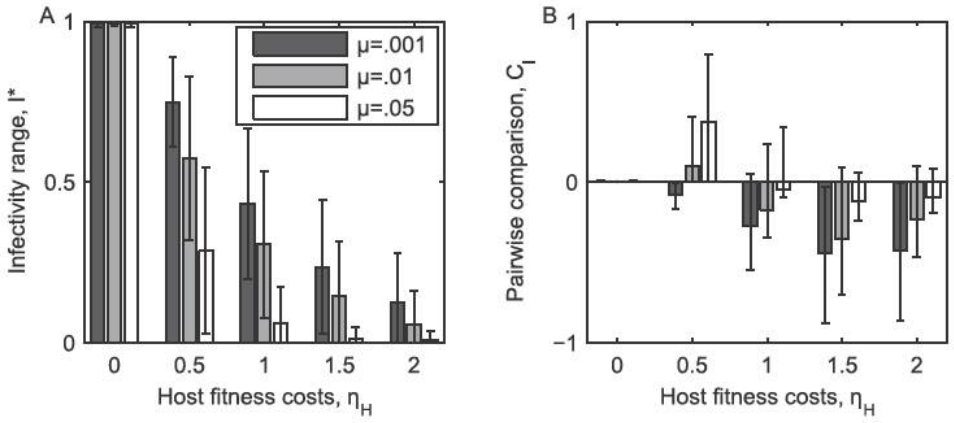

Figure 5: The effects of fitness costs, spatial structure, and the natural mortality rate on the infectivity range of the parasite population for $n = 3$. *A*, The average infectivity range (taken over both environments), I^*, tends to decrease with the host fitness cost parameter, η_H, and the natural mortality rate, μ. *B*, A pairwise comparison, C_I, of infectivity range expansion in the two environments. If $C_I > 0$, then the mixed environment tends to produce broader infectivity ranges than the spatially structured environment and vice versa if $C_I < 0$. Spatially structured populations generally exhibit broader infectivity ranges than well-mixed populations as the fitness costs for hosts increase. However, high natural mortality rates can negate this effect, as selection for infectivity decreases in both environments.

environment × fitness cost interaction for $\mu = 0.001$: $F_{1,\,3,234} = 194$, $P < .0001$; see table A3 for full ANCOVA results). The difference in evolutionary outcomes in the two environments is highlighted by the pairwise comparison (C_R) in figure 4*B*, which shows that spatially structured populations tended to evolve broader resistance ranges than comparable well-mixed populations and that this disparity increased with greater fitness costs, provided the turnover rate of the host population was not too fast. For sufficiently large values of μ, the two environments exhibited broadly similar levels of resistance in the presence of high fitness costs.

We examined the robustness of our results by varying the period during which resistance ranges were recorded. The median difference in resistance ranges between time steps 5,001–7,500 and 7,500–10,000 was 2.8%, indicating that the host population had reached a quasi-steady state prior to the beginning of the measurement window. Minor oscillations were often observed about the mean values of R^* (as shown in fig. 1), with slightly larger fluctuations more likely to occur in well-mixed populations (median difference between R^* and the peak value of R: 4.6% for well-mixed and 2.7% for spatially structured populations).

Infectivity. Parasite evolution was closely linked to changes in the host population, which meant that host fitness costs and natural mortality had similar effects on I^* to those described above for R^* (fig. 5*A*). This is unsurprising given that parasites also experienced a fitness cost associated with generalism (η_P), so that broader infectivity ranges were unlikely to be selected for unless resistance was widespread. Hence, I^* decreased with η_H and μ in both environments, but parasites in spatially structured populations tended to exhibit broader ranges than those in well-mixed populations as η_H increased, provided the host population did not exhibit rapid turnover (figs. 5*B*, 6*B*; ANCOVA, environment × fitness cost interaction for $\mu = 0.001$: $F_{1,\,3,234} = 157$, $P < .0001$; see table A3 for full ANCOVA results).

As observed in the host population, average ranges among parasites were found to be stable over different measurement windows (median difference of 3.9% between time steps 5,001–7,500 and 7,501–10,000) and minor oscillations about the mean values of I^* were common (median difference between I^* and the peak value of I within each simulation: 5% for well-mixed and 4.7% for spatially structured populations).

Probabilistic Cellular Automata. To test the generality of our results, we created a simplified version of our model based on probabilistic cellular automata (see app. B for full description and results). We found that both the epidemiological dynamics and coevolutionary outcomes were qualitatively similar to those observed in the primary model, especially when moving between environments. Most importantly, our key finding was replicated using this alternative approach: spatially structured populations tended to produce broader resistance and infectivity ranges than well-mixed populations as fitness costs increased.

Discussion

Using an individual-based model of host-parasite coevolution, we explored the effects of spatial structure on the

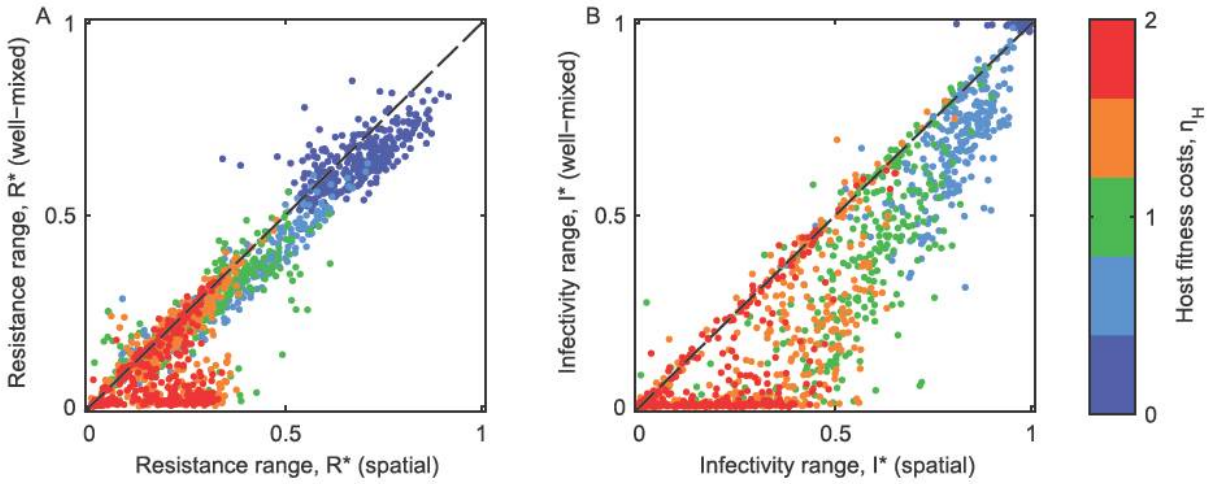

Figure 6: Scatter plots showing the resistance (*A*) and infectivity (*B*) ranges in the two environments for each simulation. Colors correspond to the host fitness cost parameter, η_H. Spatially structured populations consistently exhibit broader ranges than well-mixed populations, particularly when fitness costs are high. Parameters: $\mu = 0.001$; $n = 3$.

expansion of resistance and infectivity ranges. Consistent with previous theoretical (Hochberg and van Baalen 1998; Sasaki 2000; Agrawal and Lively 2002) and empirical studies (Bohannan et al. 2000; Forde et al. 2004, 2008; Buckling et al. 2006; Lopez-Pascua et al. 2008, 2012), we found resistance and infectivity range evolution to be constrained by fitness costs. Our novel result is that the magnitude of the constraint imposed by host fitness costs on range expansion is highly dependent on the degree of mixing in the population. More specifically, we found that high host fitness costs severely restricted range expansion in well-mixed environments, but the effect was much weaker in the presence of spatial structure. While host fitness costs did not have a direct impact on parasite fitness, we assumed that range expansion was still costly for parasites and so would only be selected for in the presence of resistant hosts (see Sasaki 2000). Thus, parasite infectivity behaved in a qualitatively similar way to resistance in response to greater host fitness costs.

To test the generality of our results we constructed a second model, replacing the free-living parasite population with one transmitted via direct contact between hosts (probabilistic cellular automata; app. B). We found the results to be broadly consistent, despite the stark contrast in transmission dynamics between the two models, which suggests that our findings are likely to have general applicability to a variety of host-parasite systems. Our study is closely linked to the work of Best et al. (2011), who reported an important effect of spatial structure on the coevolution of quantitative resistance and infectivity (as opposed to the resistance and infectivity ranges explored here): high resistance and low virulence were favored when dispersal was limited, but this trend was reversed in well-mixed environments.

The contrasting dynamics observed in the two environments can be explained by two processes. First, individuals in poorly mixed populations only compete with a subset of the population at very small spatial scales, whereas individuals in well-mixed environments are in competition with the entire population. In structured environments, the fitness of a particular genotype will vary depending on the neighboring population, which means that different genotypes may be optimal at different locations and spatial scales (Thompson 1994; Gandon et al. 1996). Second, clustering and limited dispersal can limit the extent to which a genotype can spread in spatially structured populations, even if it is optimal for a given location (Campos et al. 2008).

As a specific example of the impact of these two related processes, first consider a well-mixed population initially composed of sensitive hosts and specialist (narrow range) parasites in a single locus GFG framework. If a resistant host emerges, it is likely to have a huge advantage over the sensitive population and will therefore rapidly increase in prevalence. Specialist parasites will then be at a disadvantage, so generalists that can infect both host types will eventually dominate. Assuming resistance is associated with slower growth, but provides no protection against the generalist parasite, the sensitive population will then gradually increase relative to the resistant population. Similarly, generalist parasites have a lower fitness than specialist parasites when the host population is not resistant. Hence, the specialist parasite increases in prevalence and the cycle repeats, giving fluctuations in range. In a spatially struc-

tured environment, the emergence of resistant hosts and generalist parasites will initially follow a similar pattern. Although sensitive hosts will then have the globally optimal genotype, they may not be able to realize their growth advantage as clustering may limit the extent to which they can spread before being wiped out by parasites. In addition, if small numbers of specialist parasites are maintained by these sensitive patches then resistance could still be the locally optimal trait. Thus, local competition and clustering provide ephemeral refuges for globally suboptimal genotypes, which make spatially structured populations less likely to exhibit fluctuations in range. Similar dynamics emerge in our multilocus framework, as shown in figure 1. Resistance initially spreads in both environments, but they respond differently to the emergence of generalist parasites, with broader ranges persisting in the presence of spatial structure.

We found that the difference in coevolutionary outcomes between the two environments was dependent on the probability of natural death for the host, given by the parameter μ. Higher values of μ correspond to faster turnover rates in the host population, which increases the severity of fitness costs to the point where resistance is no longer beneficial in either environment. In addition, a faster population turnover rate will reduce the effects of clustering, allowing sensitive hosts with faster growth rates to reestablish themselves in the presence of generalist parasites.

The genetic specificity in our model was based on a well-established multilocus GFG framework that can produce arms race coevolutionary dynamics as well as fluctuations in range (Sasaki 2000; Fenton et al. 2009). While there is considerable evidence for coevolutionary arms races taking place among bacteria and phages (Bohannan and Lenski 2000; Buckling and Rainey 2002*a*; Mizoguchi et al. 2003; Brockhurst et al. 2006; Forde et al. 2008; Scanlan et al. 2011) and various other host-parasite systems (Little et al. 2006; Schulte et al. 2010), there is limited evidence of fluctuations in range except for some plant-fungus interactions (e.g., Thrall and Burdon 2003). Recent work with *Pseudomonas fluorescens* and lytic phages has shown that fluctuating selection between genotypes with similar ranges is possible, either following (Hall et al. 2011) or in the absence of a coevolutionary arms race (Gomez and Buckling 2011). In addition, the frequent occurrence of local adaptation indicates that there may be multiple routes to generalism (Buckling and Rainey 2002*b*; Morgan et al. 2005; Vos et al. 2009; Koskella et al. 2011). These data indicate that the GFG framework may only be capturing part of the genetic interactions between bacteria and phages, which has led others to propose more complex specificities (Agrawal and Lively 2002, 2003; Weitz et al. 2005; Forde et al. 2008; Fenton et al. 2012).

Furthermore, some systems appear to be based on other forms of specificity that do not permit generalism. For example, Carius et al. (2001) observed that the bacterium *Pasteuria ramosa* specializes on different lineages of the freshwater crustacean *Daphnia magna*, which can lead to fluctuating selection between different genotypes rather than an escalatory arms race (Decaestecker et al. 2007). Similarly, coevolutionary dynamics between the freshwater snail *Potamopyrgus antipodarum* and trematode parasites of the genus *Microphallus* appear to be governed by fluctuating selection between specialists, a process which has been linked to the maintenance of sexual reproduction among the host population (Lively 1987; King et al. 2009). While generalists have not yet been observed in these systems, it is possible that the fitness costs associated with broad ranges are simply too high.

Our work complements the growing body of research on the effects of spatial structure on coevolutionary dynamics (Hartvigsen and Levin 1997; Boots and Sasaki 1999; Haraguchi and Sasaki 2000; Read and Keeling 2003; Jansen and van Baalen 2006; Kerr et al. 2006; Mitchell et al. 2006; Heilmann et al. 2010, 2012; Best et al. 2011; Haerter et al. 2011; Zaman et al. 2011). There are also strong links between this study and a variety of ecological models on victim-exploiter relationships. Of particular relevance is the work on host and parasitoids, in which variation in dispersal rate can lead to a range of complex dynamics, with high rates of dispersal increasing extinction risk and leading to fluctuations in population sizes, as observed here (Hassell et al. 1991; Comins et al. 1992; Pascual 1993). Together, these studies highlight the important role that spatial structure plays in shaping both ecological and evolutionary dynamics.

Acknowledgments

We thank M. Boots, A. Gardner, B. Koskella, B. Penman, S. West, H. Williams, and two anonymous reviewers for extremely helpful comments on the manuscript. B.A. is funded by a Biotechnology and Biological Sciences Research Council Studentship. A.B. and S.G. gratefully acknowledge support from the European Research Council.

Literature Cited

Agrawal, A. F., and C. M. Lively. 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evolutionary Ecology Research 4:79–90.

———. 2003. Modelling infection as a two-step process combining gene-for-gene and matching-allele genetics. Proceedings of the Royal Society B: Biological Sciences 270:323–334.

Bak, P., and C. Chao. 1990. A forest-fire model and some thoughts on turbulence. Physics Letters A 147:297–300.

Best, A., S. Webb, A. White, and M. Boots. 2011. Host resistance and coevolution in spatially structured populations. Proceedings of the Royal Society B: Biological Sciences 278:2216–2222.

Bohannan, B. J. M., B. Kerr, C. M. Jessup, J. B. Hughes, and G. Sandvik. 2002. Trade-offs and coexistence in microbial microcosms. Antonie van Leeuwenhoek 81:107–115.

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

Boots, M., and A. Sasaki. 1999. "Small worlds" and the evolution of virulence: infection occurs locally and at a distance. Proceedings of the Royal Society B: Biological Sciences 266:1933–1938.

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

Brockhurst, M. A., A. D. Morgan, P. B. Rainey, and A. Buckling. 2003. Population mixing accelerates coevolution. Ecology Letters 6:975–979.

Buckling, A., and P. B. Rainey. 2002*a*. Antagonistic coevolution between a bacterium and a bacteriophage. Proceedings of the Royal Society B: Biological Sciences 269:931–936.

———. 2002*b*. The role of parasites in sympatric and allopatric host diversification. Nature 420:496–499.

Buckling, A., Y. Wei, R. C. Massey, M. A. Brockhurst, and M. E. Hochberg. 2006. Antagonistic coevolution with parasites increases the cost of host deleterious mutations. Proceedings of the Royal Society B: Biological Sciences 273:45–49.

Burdon, J. J., and P. H. Thrall. 1999. Spatial and temporal patterns in coevolving plant and pathogen associations. American Naturalist 153(suppl.):S15–S33.

Campos, P. R. A., P. S. C. A. Neto, V. M. de Oliveira, and I. Gordo. 2008. Environmental heterogeneity enhances clonal interference. Evolution 62:1390–1399.

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

Chao, L., B. R. Levin, and F. M. Stewart. 1977. A complex community in a simple habitat: an experimental study with bacteria and phage. Ecology 58:369–378.

Comins, H. N., M. P. Hassell, and R. M. May. 1992. The spatial dynamics of host–parasitoid systems. Journal of Animal Ecology 61:735–748.

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.

Decaestecker, E., S. Gaba, J. A. M. Raeymaekers, R. Stoks, L. Van Kerckhoven, D. Ebert, and L. De Meester. 2007. Host-parasite "Red Queen" dynamics archived in pond sediment. Nature 450:870–873.

Eames, K. T. D., and M. J. Keeling. 2002. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. Proceedings of the National Academy of Sciences of the USA 99:13330–13335.

Faruque, S. M., K. Biswas, S. M. N. Udden, Q. S. Ahmad, D. A. Sack, G. B. Nair, and J. J. Mekalanos. 2006. Transmissibility of cholera: in vivo-formed biofilms and their relationship to infectivity and persistence in the environment. Proceedings of the National Academy of Sciences of the USA 103:6350–6355.

Fenton, A., J. Antonovics, and M. A. Brockhurst. 2009. Inverse-gene-for-gene infection genetics and coevolutionary dynamics. American Naturalist 174:E230–E242.

———. 2012. Two-step infection processes can lead to coevolution between functionally independent infection and resistance pathways. Evolution 66:2030–2041.

Flor, H. H. 1956. The complementary genetic systems in flax and flax rust. Advances in Genetics 8:29–54.

Forde, S. E., R. E. Beardmore, I. Gudelj, S. S. Arkin, J. N. Thompson, and L. D. Hurst. 2008. Understanding the limits to generalizability of experimental evolutionary models. Nature 455:220–223.

Forde, S. E., J. N. Thompson, and B. J. M. Bohannan. 2004. Adaptation varies through space and time in a coevolving host-parasitoid interaction. Nature 431:841–844.

———. 2007. Gene flow reverses an adaptive cline in a coevolving host-parasitoid interaction. American Naturalist 169:794–801.

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

Gallet, R., Y. Shao, and I.-N. Wang. 2009. High adsorption rate is detrimental to bacteriophage fitness in a biofilm-like environment. BMC Evolutionary Biology 9:241.

Gandon, S., A. Buckling, E. Decaestecker, and T. Day. 2008. Host-parasite coevolution and patterns of adaptation across time and space. Journal of Evolutionary Biology 21:1861–1866.

Gandon, S., Y. Capowiez, Y. Dubois, and Y. Michalakis. 1996. Local adaptation and gene-for-gene coevolution in a metapopulation model. Proceedings of the Royal Society B: Biological Sciences 263: 1003–1009.

Gomez, P., and A. Buckling. 2011. Bacteria-phage antagonistic coevolution in soil. Science 332:106–109.

Haerter, J. O., A. Trusina, and K. Sneppen. 2011. Targeted bacterial immunity buffers phage diversity. Journal of Virology 85:10554–10560.

Hall, A. R., P. D. Scanlan, A. D. Morgan, and A. Buckling. 2011. Host-parasite coevolutionary arms races give way to fluctuating selection. Ecology Letters 14:635–642.

Haraguchi, Y., and A. Sasaki. 2000. The evolution of parasite virulence and transmission rate in a spatially structured population. Journal of Theoretical Biology 203:85–96.

Hartvigsen, G., and S. Levin. 1997. Evolution and spatial structure interact to influence plant-herbivore population and community dynamics. Proceedings of the Royal Society B: Biological Sciences 264:1677–1685.

Hassell, M. P., R. M. May, S. W. Pacala, and P. L. Chesson. 1991. The persistence of host-parasitoid associations in patchy environments. I. A general criterion. American Naturalist 138:568–583.

Heilmann, S., K. Sneppen, and S. Krishna. 2010. Sustainability of virulence in a phage-bacterial ecosystem. Journal of Virology 84: 3016–3022.

———. 2012. Coexistence of phage and bacteria on the boundary of self-organized refuges. Proceedings of the National Academy of Sciences of the USA 109:12828–12833.

Hellweger, F. L., and V. Bucci. 2009. A bunch of tiny individuals: individual-based modeling for microbes. Ecological Modelling 220: 8–22.

Hochberg, M. E., and M. van Baalen. 1998. Antagonistic coevolution over productivity gradients. American Naturalist 152:620–634.

Jansen, V. A. A., and M. van Baalen. 2006. Altruism through beard chromodynamics. Nature 440:663–666.

Keeling, M. J., M. E. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-

Topping, D. T. Haydon, S. J. Cornell, et al. 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science 294:813–817.

Kerr, B., C. Neuhauser, B. J. M. Bohannan, and A. M. Dean. 2006. Local migration promotes competitive restraint in a host-pathogen "tragedy of the commons." Nature 442:75–78.

King, K. C., L. F. Delph, J. Jokela, and C. M. Lively. 2009. The geographic mosaic of sex and the Red Queen. Current Biology 19: 1438–1441.

Koskella, B., J. N. Thompson, G. M. Preston, and A. Buckling. 2011. Local biotic environment shapes the spatial scale of bacteriophage adaptation to bacteria. American Naturalist 177:440–451.

Levin, B. R., and J. J. Bull. 2004. Population and evolutionary dynamics of phage therapy. Nature Reviews Microbiology 2:166–173.

Little, T. J., K. Watt, and D. Ebert. 2006. Parasite-host specificity: experimental studies on the basis of parasite adaptation. Evolution 60:31–38.

Lively, C. M. 1987. Evidence from a New Zealand snail for the maintenance of sex by parasitism. Nature 328:519–521.

Lopez-Pascua, L. D. C., and A. Buckling. 2008. Increasing productivity accelerates host-parasite coevolution. Journal of Evolutionary Biology 21:853–860.

Lopez-Pascua, L. D. C., S. Gandon, and A. Buckling. 2012. Abiotic heterogeneity drives parasite local adaptation in coevolving bacteria and phages. Journal of Evolutionary Biology 25:187–195.

Matz, C., D. McDougald, A. M. Moreno, P. Y. Yung, F. H. Yildiz, and S. Kjelleberg. 2005. Biofilm formation and phenotypic variation enhance predation-driven persistence of *Vibrio cholerae*. Proceedings of the National Academy of Sciences of the USA 102: 16819–16824.

Mitchell, M., M. D. Thomure, and N. L. Williams. 2006. The role of space in the success of coevolutionary learning. Pages 118–124 *in* Proceedings, Artificial Life X: Tenth International Conference on the Simulation and Synthesis of Living Systems. MIT Press, Cambridge, MA.

Mizoguchi, K., M. Morita, C. R. Fischer, M. Yoichi, Y. Tanji, and H. Unno. 2003. Coevolution of bacteriophage PP01 and *Escherichia coli* O157:H7 in continuous culture. Applied and Environmental Microbiology 69:170–176.

Morgan, A. D., M. A. Brockhurst, L. D. C. Lopez-Pascua, C. Pal, and A. Buckling. 2007. Differential impact of simultaneous migration on coevolving hosts and parasites. BMC Evolutionary Biology 7: 1.

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

Pascual, M. 1993. Diffusion-induced chaos in a spatial predator-prey system. Proceedings of the Royal Society B: Biological Sciences 251:1–7.

Poullain, V., S. Gandon, M. A. Brockhurst, and A. Buckling. 2008. The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage. Evolution 62:1–11.

Rand, D. A., M. Keeling, and H. B. Wilson. 1995. Invasion, stability and evolution to criticality in spatially extended, artificial host-pathogen ecologies. Proceedings of the Royal Society B: Biological Sciences 259:55–63.

Read, J. M., and M. J. Keeling. 2003. Disease evolution on networks: the role of contact structure. Proceedings of the Royal Society B: Biological Sciences 270:699–708.

Rhodes, C. J., and R. M. Anderson. 1996. Persistence and dynamics in lattice models of epidemic spread. Journal of Theoretical Biology 180:125–133.

Sasaki, A. 2000. Host-parasite coevolution in a multilocus gene-for-gene system. Proceedings of the Royal Society B: Biological Sciences 267:2183–2188.

Scanlan, P. D., A. R. Hall, L. D. C. Lopez-Pascua, and A. Buckling. 2011. Genetic basis of infectivity evolution in a bacteriophage. Molecular Ecology 25:1–9.

Schulte, R. D., C. Makus, B. Hasert, N. K. Michiels, and H. Schulenburg. 2010. Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. Proceedings of the National Academy of Sciences of the USA 107:7359–7364.

Shannon, C. E. 1948. A mathematical theory of communication. Bell System Technical Journal 27:379–423.

Thompson, J. N. 1994. The coevolutionary process. University of Chicago Press, Chicago.

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

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

———. 2003. Evolution of virulence in a plant host-pathogen metapopulation. Science 299:1735–1737.

Vos, M., P. J. Birkett, E. Birch, R. I. Griffiths, and A. Buckling. 2009. Local adaptation of bacteriophages to their bacterial hosts in soil. Science 325:833.

Webster, J. P., and M. E. J. Woolhouse. 1999. Cost of resistance: relationship between reduced fertility and increased resistance in a snail-schistosome host-parasite system. Proceedings of the Royal Society B: Biological Sciences 266:391–396.

Weitz, J. S., H. Hartman, and S. A. Levin. 2005. Coevolutionary arms races between bacteria and bacteriophage. Proceedings of the National Academy of Sciences of the USA 102:9535–9540.

Zaman, L., S. Devangam, and C. Ofria. 2011. Rapid host-parasite coevolution drives the production and maintenance of diversity in digital organisms. Pages 219–226 *in* Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation. ACM, New York.

Associate Editor: Pej Rohani
Editor: Judith L. Bronstein

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

rstb.royalsocietypublishing.org

Review

Cite this article: Ashby B, Gupta S. 2013 Sexually transmitted infections in polygamous mating systems. Phil Trans R Soc B 368: 20120048.
http://dx.doi.org/10.1098/rstb.2012.0048

One contribution of 14 to a Theme Issue 'The polyandry revolution'.

Subject Areas:
theoretical biology

Keywords:
sexually transmitted infections, host heterogeneity, sexual contact network, polygamy, evolution of virulence, mating strategy

Author for correspondence:
Ben Ashby
e-mail: ben.ashby@zoo.ox.ac.uk

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2012.0048 or via http://rstb.royalsocietypublishing.org.

Royal Society Publishing
Informing the science of the future

Sexually transmitted infections in polygamous mating systems

Ben Ashby and Sunetra Gupta

Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

Sexually transmitted infections (STIs) are often associated with chronic diseases and can have severe impacts on host reproductive success. For airborne or socially transmitted pathogens, patterns of contact by which the infection spreads tend to be dispersed and each contact may be of very short duration. By contrast, the transmission pathways for STIs are usually characterized by repeated contacts with a small subset of the population. Here we review how heterogeneity in sexual contact patterns can influence epidemiological dynamics, and present a simple model of polygyny/polyandry to illustrate the impact of biased mating systems on disease incidence and pathogen virulence.

1. Introduction

Evidence from anthropological and ethological studies suggests that there is much heterogeneity in sexual behaviour of humans and animals [1–4], both in rates of sexual activity and in patterns of sexual contact. Polygynous and polyandrous mating systems are particular examples, where one sex tends to have a much higher variance in partner acquisition rate compared with the other sex. It is well-established that the structure of a mating system can have a profound influence on genetic diversity [5] and the evolution of sexually selected traits [6]; here we discuss how such heterogeneities can influence epidemiological dynamics of sexually transmitted infections (STIs) and the evolution of associated pathogens.

We begin by reviewing how host heterogeneity in sexual behaviour can influence the epidemiological dynamics of STIs, mainly in the context of human diseases, such as HIV-1 and gonorrhoea. We then introduce an individual-based model for biased (polygynous and polyandrous) mating systems, where movement is based on perception of reproductive failure. In line with the results of Thrall *et al.* [7], we observe that the polygamous sex exhibits much lower levels of infection than the monogamous sex and the difference tends to increase with greater variance in attractiveness; in addition, we show that the difference between the polygynous and polyandrous scenarios depends on the probability of sterility. Finally, we present an example of how the evolution of pathogen virulence can be explored within this framework by introducing a simple dichotomy in the trade-off between transmissibility and duration of infection. Within our system, the less virulent pathogen tends to be favoured for high degrees of polygamy, demonstrating a clear link between mating patterns and pathogen evolution.

2. Influence of sexual contact patterns on epidemiology of sexually transmitted infections

From an epidemiological point of view, the transmission dynamics of STIs are fundamentally different to those of many other infectious diseases. First, sexual contact rates are usually invariant to population size, which means that there is no critical population density required for a typical STI to persist. By contrast, the rate at which non-STIs spread is often dependent on the density of the host population [8]. Second, there is often considerable variation in sexual behaviour both within [4] and between [9] populations. Highly active members of the

population (e.g. sex workers in human populations, alpha males in animal populations) will generally be at much greater risk of receiving and transmitting an infection than monogamous couples and so contribute disproportionately to the spread of disease as well as representing important targets for disease control.

In order to establish how heterogeneity in sexual behaviour can alter epidemiological dynamics, it is first useful to define the basic reproductive number, R_0, of an infectious agent in a well-mixed population:

$$R_0 = \beta D n, \tag{2.1}$$

where β is the probability of transmission per contact, D the average infectious period and n the average number of contacts (see [8] for a more detailed discussion of R_0). The basic reproductive number is essentially the average number of secondary infections that a single infectious individual will produce in an entirely susceptible population. Hence the infection will tend to spread if $R_0 > 1$, but will go extinct if $R_0 < 1$. This formulation of R_0 is based on an idealized, randomly mixing homogeneous population, but real mixing patterns are likely to be more complex due to spatial constraints and variations in host behaviour. If we imagine a continuum with well-mixed and highly structured populations at the extremes, then most real populations will fall somewhere between the two. Note that the position of a population on this continuum is dependent on the transmission pathways of a particular infection; a population may be relatively well-mixed in terms of social contacts, but might demonstrate a high degree of heterogeneity in sexual mixing patterns.

In general, casual contacts between humans tend to be ephemeral and non-repetitive, whereas sexual contacts are more stable [10]. In addition, variation in close contact rates is likely to be much smaller than variation in sexual partner acquisition rates. For example, Mossong *et al.* [11] found that adolescents had just over twice the number of close contacts than the elderly, but studies of sexual mixing patterns generally find power-law distributions in partner acquisition rates, sometimes ranging over three orders of magnitude [4,12]. Power-law distributions are also likely to be applicable to a variety of animal mating systems, particularly where a few members of one sex are dominant (i.e. polyandry or polygyny). Hence, the above formulation of R_0 may be a reasonably good indicator of epidemic spread for infections transmitted by close contact, but is likely to be a poor approximation for STIs. In addition, sexual contacts tend to be much less frequent than social contacts, lowering the value of R_0. Hence, one explanation as to why many STIs are associated with chronic, asymptomatic diseases is that this increases the value of D to compensate for lower contact rates.

Sexual transmission can be considered part of a much broader class of models with heterogeneous contact rates, usually referred to as 'super-spreader' models, where a few members of the population have a disproportionately large effect on disease spread [8,12]. For super-spreader models, we can incorporate this heterogeneity into the formula for R_0 by compartmentalizing the population according to contact rates, so that N_i is the proportion of the population that acquires i contacts per unit time [13,14]. Retaining the assumption that mixing is random,

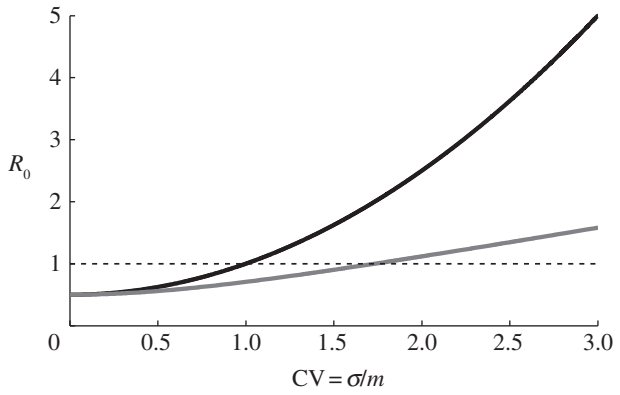

Figure 1. The relationship between the coefficient of variation, $CV = \sigma/m$, of the partner acquisition rate distribution and the basic reproductive number, R_0, for fixed values of β, D and m. ($\beta = 0.1$, $D = 5$ year and $m = 1\ yr^{-1}$, so that $\beta Dm = 0.5$). The black curve shows the relationship when there is no distinction between the sexes ($R_0 = \beta Dm(1 + CV^2)$) and the grey curve shows the relationship when variation only occurs in one sex ($R_0 = \beta Dm(1 + CV^2)^{1/2}$). The dotted line corresponds to the threshold for an epidemic ($R_0 > 1$). Adapted from May *et al.* [15].

this allows us to calculate an effective contact rate, c, over the distribution:

$$c = \frac{\sum_i i^2 N_i}{\sum_i i N_i} = m + \frac{\sigma^2}{m}, \tag{2.2}$$

where m and σ^2 are the mean and variance in contact rates, respectively [14]. The formula for the basic reproductive number now becomes $R_0 = \beta D c$; clearly, any heterogeneity in contact rates will increase the value of R_0 and hence the initial growth rate of the epidemic (figure 1).

In the context of sexual transmission, the second formulation of R_0 applies to a homosexual population, but it can be readily generalized for a heterosexual population by separating the effective partner acquisition (i.e. contact) rate into male (c_m) and female (c_f) components. If we also assume that there are differential transmission rates across the sexes (as is common with many STIs), then our equation for R_0 becomes:

$$R_0 = D\sqrt{\beta_m \beta_f c_m c_f}, \tag{2.3}$$

where β_m is the transmission rate from males to females and β_f is the transmission rate from females to males [8,16]. If $c_m = c_f$ then R_0 will asymptote towards quadratic growth with the coefficient of variation ($CV = \sigma/m$), but if variation is limited to one sex then R_0 will tend towards linear growth. Changes in the variance will be most significant when R_0 is close to unity (the epidemic threshold), as relatively small changes in the size or the behaviour of the core group can determine whether an epidemic will occur (figure 1).

The effective partner acquisition rates can also be used to estimate the ratio of cases in males (C_m) to females (C_f) during the early stages of the epidemic:

$$\frac{C_m}{C_f} = \sqrt{\frac{\beta_f c_f}{\beta_m c_m}}, \tag{2.4}$$

(see [8], §11.3.9 for a more detailed discussion; also [15]). This work was originally motivated by the spread of HIV in Africa, but the principles can be applied to other populations

that exhibit host heterogeneity. For example, the ratio C_m/C_f suggests that the dominant sex in a biased mating system (e.g. males in polygynous systems) will tend to exhibit lower than average levels of infection, although this could be counterbalanced by differences in transmission probabilities. Indeed, it is thought that $C_m/C_f \approx 1$ for HIV-1 in many parts of Africa because the partner acquisition rates ($c_f > c_m$) are more or less balanced by differences in transmission rates ($\beta_f < \beta_m$) [8,16]. Note that even if $C_m/C_f \approx 1$, the distribution of infection will still be biased towards more sexually active members of the population.

Further complications will arise if the population does not mix homogeneously, for example where people tend to show a preference for mixing with similar individuals (assortativity). Mixing patterns have been found to vary considerably between human populations, ranging from highly assortative [17], to highly disassortative (i.e. showing preference for dissimilar individuals) [18] mixing. The degree of assortative mixing may also vary within a population: for example, Wylie & Jolie [19] found that assortative mixing was common in linear components of a sexual contact network (SCN), but disassortative mixing was common in radial components.

In order to model heterogeneous mixing, we can group individuals according to their level of sexual activity (i.e. partner acquisition rate) and describe interactions between groups using a 'mixing-matrix' [13,14,16,20,21]. A simple mixing-matrix for a population split into high (H) and low (L) activity groups would be

$$\boldsymbol{p}_{ij} = \begin{pmatrix} p_{LL} & p_{LH} \\ p_{HL} & p_{HH} \end{pmatrix}, \tag{2.5}$$

where p_{ij} is the proportion of sexual contacts that individuals from group i make with members of group j. For completely assortative mixing, $\boldsymbol{p}_{ij}$ is equal to the identity matrix ($p_{ii} = 1$, $p_{ij} = 0$ for $i \neq j$). There is usually no single disassortative extreme, however, as disassortativity is maximized whenever the elements of the main diagonal of $\boldsymbol{p}_{ij}$ are minimized [21]. For a given mixing matrix, we can measure the degree of assortativity, Q, in the population as

$$Q = \frac{1}{g-1}\left(\sum_{i=1}^{g} \lambda_i - 1\right), \tag{2.6}$$

where g is the number of activity groups and λ_i are the eigenvalues of $\boldsymbol{p}_{ij}$. Gupta *et al.* [21] found that highly assortative mixing ($Q \approx 1$) tends to lead to more rapid epidemic growth and can produce multiple peaks in disease incidence. By contrast, highly disassortative mixing ($Q \approx -1/(g-1)$) is generally associated with slower epidemic growth, but will typically produce higher peaks in disease incidence (figure 2). This method highlights the importance of host heterogeneity in the spread of STIs and suggests that targeting control measures at the core group is optimal, although the efficacy of such procedures will depend on the size of this group and the degree of assortative mixing in the population.

While this approach is a useful way of capturing host heterogeneity, it cannot capture some of the complex interactions found in real populations that are imposed by other factors than level of sexual activity. Such mean-field approaches assume that sexual activity classes are well mixed so that if an infectious individual mixes with a particular activity class, then all members of that class will have an equally increased risk of infection. In reality, the risk of infection will be limited to those who have sexual contact with the infectious individual rather than the entire activity class. An alternative method is to use an SCN which captures heterogeneity at the level of individuals and provides a means of replicating more realistic transmission pathways. This approach is particularly well suited to STIs, as transmission pathways are usually much more clearly defined (i.e. sexual contact) than for non-STIs. However, there are many problems associated with collecting data on real SCNs, including biases in reporting and difficulties with linking up components in a larger network [17,22], although some attempts have been made for small populations [18,19,23–26].

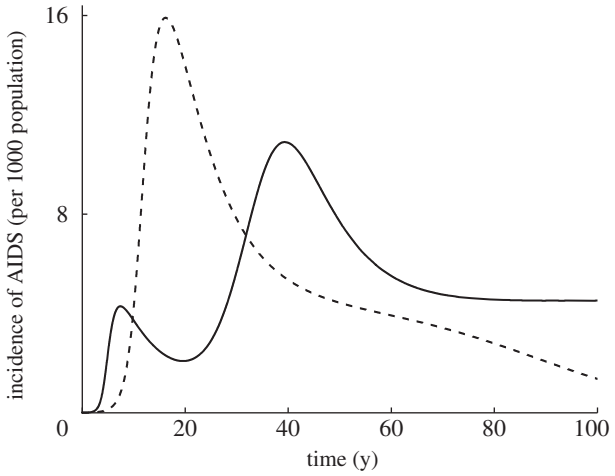

Figure 2. Incidence of AIDS as a proportion of population size in populations that exhibit highly assortative (solid curve) and highly disassortative (dotted curve) mixing. Highly assortative (solid line) mixing tends to lead to rapid growth during the early stages of an epidemic and can produce multiple peaks in disease incidence. Highly disassortative (dashed line) mixing is usually characterized by slower initial growth, but a higher peak in the incidence of AIDS. The model is adapted from Gupta *et al.* [21] (see the electronic supplementary material).

Mean-field models and SCNs may show good agreement over the main part of an epidemic, but fundamental differences in structure are likely to have significant consequences for the spread of infection during the early stages [27]. Keeling [27] showed that in a general SCN, the probability that an index case i will fail to pass on the infection to any of their n_i contacts is

$$P_{SCN}(i) = \left(1 - \frac{R_0(i)}{n_i}\right)^{n_i}, \tag{2.7}$$

where $R_0(i) = \min(\beta D n_i, n)$ is the expected number of secondary infections to be caused by the index case. The corresponding probability of extinction after the first generation using a mean-field approximation is:

$$P_{MF}(i) = \exp(-R_0(i)), \tag{2.8}$$

which satisfies $P_{MF}(i) > P_{SCN}(i)$. Hence the probability of extinction during the first generation is always higher in a mean-field approximation than it is in the corresponding SCN, but as the number of contacts increases the two values converge (i.e. $\lim_{n_i \to \infty} (P_{SCN}(i)) = P_{MF}(i)$). Averaging these values over the entire population gives the probability that a randomly introduced infection will die out after the first generation. Note that in real populations, an index case is probably more likely

to occur among highly active individuals, which will decrease the probability of extinction during the first generation.

One advantage that SCNs have over mean-field approximations is their ability to capture long-term partnerships that are commonly found in many human and animal populations. In particular, serially monogamous partnerships (common among birds as well as humans) cannot be modelled using traditional mean-field approaches. Computer-generated contact networks can be used to recreate mixing patterns observed in real populations [25], by connecting individuals (nodes) to other members of the population preferentially, based on factors such as proximity, cluster size or assortativity. Studies of simulated epidemics on SCNs have revealed that concurrent partnerships are crucially important to the spread of many STIs [28–30]. For example, Morris & Kretzschmar [29] demonstrated that the size of an epidemic grows exponentially with the relative number of concurrent partnerships in a population. Reducing the number of concurrent partnerships in a population is therefore likely to be an effective mechanism of disease control.

3. Exploring the role of mating system structure on epidemiological dynamics

We now introduce a simple SCN model to illustrate how epidemiological dynamics of STIs can be influenced by biased (i.e. polygynous or polyandrous) mating systems. Our model is similar to that of Thrall *et al.* [7], which was used to explore how disease prevalence is affected in a general biased mating system with random movement between mating groups. The authors found that disease prevalence in the two sexes tends to diverge as variance in mating success increases and that less attractive members of the population may have higher lifetime reproductive success in the presence of a sterilizing STI. We build on this study by varying the probability that an infection will cause host sterility and by basing movement decisions between mating groups on an individual's perception of reproductive failure. This simple stay-or-stray decision introduces behavioural differences between polygynous and polyandrous mating systems, as females are generally better placed to infer their reproductive success. Given that a range of complex mate choice behaviour has been observed, including the avoidance of parasitism, inbreeding and harassment (see [31] for a review of mate choice behaviour), it seems reasonable that a simple binary decision of prior reproductive success or failure could influence mate choice. In fact, a meta-analysis of mate fidelity among 35 species of monogamous birds found that divorce rates were significantly higher among unsuccessful than successful pairs [32], providing strong evidence that prior reproductive failure can reduce mate fidelity.

We consider a population of constant size, composed of N_m males and N_f females, where one sex is polygamous and the other is serially monogamous. We follow Thrall *et al.* [7] in assigning members of the serially monogamous sex to mating groups consisting of a single member of the polygamous sex. Each polygamous individual, i, is assigned a fixed level of attractiveness, $a(i)$, according to a power-law distribution with shape parameter α

$$a(i) = \frac{i^{-\alpha}}{\sum_{k=1}^{N_m} k^{-\alpha}}, \tag{3.1}$$

with $\sum_{k=1}^{N_m} a(i) = 1$. Members of the polygamous sex are then assigned non-overlapping line segments, $L(i)$, with lengths equal to their attractiveness:

$$L(i) = \begin{cases} [0, a(1)] & \text{if } i = 1 \\ \left[\sum_{k=1}^{i-1} a(k), \sum_{k=1}^{i} a(k)\right] & \text{else.} \end{cases} \tag{3.2}$$

For each serially monogamous individual j, a random number, $r(j) \epsilon (0,1)$, is then generated. The connections between males and females are given by the adjacency matrix $\boldsymbol{A}_{ij}$, where $A_{ij} = 1$ if $r(j) \epsilon L(i)$ and is zero otherwise. For small values of α, there is little variation in attractiveness and so the network approaches serial monogamy for both sexes, although some concurrent partnerships may still occur (figure 3*a*). For large values of α, the network is dominated by a single polygamous individual who is connected to a large proportion of the serially monogamous population (figure 3*b*). We refer to these scenarios as having low and high degrees of polygamy, respectively.

A randomly chosen member of the monogamous sex initiates the epidemic, after which susceptible individuals are infected with probability βI, where β is the probability of transmission per contact and I is the total number of infectious contacts for a given individual. For members of the polygamous sex, the number of infectious contacts will be equal to the total number of infected individuals in their mating group, whereas for the monogamous sex $I = 1$ if their mate is infected and is zero otherwise. We also include an external force of infection, such that susceptible individuals are randomly infected with probability κ. Individuals recover from infection with probability σ, at which point they become susceptible to infection again. Infection causes no increase in mortality, but does carry a risk of permanent sterility (probability γ).

We assume that serially monogamous individuals stay within a mating group unless they are certain that they have not successfully produced offspring in the previous mating season. For the purposes of our model, we assume that mating is only successful provided both partners are not sterile. If they are certain that they have been unsuccessful, then they will reassess their mate choice with probability $1 - \rho$, where ρ is inertia to switching mating groups. In real populations, searching for a new mate may be risky (e.g. increased chance of predation, exposure to new pathogens or risk of exclusion; [31]) and so it is reasonable to assume that the serially monogamous sex will only leave a mating group if they are certain that they have not successfully reproduced. If an individual leaves a mating group, then they are immediately reassigned to a new mating group according to the original procedure described above. Deaths occur randomly with probability μ; we keep the population size constant by assuming that there is either always a surplus of offspring, or that the immigration rate is sufficiently high to maintain this balance.

The description of polyandrous and polygynous mating systems has been identical up until this point, but they can be distinguished by the behaviour of sterile members of the monogamous sex. In a polygynous system, each female is able to independently determine whether she has successfully reproduced or not. If she has been unsuccessful, then she may opt to choose an alternative mate in future. In a polyandrous system, however, success is based on the ability of

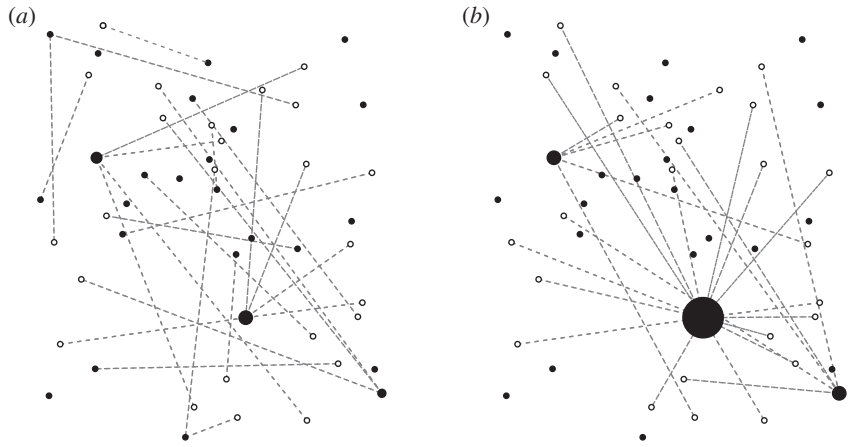

Figure 3. Example sexual contact networks (SCNs) in a biased mating system. Sexual contacts are indicated by dotted lines connecting the serially monogamous sex (empty circles) to the polygamous sex (filled circles). Members of the polygamous sex attract mates based on their relative 'attractiveness' (represented here by circle size), which is based on a power-law distribution (equation (3.1)) with (*a*) $\alpha = 1$ and (*b*) $\alpha = 2$. (*a*) Lower values of α produce more balanced SCNs, with less variation in the number of partners. (*b*) Higher values of α cause a few members of the population to dominate the network and leave many members of the polygamous sex without partners.

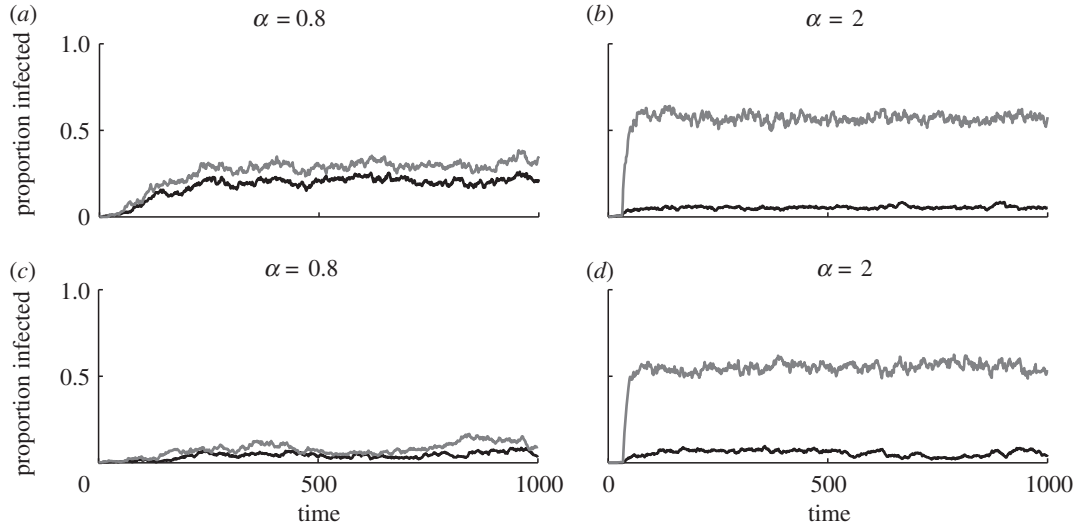

Figure 4. Dynamics of an SIS-type infection on SCNs for varying degrees of polygamy (α) and movement probabilities (ρ). (*a*,*b*) Are examples from a polygynous species, whereas (*c*,*d*) are examples from a polyandrous species. Higher values of α correspond to higher variation in the number of partners for the polygamous sex (equation (3.1)). Infection is usually more prevalent in the serially monogamous sex (grey) than in the polygamous sex (black) (parameters: $N_m = N_f = 500$; $\beta = 0.1$; $\gamma = 0.5$; $\kappa = 10^{-4}$; $\mu = 0.01$; $\rho = 0.5$; $\sigma = 0.05$).

the sole female in the mating group to reproduce. As long as there is a chance that a male has successfully reproduced (i.e. the female and at least one male in the group are both fertile), then the benefits of staying within a mating group may outweigh the costs of leaving, even if males are unable to detect who is the father of an infant.

Figure 4 shows example simulation dynamics for various degrees of polygamy (α) in polygynous (figure 4*a*,*b*) and polyandrous (figure 4*c*,*d*) mating systems. For relatively low values of α, the infection is able to spread to a reasonably large proportion of a polygynous population, but it is unable to spread extensively in a polyandrous population. For higher values of α, the infection is able to propagate through both polygynous and polyandrous populations, with little difference between the two scenarios. At this extreme, most of the monogamous population mates with a small set of individuals, leaving the rest of the polygamous sex disconnected from the network. This simultaneously increases the average exposure of the monogamous sex to infection, while decreasing the average exposure of the polygamous sex. As in Thrall *et al.* [7], we observe that the polygamous sex tends to exhibit much lower levels of infection than the monogamous sex and the difference tends to increase with greater variance in attractiveness. This is further emphasized in figure 5, where the average prevalence of infection in the monogamous sex generally increases with larger values of α, but peaks at intermediate values of α for the polygamous sex.

The number of cases in the monogamous sex is largely invariant to changes in the level of ρ (inertia); this is true for both polygynous and polyandrous scenarios (figure 5*b*,*d*). Similarly, the prevalence of infection in polyandrous females is only marginally influenced by ρ (figure 5*c*). In accordance with Thrall *et al.* [7], the average prevalence of

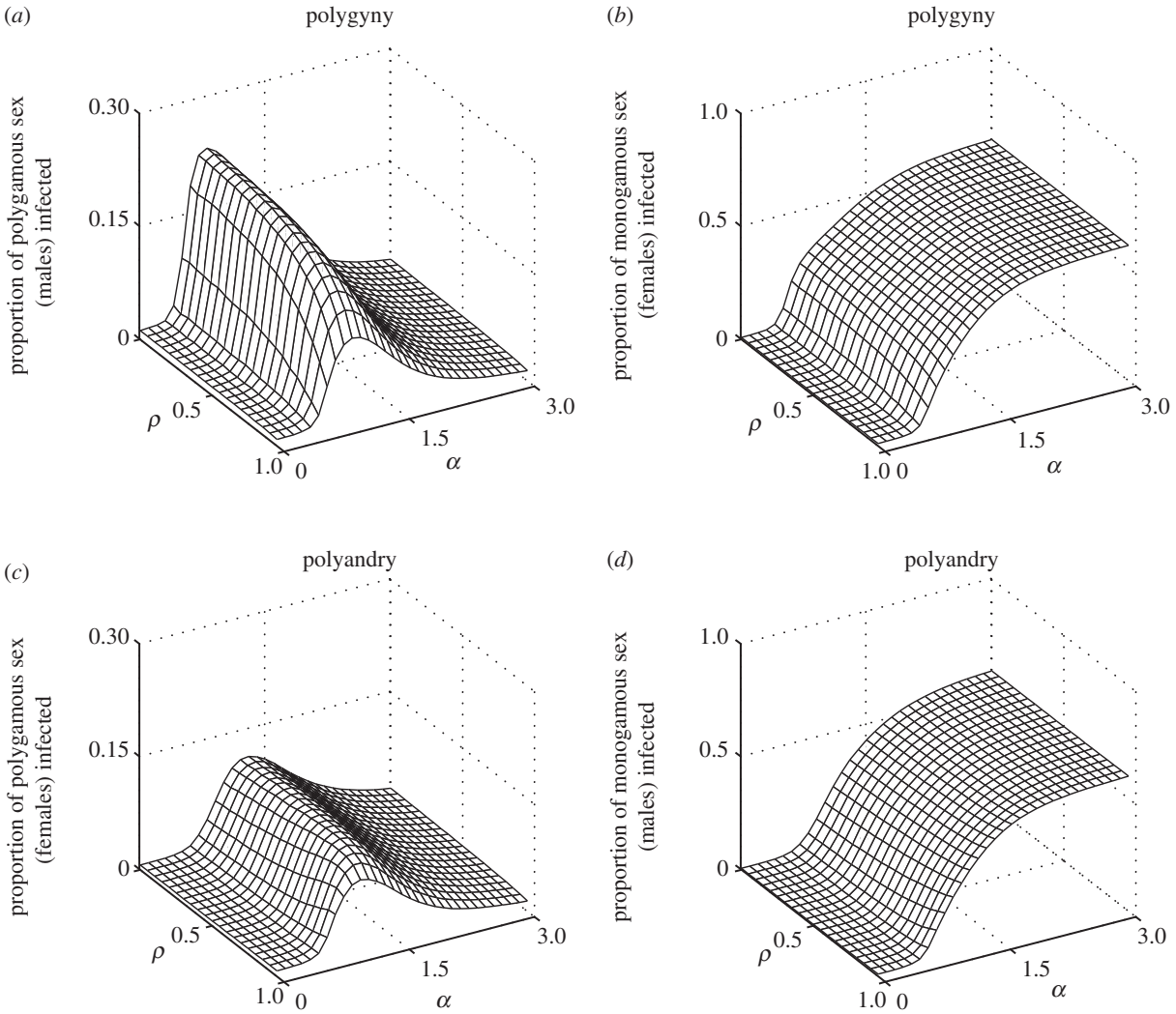

Figure 5. The average proportion of the (*a*,*c*) polygamous sex and (*b*,*d*) the monogamous sex that are infected for varying degrees of polygamy (α) and inertia to switching mating groups (ρ). (*a*,*b*) Data from a polygynous system; (*c*,*d*) data from a polyandrous system. Higher values of α correspond to higher variation in the number of partners for the polygamous sex (equation (3.1)). For the monogamous sex, the prevalence of infection tends to increase with higher degrees of polygamy and lower inertia. By contrast, the prevalence of infection in the polygamous sex peaks at intermediate values of α. This peak is higher for polygyny than for polyandry (parameters: $N_m = N_f = 500$; $\beta = 0.1$; $\gamma = 0.5$; $\kappa = 10^{-4}$; $\mu = 0.01$; $\sigma = 0.05$).

infection in polygynous males decreases with greater inertia, but only for intermediate values of α (figure 5*a*). In fact, the average number of males infected in a highly mobile population ($\rho = 0$) is approximately double that when movement is more limited ($\rho = 0.9$). Lower values of ρ can lead to more mixing between groups and increases mating opportunities for less attractive members of the polygamous sex. Hence, equilibrium levels of infection tend to increase with lower inertia, but the polygamous sex is disproportionately affected. Mixing tends to be much less common in our polyandrous system, as males are less able to determine whether or not they have successfully produced offspring and so generally choose to stay in a mating group. The opposite is true in the polygynous scenario, as females are always able to distinguish success from failure. This may well be a double-edged sword: although polyandry may restrict movement and limit the spread of infection within the population as a whole, it may increase infection locally.

We find that the probability of sterility (γ) is also an important factor in determining disease prevalence when movement between groups is based on reproductive failure. In particular, the difference between the polygynous and polyandrous scenarios was found to be maximized for intermediate values of γ, but only for low to intermediate values of α (figure 6).

In §2, we discussed how the ratio of male to female cases during the early stages of an epidemic could be predicted based on the transmission rates between sexes and partner acquisition rates (equation (2.4)). Figure 7 compares this prediction (using attractiveness, $a(i)$, as a proxy for partner acquisition rates) with the actual ratio of cases between the monogamous sex (C_M) and the polygamous sex (C_P) for our model. It is clear that the polyandrous system tends to have a greater bias towards infection in the monogamous sex compared with polygynous systems. As discussed earlier, this is due to increased mixing in polygynous systems exposed to a sterilizing pathogen. For low to moderate degrees of polygamy ($\alpha < 1.5$), there is very good agreement between the predicted and actual ratios for the polyandrous scenario, but is generally an overestimate for the polygynous scenario. For higher degrees of polygamy ($\alpha > 1.5$), the predicted and actual ratios tend to diverge, with the prediction increasingly underestimating the actual ratios (for the predicted values, the variance in attractiveness grows linearly with α, giving

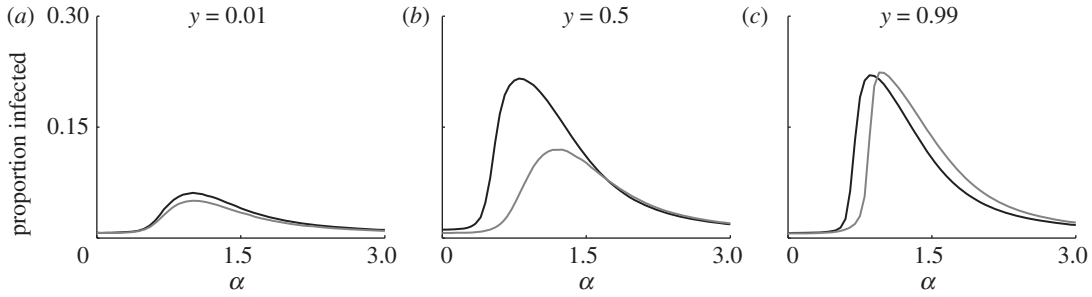

Figure 6. The average proportion of the polygamous sex that is infected for polygynous (black) and polyandrous (grey) mating systems. The difference in epidemic size between the two mating systems peaks at intermediate values of α (degree of polygamy) and γ (probability of sterilization) (parameters: $N_m = N_f = 500$; $\beta = 0.1$; $\kappa = 10^{-4}$; $\mu = 0.01$; $\rho = 0.5$; $\sigma = 0.05$).

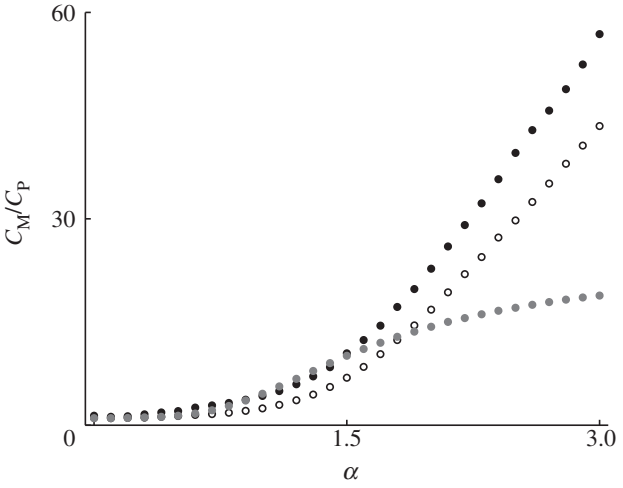

Figure 7. The ratio of cases in the monogamous sex (C_M) to the polygamous sex (C_P) during the early stages of the epidemic for varying degrees of polygamy (α). Empty circles correspond to a polygynous system, filled black circles correspond to a polyandrous system and grey circles correspond to the predicted ratio (as per equation (2.4)). The ratio C_M/C_P is generally higher in polyandrous systems than in polygynous systems. For low to moderate values of α, the prediction and actual ratios are generally in good agreement, but this breaks down as α increases.

$C_M/C_P \sim O(\sqrt{\alpha})$); this is because the network is increasingly dominated by a very small number of individuals who are in contact with almost the entire monogamous population (i.e. highly disassortative mating).

Thus far, we have only been concerned with the epidemiological dynamics of our model. We now introduce a second pathogen strain into our model in order to consider the evolutionary implications for pathogens in biased mating systems. Each pathogen strain, p, has a transmission probability per contact β_p and recovery rate σ_p, and it is assumed that there is a trade-off between these two values, such that $\beta_p = r(\sigma_p + \mu)^s$ with $r, s > 0$ parameters describing the trade-off and μ equal to the natural death rate. For human populations, the trade-off for more transmissible strains could be interpreted as an increased likelihood of seeking medical treatment due to more visible signs of disease. For simplicity, we do not allow co-infection to occur: if an individual is challenged by two different pathogens in a single time-step, then one pathogen is randomly chosen to establish an infection. Both strains are introduced at the start of each simulation and susceptible individuals are infected with probability $\beta_p I + \kappa$, where β_p is the probability of transmission per contact for pathogen p, I the total number of infectious contacts for a given individual and κ the external force of infection, as before. We assume that the two strains are equally likely to cause sterility.

Figure 8 shows the probability that each strain will account for at least 95 per cent of infections for various degrees of polygamy (α) in the polygynous and polyandrous scenarios when the trade-off between transmission probability per contact (β_p) and recovery rate (σ_p) is superlinear ($s = 1.1$). Under polygyny (figure 8*a*), the less virulent pathogen (strain 1) tends to dominate for $\alpha > 1$, the more virulent pathogen (strain 2) dominates when $\alpha < 1$ and coexistence is most common when $\alpha = 1$. The pattern is similar for polyandry (figure 8*b*), but for $\alpha < 1$ neither strain is able to become widely established, allowing both strains to coexist at low levels. For linear and sublinear trade-offs, the less virulent pathogen tends to dominate, but coexistence is often still possible (see the electronic supplementary material, figures S1 and S2).

A possible reason that the less virulent pathogen tended to be favoured for high degrees of polygamy is that maintaining infection in the dominant member of the mating group is likely to significantly contribute to the survival of a particular strain. Hence long infectious periods (low σ) are likely to be favoured when the degree of polygamy is high, even if this results in a lower basic reproductive number (i.e. if $s > 1$). By contrast, a higher transmission rate may be favoured when the distribution of partners is much more even (small α), as maintaining infection in the most dominant individual in the population becomes less important.

Typically, when multiple pathogens compete for the same pool of hosts, it is predicted that the pathogen with the highest basic reproductive number (R_0) will drive all others to extinction [8]. Host heterogeneity can complicate this picture for two reasons. Firstly, different behavioural patterns between groups will select for different traits in pathogens [33]. Secondly, components of real SCNs are likely to remain unconnected for long periods of time, owing to a combination of spatial constraints, assortative mixing and serial monogamy. Distinct components may increase divergent selection owing to behavioural differences, but they may also provide spatial refugia for less competitive strains to persist.

4. Discussion

Host heterogeneity is known to play an important role in the epidemiological dynamics of many infectious diseases, but is

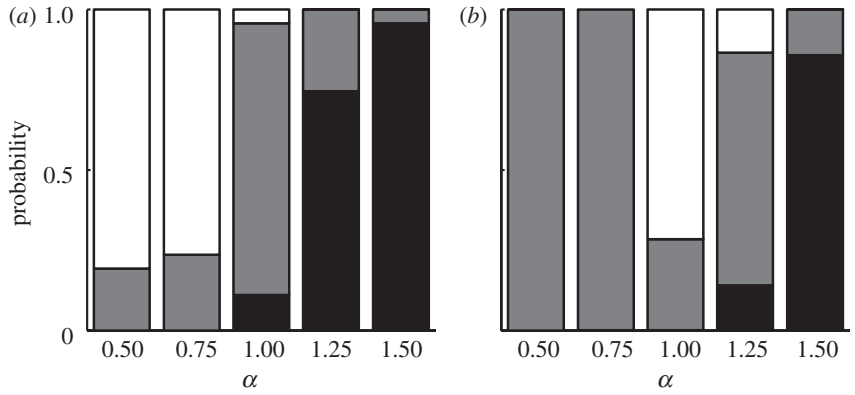

Figure 8. Probabilities of different outcomes when two strains compete, with varying degrees of polygamy (α) for (*a*) polygynous and (*b*) polyandrous mating systems. Black bars correspond to strain 1 (less virulent) dominating, white bars to strain 2 (more virulent) dominating and grey bars to coexistence. A strain is defined to be dominating if it accounts for at least 95% of infections during the final 20% of a simulation (parameters: $N_m = N_f = 500$; $r = 1.5$; $s = 1.1$; $\beta_1 \approx 0.07$; $\beta_2 \approx 0.56$; $\gamma = 0.5$; $\kappa = 10^{-4}$; $\mu = 0.01$; $\rho = 0.5$; $\sigma_1 = 0.05$; $\sigma_2 = 0.4$).

particularly relevant to STIs due to the potential for high variability in partner acquisition rates and sexual mixing patterns. High levels of host heterogeneity are usually found in biased mating systems, where large variance in mating success for one sex may leave many individuals isolated from the SCN. Such isolation has been observed in a variety of polygynous and polyandrous mating systems. For example, observations of elephant seals (*Mirounga angustirostris*) indicate that two-thirds of males do not mate during a breeding season and that a high proportion of males are unlikely to breed at all during their lifetime [34]. Similarly, approximately 75 per cent of matings in the insect *Zorotypus gurneui* are carried out by dominant males, leaving most males with only sporadic opportunities for mating [35]. Other mating systems may contain groups of males and females where mating opportunities are limited for subordinate members of the group, as observed in tamarins (*Saguinus*) [36] and white-winged trumpeters (*Psophia leucoptera*) [37]; in both cases, subordinate females are much less likely to copulate than the dominant female. Even in the absence of a strong dominance hierarchy, populations may still exhibit high variance in partner acquisition rates between the sexes. This is particularly evident among human populations in parts of Central, Eastern and Southern Africa, where male migrant workers are often separated from stable female partners for long periods of time and are concentrated in populations with highly unbalanced sex ratios [38]. These factors, along with high unemployment levels for young women, are thought to be responsible for the presence of large numbers of casual sex workers [38]. For these populations, the distribution of partner acquisition rates for males is usually characterized by a high mean and low variance, whereas the distribution for females will have a low mean and high variance.

In general, large skews in risk and reward for level of sexual activity will lead to high levels of host heterogeneity and will be conducive to the evolution of mate choice. Similarly, the complex decisions that govern mate choice and movement between mating groups are likely to be of crucial importance for the evolution of virulence in STIs. Although the extension of our model to investigate pathogen competition was fairly simplistic, it lends some credence to the notion of coevolution between mating systems and STIs. In particular, the finding that less virulent strains with long infectious periods may be favoured in highly skewed mating systems provides an interesting contrast to studies that have found virulence to increase with greater potential for pathogen dispersal [33,39,40].

We made the assumption that individuals are generally averse to leaving a mating group, and will only do so if they are certain that they have not produced offspring. Inertia to change could be due to a number of social and environmental factors, such as competition for mates, exposure to new pathogens, increased predation risk or time and energy costs of mate searching [31]. Conversely, we could have assumed that individuals only stay within a mating group if they are certain of their success. In this alternative scenario, we would expect levels of infection to be generally higher under polyandry than under polygyny.

We based movement between mating groups on a binary decision (perception of reproductive success/failure), but real systems are likely to demonstrate more complex decisions with regard to mate choice. For example, female fur seals (*Arctocephalus gazelle*) show preference for unrelated, heterozygous males, even if this choice requires increased movement [41], and monogamous bird pairs may still divorce even if they have successfully reared young together, indicating that other factors contribute to mate choice [32]. Still, one might speculate that by tending to cause sterility rather than mortality [42], sexually transmitted pathogens could increase divorce rates and movement between mating groups, thereby leading to a higher incidence of disease.

In our model, we chose to hold the host population size constant, so as to keep the degree of polygamy (i.e. the distribution of $a(i)$, equation (3.1)) fixed and to ensure that any variation in epidemiological dynamics could be unequivocally ascribed to behavioural differences between mating systems. However, it is possible that widespread sterilization could put host populations under considerable pressure: for example, it has been suggested that sterilizing *Chlamydia psittaci* infections may have contributed to declines in koala (*Phascolarctos cinereus*) populations [43]. Similarly, parasitic nematodes (*Trichostrongylus tenuis*) can reduce host fecundity and are believed to be the primary cause of population crashes among red grouse (*Lagopus lagopus scoticus*) [44]. Thus, it is probable that sterilizing infections could lead to counter-adaptations in the host, such as resistance, tolerance and more advanced mate inspection.

It is widely believed that bright plumage and other sexually selected ornaments could be indicators of resistance to parasitism, so that high-quality mates can be readily identified [45]. In proposing the 'good-genes' theory, Hamilton and Zuk avoid the lek paradox (the depletion of genetic variation) by arguing that genetic variation can be maintained by parasite counter-adaptation, resulting in coevolutionary cycling [45]. An alternative resolution to the lek paradox is the pathogen avoidance argument, which suggests that there is a trade-off between obtaining high-quality genes (maximizing the fitness of offspring) and risking exposure to infection (potential reproductive failure) [7,46–49]. Neither theory is likely to be universally applicable [50], but it is conceivable that the presence or the absence of virulent STIs could be used to distinguish between the two in some circumstances. STIs are highly relevant to the pathogen avoidance theory, but are less important to the good-genes theory [45] due to the tendency for STIs to be asymptomatic [42]. As such, one might expect to witness the good-genes theory in action if STIs result in obvious host deterioration, are uncommon or are avirulent, but pathogen avoidance behaviour should be more conspicuous among species with virulent, less visible STIs.

5. Future directions

Mathematical models have shown how heterogeneity in sexual behaviour can shape epidemiological dynamics [8,13,21] and influence the efficacy of intervention programmes [51–53]. Still, there are many consequences of host heterogeneity that are yet to be fully understood. The models introduced here and elsewhere [7,54,55] suggest that different mating strategies between the sexes can lead to considerable variation in the dynamics of STIs, but many of the evolutionary consequences of such heterogeneity are yet to be determined. For instance, Boots & Knell [56] explored a system where hosts exhibited either risky (highly active) or safe (less active) mating strategies in the presence of a sterilizing STI and found that both strategies are able to coexist for a wide range of parameters, provided the risky strategy carries a fitness benefit in the absence of disease. However, the authors did not explore how non-random mixing might affect the coexistence of risky and safe mating strategies, or whether coexistence is possible when there is a greater degree of host heterogeneity.

Simulated epidemics on SCNs are the most realistic models available for the spread of STIs in human and animal populations, but they are computationally intensive and are often difficult to parametrize. Pairwise approximations [57] offer some of the realism of SCNs by tracking the formation and break-up of partnerships, making them analytically and computationally more manageable than SCNs, but at the cost of neglecting wider population structure. Pairwise approximations are still in their infancy, but have been shown to exhibit dynamics similar to full simulations on SCNs [30,33,57]. Various degrees of host heterogeneity [57] and assortative mixing [33] have been modelled using this approach, but there is considerable scope for further research on these topics. In addition, pairwise approximations for polygynous/polyandrous mating systems are noticeably absent from the literature.

Various models incorporating heterogeneity in host contact structure have been used to study the evolution of pathogen virulence [33,39,58–60] and of antigenic diversity [61,62], but these have not been widely applied to STIs. Models have also been used to explore the role of STIs in the evolution of host-mating strategies [7,54,63], but there have been very few studies that have combined these approaches to explore coevolution between hosts and STIs. As an exception, Prado *et al.* [64] explored how host sociality (i.e. contact frequency) and pathogen virulence may coevolve on a contact network. The authors found that high levels of sociality tend to benefit more virulent pathogens, but then selection will favour more cautious hosts and subsequent reductions in virulence, which can lead to coevolutionary cycling in these traits. An exciting avenue for future work in this area would be to explicitly incorporate host and pathogen genetics within a coevolutionary framework with some plasticity in mating strategies, particularly as the predictions of such models may be amenable to testing in a wide variety of animal systems.

We thank A. Gardner and S. West for comments on the manuscript. We are grateful to the BBSRC and the ERC (ERC Advanced Grant, Diversity) for funding this work. S.G. is a Royal Society Wolfson Research Fellow.

References

1. Sherman PW, Seeley TD, Reeve HK. 1988 Parasites, pathogens, and polyandry in social Hymenoptera. *Am. Nat.* **131**, 602–610. (doi:10.2307/2461747)
2. Dixson AF. 1998 *Primate sexuality: comparative studies of the prosimians, monkeys, apes, and humans*. Oxford, UK: Oxford University Press.
3. Jennions MD, Petrie M. 2000 Why do females mate multiply? A review of the genetic benefits. *Biol. Rev. Camb. Philos.* **75**, 21–64. (doi:10.1111/j.1469-185X.1999.tb00040.x)
4. Liljeros F, Edling CR, Amaral LAN, Stanley HE, Aberg Y. 2001 The web of human sexual contacts. *Nature* **411**, 907–908. (doi:10.1038/35082140)
5. Boomsma JJ, Fjerdingstad EJ, Frydenberg J. 1999 Multiple paternity, relatedness and genetic diversity in *Acromyrmex* leaf-cutter ants. *Proc. R. Soc. Lond. B* **266**, 249–254. (doi:10.1098/rspb.1999.0629)
6. Emlen ST, Oring LW. 1977 Ecology, sexual selection, and the evolution of mating systems. *Science* **197**, 215–223. (doi:10.1126/science.327542)
7. Thrall PH, Antonovics J, Dobson AP. 2000 Sexually transmitted diseases in polygynous mating systems: prevalence and impact on reproductive success. *Proc. R. Soc. Lond. B* **267**, 1555–1563. (doi:10.1098/rspb.2000.1178)
8. Anderson RM, May RM. 1991 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.
9. Blanchard JF. 2002 Populations, pathogens, and epidemic phases: closing the gap between theory and practice in the prevention of sexually transmitted diseases. *Sex. Transm. Infect.* **78**, i183–i188. (doi:10.1136/sti.78.suppl_1.i183)
10. Edmunds WJ, Kafatos G, Wallinga J, Mossong JR. 2006 Mixing patterns and the spread of close-contact infectious diseases. *Emerg. Themes Epidemiol.* **3**, 10. (doi:10.1186/1742-7622-3-10)
11. Mossong J *et al*. 2008 Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74. (doi:10.1371/journal.pmed.0050074)
12. Woolhouse MEJ *et al*. 1997 Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* **94**, 338–342. (doi:10.1073/pnas.94.1.338)

13. Hethcote HW, Yorke JA. 1984 *Gonorrhea: transmission dynamics and control*. Springer Lecture Notes in Biomathematics, vol. 56. Berlin, Germany: Springer.
14. Anderson RM, Medley GF, May RM, Johnson AM. 1986 A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. *IMA J. Math. Appl. Med.* **3**, 229–263. (doi:10.1093/imammb/3.4.229)
15. May RM, Gupta S, McLean AR. 2001 Infectious disease dynamics: what characterizes a successful invader? *Phil. Trans. Soc. Lond. B* **356**, 901–910. (doi:10.1098/rstb.2001.0866)
16. Anderson RM, May RM, Boily MC, Garnett GP, Rowley JT. 1991 The spread of HIV-1 in Africa: sexual contact patterns and the predicted demographic impact of AIDS. *Nature* **352**, 581–589. (doi:10.1038/352581a0)
17. Ghani AC, Garnett GP. 1998 Measuring sexual partner networks for transmission of sexually transmitted diseases. *J. Roy. Stat. Soc. A. Stat.* **161**, 227–238. (doi:10.1111/1467-985X.00101)
18. Haraldsdottir S, Gupta S, Anderson RM. 1992 Preliminary studies of sexual networks in a male homosexual community in Iceland. *J. AIDS* **5**, 374–381.
19. Wylie JL, Jolly A. 2001 Patterns of chlamydia and gonorrhea infection in sexual networks in Manitoba, Canada. *Sex. Transm. Dis.* **28**, 14–24. (doi:10.1097/00007435-200101000-00005)
20. Jacquez JA, Simon CP, Koopman J, Sattenspiel L, Perry T. 1988 Modeling and analyzing HIV transmission: the effect of contact patterns. *Math. Biosci.* **92**, 119–199. (doi:10.1016/0025-5564(88)90031-4)
21. Gupta S, Anderson RM, May RM. 1989 Networks of sexual contacts: implications for the pattern of spread of HIV. *AIDS* **3**, 807–817. (doi:10.1097/00002030-198912000-00005)
22. Ghani AC, Donnelly CA, Garnett GP. 1998 Sampling biases and missing data in explorations of sexual partner networks for the spread of sexually transmitted diseases. *Stat. Med.* **17**, 2079–2097. (doi:10.1002/(SICI)1097-0258(19980930)17: 18<2079::AID-SIM902>3.0.CO;2-H)
23. Woodhouse DE *et al*. 1994 Mapping a social network of heterosexuals at high risk for HIV infection. *AIDS* **8**, 1331–1336. (doi:10.1097/00002030-199409000-00018)
24. Potterat JJ, Phillips-Plummer L, Muth SQ, Rothenberg RB, Woodhouse DE, Maldonado-Long TS, Zimmerman HP, Muth JB. 2002 Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs. *Sex. Transm. Infect.* **78**, i159–i163. (doi:10.1136/sti.78.suppl_1.i159)
25. Bearman PS, Moody J, Stovel K. 2004 Chains of affection: the structure of adolescent romantic and sexual networks. *Am. J. Sociol.* **110**, 44–91. (doi:10.1086/386272)
26. De P, Singh AE, Wong T, Yacoub W, Jolly AM. 2004 Sexual network analysis of a gonorrhoea outbreak. *Sex. Transm. Infect.* **80**, 280–285. (doi:10.1136/sti.2003.007187)
27. Keeling M. 2005 The implications of network structure for epidemic dynamics. *Theor. Popul. Biol.* **67**, 1–8. (doi:10.1016/j.tpb.2004.08.002)
28. Watts CH, May RM. 1992 The influence of concurrent partnerships on the dynamics of HIV/AIDS. *Math. Biosci.* **108**, 89–104. (doi:10.1016/0025-5564(92)90006-I)
29. Morris M, Kretzschmar M. 1997 Concurrent partnerships and the spread of HIV. *AIDS* **11**, 641–648. (doi:10.1097/00002030-199705000-00012)
30. Eames KTD, Keeling MJ. 2004 Monogamous networks and the spread of sexually transmitted diseases. *Math. Biosci.* **189**, 115–130. (doi:10.1016/j.mbs.2004.02.003)
31. Jennions MD, Petrie M. 1997 Variation in mate choice and mating preferences: a review of causes and consequences. *Biol. Rev. Camb. Philos.* **72**, 283–327. (doi:10.1111/j.1469-185X.1997.tb00015.x)
32. Dubois F, Cézilly F. 2002 Breeding success and mate retention in birds: a meta-analysis. *Behav. Ecol. Sociobiol.* **52**, 357–364. (doi:10.1007/s00265-002-0521-z)
33. Eames KTD, Keeling MJ. 2006 Coexistence and specialization of pathogen strains on contact networks. *Am. Nat.* **168**, 230–241. (doi:10.1086/505760)
34. Le Boeuf BJ. 1974 Male-male competition and reproductive success in elephant seals. *Am. Zool.* **14**, 163–176. (doi:10.1093/icb/14.1.163)
35. Choe JC. 1994 Sexual selection and mating system in Zorotypus *gurneyi* Choe (Insecta: Zoraptera): I. Dominance hierarchy and mating success. *Behav. Ecol. Sociobiol.* **34**, 87–93. (doi:10.1007/BF00164179)
36. Garber PA. 1997 One for all and breeding for one: cooperation and competition as a tamarin reproductive strategy. *Evol. Anthropol.* **5**, 187–199. (doi:10.1002/(SICI)1520-6505(1997)5:6<187::AID-EVAN1>3.0.CO;2-A)
37. Sherman PT. 1995 Social organization of cooperatively polyandrous white-winged trumpeters (*Psophia leucoptera*). *Auk* **112**, 296–309. (doi:10.2307/4088718)
38. Hunt CW. 1989 Migrant labor and sexually transmitted disease: AIDS in Africa. *J. Health. Soc. Behav.* **30**, 353–373. (doi:10.2307/2136985)
39. Lipsitch M, Nowak MA. 1995 The evolution of virulence in sexually transmitted HIV/AIDS. *J. Theor. Biol.* **174**, 427–440. (doi:10.1006/jtbi.1995.0109)
40. Boots M, Sasaki A. 1999 'Small worlds' and the evolution of virulence: infection occurs locally and at a distance. *Proc. R. Soc. Lond. B* **266**, 1933–1938. (doi:10.1098/rspb.1999.0869)
41. Hoffman JI, Forcada J, Trathan PN, Amos W. 2007 Female fur seals show active choice for males that are heterozygous and unrelated. *Nature* **445**, 912–914. (doi:10.1038/nature05558)
42. Lockhart AB, Thrall PH, Antonovics J. 1996 Sexually transmitted diseases in animals: ecological and evolutionary implications. *Biol. Rev. Camb. Philos.* **71**, 415–471. (doi:10.1111/j.1469-185X.1996.tb01281.x)
43. Weigler BJ, Girjes AA, White NA, Kunst ND, Carrick FN, Lavin MF. 1988 Aspects of the epidemiology of *Chlamydia psittaci* infection in a population of koalas (*Phascolarctos cinereus*) in southeastern Queensland, Australia. *J. Wildlife Dis.* **24**, 282–291.
44. Hudson PJ, Dobson AP, Newborn D. 1998 Prevention of population cycles by parasite removal. *Science* **282**, 2256–2258. (doi:10.1126/science.282.5397.2256)
45. Hamilton WD, Zuk M. 1982 Heritable true fitness and bright birds: a role for parasites? *Science* **218**, 384–387. (doi:10.1126/science.7123238)
46. Hamilton WD. 1990 Mate choice near or far. *Am. Zool.* **30**, 341–352. (doi:10.1093/icb/30.2.341)
47. Loehle C. 1997 The pathogen transmission avoidance theory of sexual selection. *Ecol. Mod.* **103**, 231–250. (doi:10.1016/S0304-3800(97)00106-3)
48. Lombardo MP. 1998 On the evolution of sexually transmitted diseases in birds. *J. Avian Biol.* **29**, 314–321. (doi:10.2307/3677114)
49. Qvarnström A, Forsgren E. 1998 Should females prefer dominant males? *Trends Ecol. Evol.* **13**, 498–501. (doi:10.1016/S0169-5347(98)01513-4)
50. Hamilton WD, Poulin R. 1997 The Hamilton and Zuk hypothesis revisited: a meta-analytical approach. *Behaviour* **134**, 299–320. (doi:10.1163/156853997X00485)
51. Boily MC, Lowndes C, Alary M. 2002 The impact of HIV epidemic phases on the effectiveness of core group interventions: insights from mathematical models. *Sex. Transm. Infect.* **78**, i78–i90. (doi:10.1136/sti.78.suppl_1.i78)
52. Eames KTD, Keeling MJ. 2003 Contact tracing and disease control. *Proc. R. Soc. Lond. B* **270**, 2565–2571. (doi:10.1098/rspb.2003.2554)
53. Eames KTD. 2007 Contact tracing strategies in heterogeneous populations. *Epidemiol. Infect.* **135**, 443–454. (doi:10.1017/S0950268806006923)
54. Thrall PH, Antonovics J, Bever JD. 1997 Sexual transmission of disease and host mating systems: within-season reproductive success. *Am. Nat.* **149**, 485–506. (doi:10.1086/286001)
55. Gouveia-Oliveira R, Pedersen AG. 2009 Higher variability in the number of sexual partners in males can contribute to a higher prevalence of sexually transmitted diseases in females. *J. Theor. Biol.* **261**, 100–106. (doi:10.1016/j.jtbi.2009.06.028)
56. Boots M, Knell RJ. 2002 The evolution of risky behaviour in the presence of a sexually transmitted disease. *Proc. R. Soc. Lond. B* **269**, 585–589. (doi:10.1098/rspb.2001.1932)
57. Eames KTD, Keeling MJ. 2002 Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc. Natl Acad. Sci. USA* **99**, 13 330–13 335. (doi:10.1073/pnas.202244299)

58. Lipsitch M, Herre EA, Nowak MA. 1995 Host population structure and the evolution of virulence: A 'law of diminishing returns'. *Evolution* **49**, 743–748. (doi:10.2307/2410327)
59. Read JM, Keeling MJ. 2003 Disease evolution on networks: the role of contact structure. *Proc. R. Soc. Lond. B* **270**, 699–708. (doi:10.1098/rspb.2002.2305)
60. Messinger SM, Ostling A. 2009 The consequences of spatial structure for the evolution of pathogen transmission rate and virulence. *Am. Nat.* **174**, 441–454. (doi:10.1086/605375)
61. Buckee COF, Koelle K, Mustard MJ, Gupta S. 2004 The effects of host contact network structure on pathogen diversity and strain structure. *Proc. Natl Acad. Sci. USA* **101**, 10 839–10 844. (doi:10.1073/pnas.0402000101)
62. Buckee C, Danon L, Gupta S. 2007 Host community structure and the maintenance of pathogen diversity. *Proc. R. Soc. B* **274**, 1715–1721. (doi:10.1098/rspb.2007.0415)
63. Kokko H, Ranta E, Ruxton G, Lundberg P. 2002 Sexually transmitted disease and the evolution of mating systems. *Evolution* **56**, 1091–1100.
64. Prado F, Sheih A, West JD, Kerr B. 2009 Coevolutionary cycling of host sociality and pathogen virulence in contact networks. *J. Theor. Biol.* **261**, 561–569. (doi:10.1016/j.jtbi.2009.08.022)

Pathogen selection drives nonoverlapping associations between HLA loci

Bridget S. Penman[a], Ben Ashby[a], Caroline O. Buckee[b], and Sunetra Gupta[a,1]

[a]Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom; and [b]Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115

Edited* by Robert M. May, University of Oxford, Oxford, United Kingdom, and approved October 14, 2013 (received for review March 21, 2013)

Pathogen-mediated selection is commonly invoked as an explanation for the exceptional polymorphism of the HLA gene cluster, but its role in generating and maintaining linkage disequilibrium between HLA loci is unclear. Here we show that pathogen-mediated selection can promote nonrandom associations between HLA loci. These associations may be distinguished from linkage disequilibrium generated by other population genetic processes by virtue of being nonoverlapping as well as nonrandom. Within our framework, immune selection forces the pathogen population to exist as a set of antigenically discrete strains; this then drives nonoverlapping associations between the HLA loci through which recognition of these antigens is mediated. We demonstrate that this signature of pathogen-driven selection can be observed in existing data, and propose that analyses of HLA population structure can be combined with laboratory studies to help us uncover the functional relationships between HLA alleles. In a wider coevolutionary context, our framework also shows that the inclusion of memory immunity can lead to robust cyclical dynamics across a range of host–pathogen systems.

infectious disease | major histocompatibility complex | mathematical model | human evolution | population genetics

HLAs, found on the surface of all nucleated cells, present pathogen peptides to T lymphocytes and are thus a keystone of adaptive immunity. Demonstrable associations of particular HLA alleles with resistance or susceptibility to severe disease (1, 2) underscore the importance of their role in protection against death from infection. The genes encoding HLAs are found in the 3.6-Mb-long MHC on chromosome 6 and are distinguished by their exceptional polymorphism (3), which is likely the result of selection from pathogens (4–6) Despite this enormous diversity, most human populations are dominated by a relatively small number of combinations of the alleles present at the class I HLA (A, B, C) and the principal class II HLA (DP, DQ, and DR) loci (7–12). Here we present a coevolutionary model demonstrating that pathogen selection can drive such long-term, long-range associations between HLAs. We show that this mechanistic process can be distinguished from other evolutionary effects by virtue of generating a higher degree of nonoverlap between HLA repertoires than might be expected under founder effects or hitchhiking.

A Multilocus Model for Host–Pathogen Coevolution with Allele-Specific Adaptive Immunity

We first explored the properties of a deterministic epidemiological model (*Methods*) in which (*i*) the pathogen population was represented by four potential strains defined by two antigenic loci containing alleles (a, b) and (x, y), respectively, and (*ii*) we defined within a diploid host, alleles (A,B) and (X,Y) at two linked "recognition loci" (i.e., HLA loci), each only capable of responding to the corresponding parasite allele (or epitope) given in lowercase above. We assumed that immunity developed in an allele-specific manner conferring complete protection against infection by any other antigenic type containing that allele, but that there was a risk of death if a host was incapable of recognizing either allele of the infecting pathogen strain (Fig. 1).

In line with previous observations, the pathogen population was observed to adopt a discrete, nonoverlapping strain structure (13). However, once any two strains (e.g., *ax* and *by*) achieve dominance, host homozygotes AX/AX and BY/BY suffer from increased mortality because each is only able to mount an immune response against one of the two circulating pathogen strains (all other host genotypes can recognize at least one allele of both strains). The numbers of these homozygotes fall until eventually the only host haplotypes left in the population are AY and BX. Thus, the strain structuring of the pathogen population by host immune selection generates nonrandom associations among the immune recognition genes of the host.

This scenario will be stable (Fig. 2*A*) in the absence of pathogen mutation, or when the basic reproduction number of the pathogen (R_o; a measure of its fundamental transmission potential) (14) is low. Conversely, if R_o is above a certain threshold, no genetic structuring is possible in either pathogen or host (*SI Appendix*, Fig. S1). Between these two extremes, we observe coevolutionary cycling (Fig. 2*B*) in place of permanent structuring. This dynamic emerges due to the fact that as soon as the pathogen population becomes dominated by a particular set of strains (say *ax* and *by*), haplotypes that are incapable of recognizing any one of the dominant pathogen strains (i.e., BY and AX) start to go down in frequency, and haplotypes that can recognize both the dominant pathogen strains (i.e., BX and AY) increase in frequency. Eventually the proportion of BX and AY in the population will be so high that it will be in the pathogen's interest to switch its strain structure to (*ay, bx*) so as to exploit the infection reservoirs created by homozygotes of these haplotypes (BX/BX cannot become immune to *ay*). The system is capable of generating nonrandom associations between recognition alleles

Significance

Human leukocyte antigens (HLA), first identified in tissue-matching for transplantation, play a critical role in immunity. HLAs are extraordinarily diverse, but certain sets of HLA genes are more likely to be found together than others. Here, we show that associations between HLA genes can arise through their coevolutionary interaction with pathogens. Technological advances are making it easier to determine HLA types, but DNA sequence alone cannot fully predict an HLA's functional properties. Our work offers a new evolutionary approach to tackling this problem.

Author contributions: S.G. designed research; B.S.P. performed research; B.A. and C.O.B. contributed new reagents/analytic tools; B.S.P. analyzed data; and B.S.P. and S.G. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: sunetra.gupta@zoo.ox.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1304218110/-/DCSupplemental.

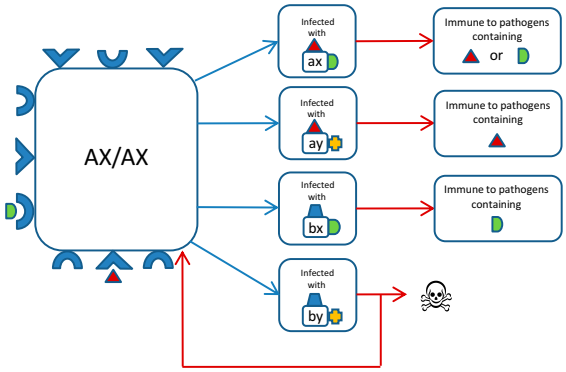

Fig. 1. A schematic representation of the key properties of the model. This flowchart illustrates the possible outcomes of infection by different pathogen strains for a host of genotype AX/AX who can only mount immune responses against pathogen epitopes *a* and *x*. Blue arrows indicate infection, and red arrows indicate recovery or death.

(both stable and cyclical) even when recombination is introduced between the host's recognition loci (*SI Appendix*, Fig. S1).

To investigate the generality of these conclusions, we generated a stochastic equivalent of this deterministic model and found it to produce the same behavior in the two-locus, two-allele case, and in higher dimensional systems (*SI Appendix*). Nonoverlapping combinations of alleles or high complementarity equilibria (HCE) have been shown to arise in multilocus population genetic models (15) where the fitness of both host and parasite depends on the number of "matching alleles." Our results are qualitatively different to HCE because although host recognition of pathogen epitopes has analogies with a matching allele mechanism, the structuring of the pathogen population in our model occurs through immune selection exerted by all host genotypes and would be maintained even in the absence of host heterogeneity (13). Matching allele models have so far not been shown to alternate between different nonoverlapping population structures; our results demonstrate that the integration of memory immunity can precipitate this form of coevolutionary cycling.

Structuring of HLA

Our model provides a unique mechanistic basis for the observation of nonrandom associations between host recognition loci such as HLA. Recombination between markers flanking a 7-Mb region containing the human MHC has been estimated at between 1.66% and 6.54% (16). Within our framework, pathogen selection is capable of generating nonrandom associations between host recognition alleles in the presence of recombination frequencies of up to 10% (*SI Appendix*, Fig. S1), and thus has the potential to maintain even long range HLA associations across recombination hotspots.

Furthermore, our model predicts that if selection from a multiepitope, strain-structured pathogen is maintaining associations between host recognition loci, alleles at those loci should not only be nonrandomly associated [i.e., in linkage disequilibrium (LD)], but also exhibit nonoverlapping repertoires (i.e., where A is principally associated with X, and B with Y). Standard metrics of LD such as D′ (17) will not capture this nonoverlapping pattern, but a previously introduced metric, *f** (18), has been used to measure nonoverlapping associations among pathogen epitopes. We made a slight modification to *f** (*Methods*) to produce a metric f^*_{adj}, which we propose can be used as an additional feature alongside LD to begin to identify the specific effects of pathogen selection.

Qualitative evidence for nonoverlapping patterns between HLA is apparent in existing studies: the Burusho population of Pakistan provides a particularly striking example (Fig. 3)

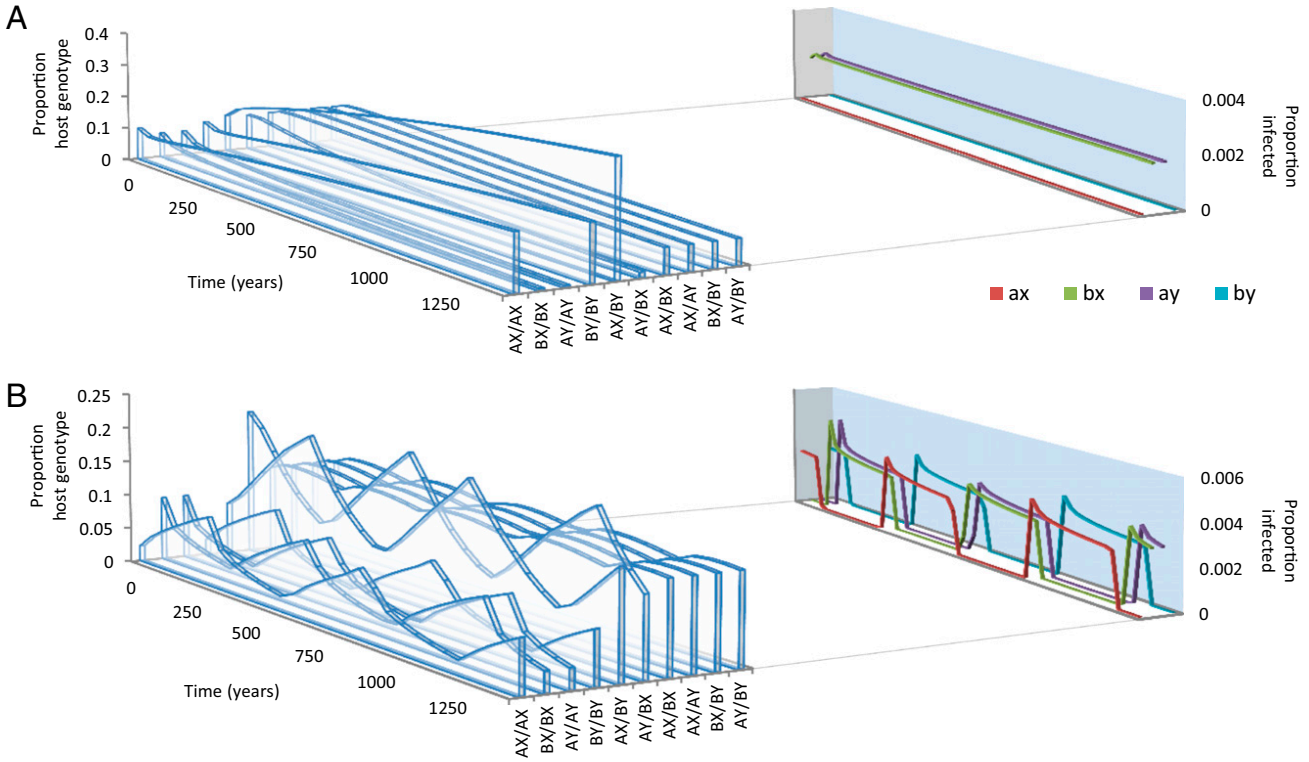

Fig. 2. The two key behaviors of the model. See *Methods* for a full description of the model and its parameters. In both panels, $\mu_1 = 0.03$; $\mu_2 = 2$; $r = 0$; $m = 0.0001$ and $\sigma = 10$. (*A*) The basic reproductive number of the pathogen, $R_o = 2.5$; (*B*) $R_o = 5$.

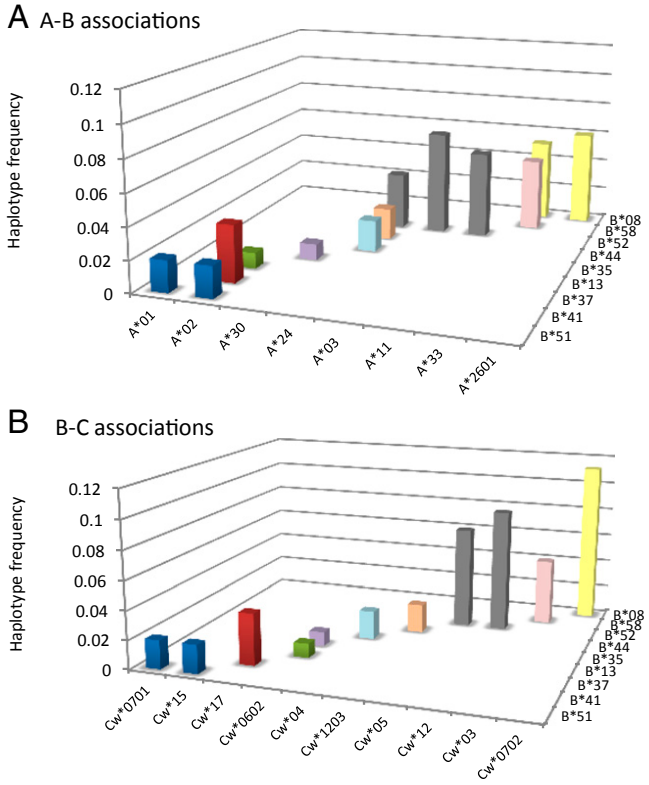

Fig. 3. Nonoverlapping HLA associations in the Burusho population in Pakistan. (*A*) HLA-A-B associations and (*B*) HLA-B-C associations.

(12). In their survey of US bone marrow donors, Cao et al. (7) point out that whenever sequence-related (i.e., serologically indistinguishable, but with different amino acid sequences) HLA-B alleles occur at similar, moderate frequencies in a population, they tend to have nonoverlapping associations with alleles at the HLA-A or HLA-C loci. However, only by considering the entire MHC region can we establish whether HLAs are especially nonoverlapping relative to other loci with a similar demographic history. A study of 962 members of the Hutterite population of South Dakota, in which 16 loci were typed, allows us to make such comparisons (19). When we apply f^*_{adj} to all possible pairwise combinations involving HLA-C (Fig. 4*A*), we find that HLA-C is highly nonoverlapping with its close neighbor HLA-B, but is also highly nonoverlapping with the much further HLA DRB1, DQA1, and DQB1 loci. Non-HLA loci in the intervening region do not display such a nonoverlapping relationship. If we repeat the exercise for TNF-α (which we would not predict to be under the same kind of pathogen selection as the HLAs), there is no particular peak in its degree of nonoverlap with physically distant loci in the region (Fig. 4*B*).

Other processes that generate LD, such as founder effects or hitchhiking, could potentially generate high f^*_{adj} between two loci; to address this, we introduced into our stochastic framework an alternative locus physically linked to one of our simulated HLA loci, but not subject to the pathogen selection acting on the HLAs (Fig. 5*A*). Randomly chosen alleles at the alternative locus were deemed favored at any given time, generating the sequential dominance of alleles in a manner that was entirely unrelated to the antigenic structure of the pathogen population (*SI Appendix*). We found that selection on the non-HLA locus was capable of generating results where HLA1/HLA2 f^*_{adj} scores ($f^*_{adj}HLA$) were greater than HLA2/non-HLA f^*_{adj} scores ($f^*_{adj}ALT$), but only when very little recombination occurred between any of the loci in the system (Fig. 5 *B* and *D*). Pathogen-driven coevolution, by contrast, ensured that $f^*_{adj}HLA > f^*_{adj}ALT$ even when the HLA loci were separated by recombination (Fig. 5 *B* and *E*). We can thus be more confident of pathogen selection having driven nonoverlap when considering long-range associations such as the HLA B/DRB pattern shown in Fig. 4.

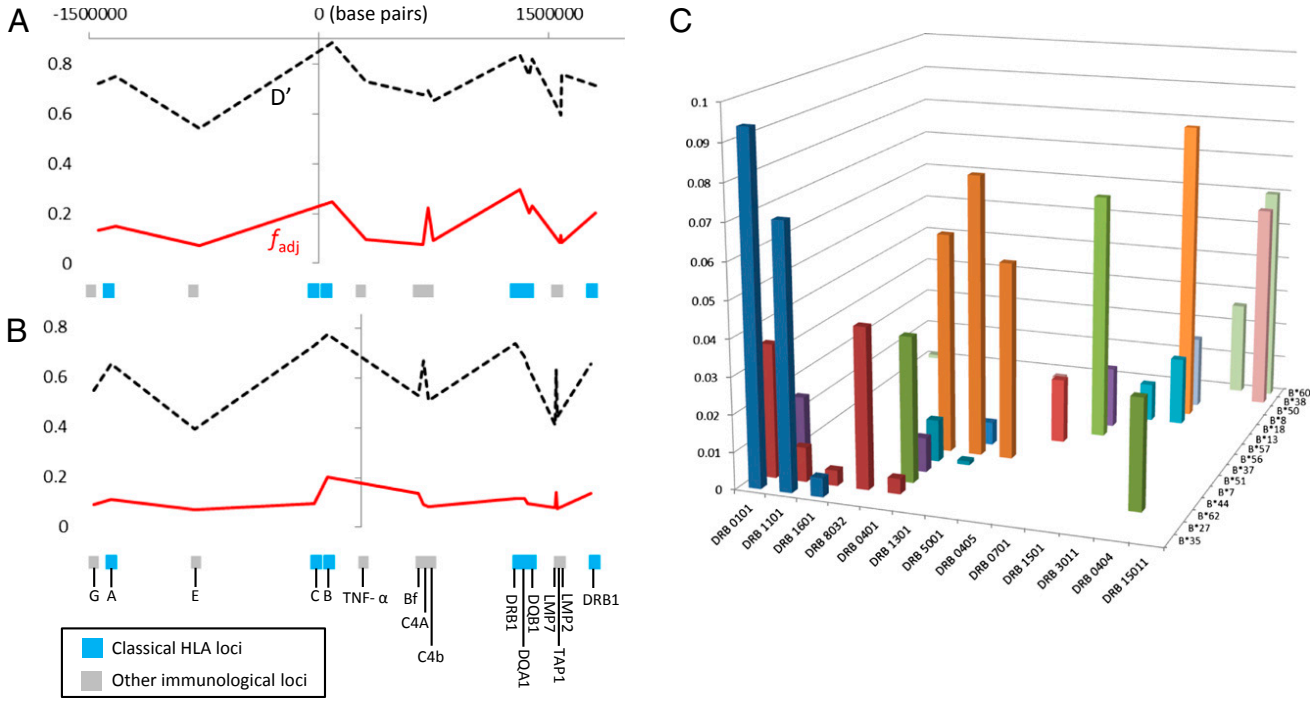

Fig. 4. Nonoverlapping MHC associations in the Hutterite population. Here we illustrate the pairwise f^*_{adj} and D′ values calculated from 16-locus HLA haplotypes recorded in the Hutterite population of South Dakota (19) for each indicated locus relative to HLA –C (*A*) or TNF-α (*B*). Only haplotypes for which all 16 loci were resolved were included in the analysis (97% of the total sample). (*C*) Frequencies of HLA-B/HLA-DRB associations within the dataset.

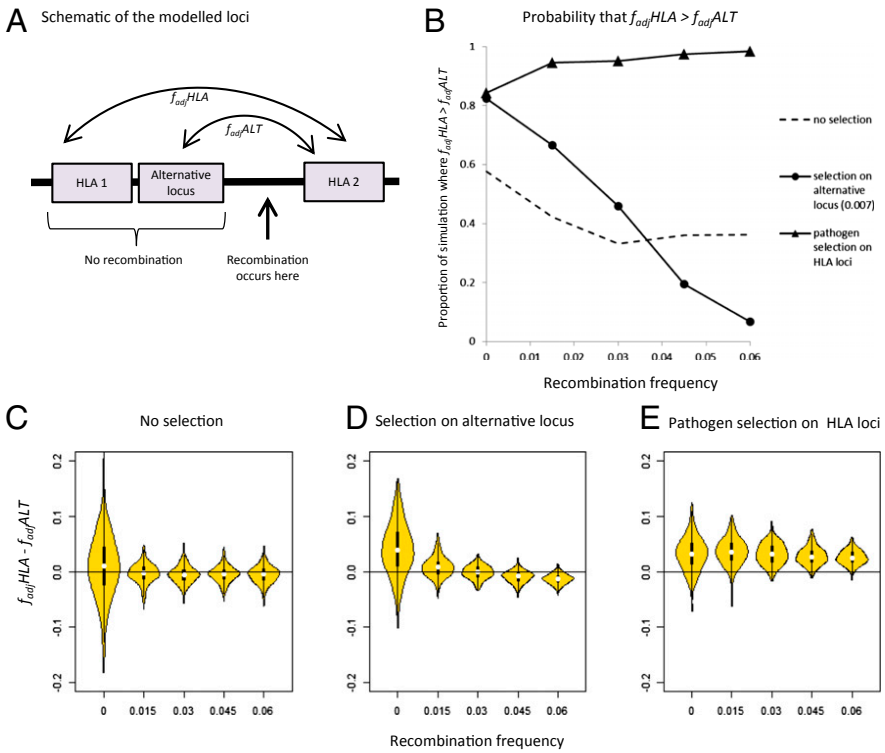

Fig. 5. Identifying a unique signature of HLA/pathogen coevolution. (*A*) Schematic representation of the loci within our model framework, indicating the pairs of loci between which f_{adj}HLA and f_{adj}ALT measure the degree of nonoverlap. (*B*) Probability that f_{adj}HLA > f_{adj}ALT under different selection regimes and levels of recombination. (*C–E*) Violin plots (32) produced using the vioplot package in R version 2.15.2, representing the distribution of values of f_{adj}HLA − f_{adj}ALT for 350 simulations at each indicated selection regime and level of recombination. Only results for surviving populations are shown. The f_{adj}HLA − f_{adj}ALT values displayed are average values calculated over the final 2,500 y of 5,000-y simulations (*Methods*). *SI Appendix* describes the stochastic framework used for the simulations, and provides parameter definitions. We let there be five possible alleles at each of the two HLA loci. Parameter values were as follows: $b = 0.07$; $m = 0.0001$; $\varphi = 0.0015$; $\alpha = 0.002$; $Q = 5$; $\theta = 1.1$; $\Omega = 0.1$; $C = 2{,}000$; $k = 0.004$. r was varied between 0 and 0.06, as indicated in the x axes of each panel. For "no selection," $\varpi = 0$ and $d = 0$; for "selection on alternative locus," $\varpi = 0$ and $d = 0.007$; and for "pathogen selection on HLA," $\varpi = 0.012$ and $d = 0$.

Future Directions

The system we present is necessarily a minimal caricature of the MHC, and suffers from a number of limitations. Most importantly, we have only considered the effects of interaction with a single pathogen. However, though a single HLA locus undoubtedly presents peptides from a variety of pathogens (as well as self), the selective pressure upon it will mainly arise from the pathogens causing the highest mortality. Take, for example, an HLA system as described by Fig. 1 and assume that it is under assault from n pathogens whose allelic variants may be represented according the convention we have established as (a_i,b_i) and (x_i,y_i); if the most deleterious pathogen adopts the configuration ($a_i x_i$, $b_i y_i$), then the homozygotes that are most disadvantaged will still be AX/AX and BY/BY.

A second important limitation of this model is that specific host recognition loci "target" specific pathogen epitopes—in other words, why should all variants at locus 1 of the pathogen specifically be recognized by locus 1 within the host? When considering associations between class I and class II HLAs, it seems justifiable to assume that different epitopes from any given pathogen are displayed by each, but it may not be strictly correct to distinguish between class I loci (particularly A and B) on this basis.

Future work in this area should also place the HLA in its wider genomic context. The very architecture of the MHC will have an effect: in the chicken, for example, the relative proximity of the TAP and class I MHC loci may have led to tight coevolution between them, limiting the possible coexpression of class I genes (20). Furthermore, in humans, HLAs interact directly with a second family of immune system genes: Killer-cell Ig-like receptors (KIRs). KIRs display a striking haplotypic structure (21); particular KIR/HLA genotypes have been associated with different infectious disease outcomes (22, 23), and a direct effect of KIR/HLA coevolution on HLA haplotypes has recently been suggested (24).

If proven to be robust, this framework may, in principle, be able to assist in developing functional classifications of HLA alleles. It is possible to categorize HLA alleles into broad "supertypes," based on their binding properties (25); at the same time, it is clear that a very small change in sequence (e.g., a single amino acid) can have very significant functional consequences (26). Furthermore, the ability of an HLA to bind to a specific pathogen epitope is not in itself a guarantee of an effective T-cell response to that epitope (27). If nonoverlapping allelic patterns are a signature of disease selection, they offer an alternative evolutionary approach to solving this problem. The multilocus framework described here provides a flexible platform for investigating the population-level consequences of interactions between diverse immune system genes and the pathogens they help recognize.

Methods

Deterministic Model. We used a system of linked ordinary differential equations to capture both the population genetics of the host and the disease dynamics of the pathogen. A range of coevolutionary frameworks have been developed to combine population genetics and epidemiology (28–30); the differential equation approach, first used by Gupta and Hill (31), offers a highly flexible framework that is especially amenable to the inclusion of immunological memory.

The pathogen population was represented by four potential strains (P = 1–4) defined by two antigenic loci containing epitopes (a, b) and (x, y) respectively (*SI Appendix*, Table S1). Our host population was diploid, possessing recognition alleles at two linked loci (A, B) and (X, Y), making up four possible host haplotypes (h = 1–4; *SI Appendix*, Table S2) and giving 10 possible host genotypes (i = 1–10; *SI Appendix*, Table S3). To mount an immune response against a pathogen epitope represented by a particular lowercase letter, a host must possess the recognition allele represented by the corresponding uppercase letter. The various combinations of epitopes, E_j, to which a host could be immune are shown in *SI Appendix*, Table S4; of these, only a subset $\{E_k\}^i$ will be accessible to host genotype i (e.g., host genotype AXAX can only be immune to epitope sets E_1, E_3, or E_5). A host immune to the epitopes in E_j can be infected by any pathotype not displaying those epitopes. Hosts immune to the epitopes in E_k can become immune to the epitopes of E_j by being infected by strain p, where strain p contains epitopes in E_j but not in E_k.

The dynamics of this system can be described by the following set of equations:

$$\frac{dN_j^i}{dt} = \alpha_j\omega_i + (1-\alpha_j)\sum_{p,k}\left(\lambda_p N_k^i\right) - \left(\sum_{q,q\neq v}\lambda_q + \mu_1\right)N_j^i - \delta_i\mu_2 G_j^i$$

$$\frac{dG_j^i}{dt} = \lambda_v\left(N_j^i - G_j^i\right) - (\sigma+\mu_1+\mu_2)G_j^i$$

$$\frac{dI_u}{dt} = \lambda_u S_u - (\sigma+\mu_1)I_u$$

Here, N_j^i is the number of hosts of genotype i who are immune to the set of epitopes E_j. G_j^i is the number of these hosts who are infected with strain v, to which they can never mount an immune response and from which they risk dying at a rate μ_2 ; this only applies to homozygous hosts in this system (Fig. 1), so $\delta_i = 0$ for all heterozygous host genotypes. I_u is the number of hosts who are currently infected with pathogen strain u and will become immune to at least one of the epitopes of strain u. S_u is the sum of all those hosts who are not yet immune to strain u but are capable of becoming immune to at least one of the epitopes of u. All individuals recover from infection at rate σ and suffer a natural mortality rate μ_1.

The force of infection with strain p is $\lambda_p = \frac{\beta(I_p+G_p)}{\sum_{i,j}N_j^i}$, where β is a transmission coefficient, such that $R_0 = \frac{\beta}{\sigma+\mu_1}$ for the pathogen in a population of hosts that can mount an immune response against it, and $R_0 = \frac{\beta}{\sigma+\mu_1+\mu_2}$ in a population of hosts that cannot mount an immune response against it. In the figures and figure legends, we always quote R_0 values for a pathogen in a host population that *can* mount an immune response against it.

Pathogen mutation can be included in the model by allowing small perturbations in the force of infection. In the model presented here we included pathogen mutation at rate m by adjusting the force of infection term, thus

$$\lambda_p^m = (1-m)\lambda_p + \frac{1}{3}\sum_{q\neq p}\lambda_q.$$

The term ω_i represents the births into the fully susceptible compartment of genotype i (thus if j = 0, α_j = 1, if j > 0, α_j = 0). The birth term for host genotype i is given by the following:

$$\omega_i = \kappa f_h f_g(1+\delta_i),$$

where κ is the total death rate for the entire population; δ_i is defined as above, and f_h and f_g are the frequencies of the haplotypes that make up host genotype i.

Haplotype frequencies are calculated as follows, where r is the host recombination rate. If r = 0.5, the two host loci are effectively unlinked.

$$f_h = \frac{2\sum_j N_j^{c_1} + \sum_j N_j^{c_2} + \sum_j N_j^{c_3} + (1-r)\sum_j N_j^{c_4} + r\sum_j N_j^{c_5}}{2\sum_{i,j}N_j^i}$$

See *SI Appendix*, Table S5 for the values of c_{1-5} that correspond to a particular haplotype.

The total death rate is calculated as follows:

$$\kappa = \mu_1\left(\sum_{i,j}N_j^i\right) + \sum_{i,j}G_j^i\mu_2.$$

Numerical simulations were carried out using the ode45 solver in MatLab version 7.10.0 (R2010b).

Stochastic Model. A full description of the stochastic model is provided in *SI Appendix*, section 1. Briefly, the population was made up of N hosts, where $N < C$, the population carrying capacity. Each host was represented by a 19-element identifier code that recorded age, genotype, infection, and immunity status. As in the deterministic model, host genotype AX/AX was only capable of becoming immune to pathogen epitopes a and x, and risked death when infected with a pathogen it could not recognize. Infection, recovery, mortality, and reproduction were all probabilistic events.

Metrics. We used a standard metric (Lewontin's D′, normalized where necessary for >2 alleles per locus, as described in ref. 17) to measure LD.

The f^* metric for nonoverlap between two loci was calculated as described in ref. 18 and adjusted as follows:

$$f^*_{adj} = (1-H_{\max})f^*,$$

where $H_{\max}$ = the frequency of the most frequent haplotype in the population. f^* takes values between 0 and 1, where values closer to 1 indicate a more nonoverlapping pattern. However, $f^* = 1$ for a population that consists of one haplotype only, which is not a case of true nonoverlap. For f^*_{adj}, by contrast, populations containing relatively balanced frequencies of nonoverlapping haplotypes will receive the highest scores.

To calculate $f^*_{adj}HLA - f^*_{adj}ALT$ from our simulations in Fig. 5, we measured $f^*_{adj}HLA - f^*_{adj}ALT$ every 20 y during the final 2,500 y of a 5,000-y simulation, and took the mean of those measurements.

ACKNOWLEDGMENTS. We thank Adrian Hill, Angus Buckling, Paul Harvey, and Oliver Pybus for their comments on the manuscript, and Adrian Smith for general guidance on this project. Funding for this work was provided by the Wellcome Trust, the European Research Council (ERC Advanced Grant – DIVERSITY), the Biotechnology and Biological Sciences Research Council, and the Christopher Welch Trust. B.S.P. is a Sir Henry Wellcome Postdoctoral Fellow (Grant 096063/Z/11/Z) and a Junior Research Fellow at Merton College, Oxford. S.G. is a Royal Society Wolfson Research Fellow and an ERC Advanced Investigator.

1. Hill AVS, et al. (1991) Common west African HLA antigens are associated with protection from severe malaria. *Nature* 352(6336):595–600.
2. Carrington M, O'Brien SJ (2003) The influence of HLA genotype on AIDS. *Annu Rev Med* 54:535–551.
3. Robinson J, et al. (2011) The IMGT/HLA database. *Nucleic Acids Res* 39(Database issue, SUPPL. 1):D1171–D1176.
4. Jeffery KJM, Bangham CRM (2000) Do infectious diseases drive MHC diversity? *Microbes Infect* 2(11):1335–1341.
5. Hedrick PW (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution* 56(10):1902–1908.
6. Trowsdale J (2011) The MHC, disease and selection. *Immunol Lett* 137(1-2):1–8.
7. Cao K, et al. (2001) Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Hum Immunol* 62(9):1009–1030.
8. Cao K, et al. (2004) Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. *Tissue Antigens* 63(4):293–325.
9. Shaw CK, Chen LL, Lee A, Lee TD (1999) Distribution of HLA gene and haplotype frequencies in Taiwan: A comparative study among Min-nan, Hakka, Aborigines and Mainland Chinese. *Tissue Antigens* 53(1):51–64.
10. Cox ST, et al. (1999) HLA-A, -B, -C polymorphism in a UK Ashkenazi Jewish potential bone marrow donor population. *Tissue Antigens* 53(1):41–50.
11. Buhler S, Nunes JM, Nicoloso G, Tiercy JM, Sanchez-Mazas A (2012) The heterogeneous HLA genetic makeup of the Swiss population. *PLoS ONE* 7(7):e41400.
12. Mohyuddin A, et al. (2002) HLA polymorphism in six ethnic groups from Pakistan. *Tissue Antigens* 59(6):492–501.
13. Gupta S, Ferguson N, Anderson R (1998) Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science* 280(5365):912–915.
14. Anderson RM, May RM (1991) *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ Press, New York).
15. Kouyos RD, Salathé M, Otto SP, Bonhoeffer S (2009) The role of epistasis on the evolution of recombination in host-parasite coevolution. *Theor Popul Biol* 75(1):1–13.
16. Carrington M (1999) Recombination within the human MHC. *Immunol Rev* 167:245–256.
17. Hedrick PW (1987) Gametic disequilibrium measures: Proceed with caution. *Genetics* 117(2):331–341.
18. Buckee CO, Gupta S, Kriz P, Maiden MCJ, Jolley KA (2010) Long-term evolution of antigen repertoires among carried Meningococci. *Proc R Soc B Biol Sci* 277(1688):1635–1641.
19. Weitkamp LR, Ober C (1999) Ancestral and recombinant 16-locus HLA haplotypes in the Hutterites. *Immunogenetics* 49(6):491–497.

20. Walker BA, et al. (2011) The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. *Proc Natl Acad Sci USA* 108(20):8396–8401.
21. Parham P (2005) MHC class I molecules and KIRs in human history, health and survival. *Nat Rev Immunol* 5(3):201–214.
22. Martin MP, et al. (2007) Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. *Nat Genet* 39(6):733–740.
23. Seich Al Basatena NK, et al. (2011) KIR2DL2 enhances protective and detrimental HLA class I-mediated immunity in chronic viral infection. *PLoS Pathog* 7(10):e1002270.
24. Capittini C, et al. (2012) Possible KIR-driven genetic pressure on the genesis and maintenance of specific HLA-A,B haplotypes as functional genetic blocks. *Genes Immun* 13(6):452–457.
25. Sidney J, Peters B, Frahm N, Brander C, Sette A (2008) HLA class I supertypes: A revised and updated classification. *BMC Immunol* 9:1.
26. Kløverpris HN, et al. (2012) HIV control through a single nucleotide on the HLA-B locus. *J Virol* 86(21):11493–11500.
27. Assarsson E, et al. (2007) A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. *J Immunol* 178(12):7890–7901.
28. Gillespie JH (1975) Natural selection for resistance to epidemics. *Ecology* 56:493–495.
29. May RM, Anderson RM (1983) Parasite–host coevolution. *Coevolution*, eds Futuyma DJ, Slatkin M (Sinauer, Sunderland, MA).
30. Antonovics J, Thrall PH (1994) The cost of resistance and the maintenance of genetic polymorphism in host-pathogen systems. *Proc Biol Sci* 257(1349):105–110.
31. Gupta S, Hill AVS (1995) Dynamic interactions in malaria: Host heterogeneity meets parasite polymorphism. *Proc Biol Sci* 261(1362):271–277.
32. Hintze JL, Nelson RD (1998) Violin plots: A box plot-density trace synergism. *Am Stat* 52(2):181–184.

References

Able, D. J., 1996. The contagion indicator hypothesis for parasite-mediated sexual selection. Proc. Natl. Acad. Sci. U. S. A. 93:2229–2233.

Adamo, S. a., I. Kovalko, R. H. Easy, and D. Stoltz, 2014. A viral aphrodisiac in the cricket Gryllus texensis. J. Exp. Biol. 217:1970–1976.

Agrawal, A. F., 2009. Differences between selection on sex versus recombination in red queen models with diploid hosts. Evolution 63:2131–2141.

Agrawal, A. F. and C. M. Lively, 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evol. Ecol. Res. 4:79–90.

———, 2003. Modelling infection as a two-step process combining gene-for-gene and matching-allele genetics. Proc. R. Soc. B 270:323–334.

Agrawal, A. F. and S. P. Otto, 2006. host–parasite coevolution and selection on sex through the effects of segregation. Am. Nat. 168:617–629.

Agrios, G. N., 1997. Plant Pathology. 5th ed. Academic Press, London, UK.

Allison, A. C., 1954. Protection afforded by sickle-cell trait against subtertian malareal infection. Br. Med. J. 1:290–294.

Anderson, R. M. and R. M. May, 1982. Coevolution of hosts and parasites. Parasitology 85:411–426.

———, 1991. Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford, UK.

Anderson, R. M., R. M. May, M. C. Boily, G. P. Garnett, and J. T. Rowley, 1991. The spread of HIV-1 in Africa: sexual contact patterns and the predicted demographic impact of AIDS. Nature 352:581–589.

Anderson, R. M., G. F. Medley, R. M. May, and A. M. Johnson, 1986. A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. IMA J. Math. Appl. Med. Biol. 3:229–263.

Antonovics, J., M. Boots, D. Ebert, B. Koskella, M. Poss, and B. M. Sadd, 2013. The origin of specificity by means of natural selection: evolved and nonhost resistance in host–pathogen interactions. Evolution 67:1–9.

Aplin, L. M., D. R. Farine, J. Morand-Ferron, E. F. Cole, A. Cockburn, and B. C. Sheldon, 2013. Individual personalities predict social behaviour in wild networks of great tits (Parus major). Ecol. Lett. 16:1365–1372.

Aplin, L. M., D. R. Farine, J. Morand-Ferron, and B. C. Sheldon, 2012. Social networks predict patch discovery in a wild population of songbirds. Proc. R. Soc. B 279:4199–4205.

Ashby, B. and S. Gupta, 2013. Sexually transmitted infections in polygamous mating systems. Philos. Trans. R. Soc. London. B, Biol. Sci. 368:20120048.

———, 2014. Parasitic castration promotes coevolutionary cycling but also imposes a cost on sex. Evolution 68:2234–2244.

Ashby, B., S. Gupta, and A. Buckling, 2014a. Effects of epistasis on infectivity range during host–parasite coevolution. Evolution 68:2972–2982.

———, 2014b. Spatial structure mitigates fitness costs in host–parasite coevolution. Am. Nat. 183:E64–E74.

van Baalen, M., 1998. Coevolution of recovery ability and virulence. Proc. R. Soc. B 265:317–325.

Bak, P. and C. Chao, 1990. A forest-fire model and some thoughts on turbulence. Phys. Lett. A 147:297–300.

Balenger, S. L. and M. Zuk, 2014. Testing the Hamilton-Zuk Hypothesis: Past, Present, and Future. Integr. Comp. Biol. P. icu059.

Bangham, J., D. J. Obbard, K. W. Kim, P. R. Haddrill, and F. M. Jiggins, 2007. The age and evolution of an antiviral resistance mutation in Drosophila melanogaster. Proc. R. Soc. B 274:2027–2034.

Barton, N. H., 1995. A general model for the evolution of recombination. Genet. Res. 65:123–145.

Baudoin, M., 1975. Host castration as a parasitic strategy. Evolution 29:335–352.

Bearman, P. S., J. Moody, and K. Stovel, 2004. Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks. Am. J. Sociol. 110:44–91.

Beauchamp, G., 2008. What is the magnitude of the group-size effect on vigilance? Behav. Ecol. 19:1361–1368.

Bell, G., 1982. The masterpiece of nature: the evolution and genetics of sexuality. University of California Press, Berkeley, CA.

Benmayor, R., D. J. Hodgson, G. G. Perron, and A. Buckling, 2009. Host mixing and disease emergence. Curr. Biol. 19:764–767.

Best, A., S. Webb, A. White, and M. Boots, 2011. Host resistance and coevolution in spatially structured populations. Proc. R. Soc. B 278:2216–2222.

Best, A., A. White, and M. Boots, 2009. The implications of coevolutionary dynamics to host–parasite interactions. Am. Nat. 173:779–791.

———, 2010a. Resistance is futile but tolerance can explain why parasites do not always castrate their hosts. Evolution 64:348–357.

Best, A., A. White, E. Kisdi, J. Antonovics, and M. A. Brockhurst, 2010b. The Evolution of Host–Parasite Range. Am. Nat. 176:63–71.

Blanchard, J. F., 2002. Populations, pathogens, and epidemic phases: closing the gap between theory and practice in the prevention of sexually transmitted diseases. Sex. Transm. Infect. 78:i183–i188.

Blower, S. M. and J. Roughgarden, 1989. Population dynamics and parasitic castration: test of a model. Am. Nat. 134:848–858.

Bohannan, B. J. M., B. Kerr, C. M. Jessup, J. B. Hughes, and G. Sandvik, 2002. Trade-offs and coexistence in microbial microcosms. Antonie Van Leeuwenhoek 81:107–115.

Bohannan, B. J. M. and R. E. Lenski, 2000. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. Ecol. Lett. 3:362–377.

Boily, M. C., C. Lowndes, and M. Alary, 2002. The impact of HIV epidemic phases on the effectiveness of core group interventions: insights from mathematical models. Sex. Transm. Infect. 78:i78–i90.

Bonds, M. H., D. C. Keenan, A. J. Leidner, and P. Rohani, 2005. Higher disease prevalence can induce greater sociality: a game theoretic coevolutionary model. Evolution 59:1859–1866.

Boomsma, J. J., E. J. Fjerdingstad, and J. Frydenberg, 1999. Multiple paternity, relatedness and genetic diversity in Acromyrmex leaf-cutter ants. Proc. R. Soc. B 266:249–254.

Boots, M., P. J. Hudson, and A. Sasaki, 2004. Large Shifts in Pathogen Virulence Relate to Host Population Structure. Science 303:842–844.

Boots, M. and R. J. Knell, 2002. The evolution of risky behaviour in the presence of a sexually transmitted disease. Proc. R. Soc. B 269:585–589.

Boots, M. and R. Norman, 2000. Sublethal infection and the population dynamics of host–microparasite interactions. J. Anim. Ecol. 69:517–524.

Boots, M. and A. Sasaki, 1999. 'Small worlds' and the evolution of virulence: infection occurs locally and at a distance. Proc. R. Soc. B 266:1933–1938.

Boots, M., A. White, A. Best, and R. Bowers, 2014. How Specificity and Epidemiology Drive the Coevolution of Static Trait Diversity in Hosts and Parasites. Evolution 68:1594–1606.

Borgia, G. and K. Collis, 1990. Parasites and bright male plumage in the satin bowerbird (Ptilonorhynchus violaceus). Am. Zool. 30:279–285.

Boyce, M. S., 1990. The Red Queen Visits Sage Grouse Leks. Am. Zool. 30:263–270.

Brault, A. C., A. M. Powers, D. Ortiz, J. G. Estrada-Franco, R. Navarro-Lopez, and S. C. Weaver, 2004. Venezuelan equine encephalitis emergence: enhanced vector infection from a single amino acid substitution in the envelope glycoprotein. Proc. Natl. Acad. Sci. U. S. A. 101:11344–11349.

Bremermann, H. J. and J. Pickering, 1983. A game-theoretical model of parasite virulence. J. Theor. Biol. 100:411–426.

Brockhurst, M. A., A. Buckling, and P. B. Rainey, 2006. Spatial heterogeneity and the stability of host–parasite coexistence. J. Evol. Biol. 19:374–379.

Brockhurst, M. A., N. Colegrave, and D. E. Rozen, 2010. Next-generation sequencing as a tool to study microbial evolution. Mol. Ecol. 20:972–980.

Brockhurst, M. A., A. D. Morgan, P. B. Rainey, and A. Buckling, 2003. Population mixing accelerates coevolution. Ecol. Lett. 6:975–979.

Brown, J. K. M. and A. Tellier, 2011. Plant-parasite coevolution: bridging the gap between genetics and ecology. Annu. Rev. Phytopathol. 49:345–367.

Buckee, C., L. Danon, and S. Gupta, 2007. Host community structure and the maintenance of pathogen diversity. Proc. R. Soc. B 274:1715–1721.

Buckee, C. O., K. Koelle, M. J. Mustard, and S. Gupta, 2004. The effects of host contact network structure on pathogen diversity and strain structure. Proc. Natl. Acad. Sci. U. S. A. 101:10839–10844.

Buckling, A., R. Craig Maclean, M. A. Brockhurst, and N. Colegrave, 2009. The Beagle in a bottle. Nature 457:824–829.

Buckling, A. and P. B. Rainey, 2002a. Antagonistic coevolution between a bacterium and a bacteriophage. Proc. R. Soc. B 269:931–936.

———, 2002b. The role of parasites in sympatric and allopatric host diversification. Nature 420:496–499.

Buckling, A., Y. Wei, R. C. Massey, M. A. Brockhurst, and M. E. Hochberg, 2006. Antagonistic coevolution with parasites increases the cost of host deleterious mutations. Proc. R. Soc. B 273:45–49.

Burdon, J. J. and P. H. Thrall, 1999. Spatial and Temporal Patterns in Coevolving Plant and Pathogen Associations. Am. Nat. 153:S15–S33.

Campos, P. R. A., P. S. C. A. Neto, V. M. de Oliveira, and I. Gordo, 2008. Environmental heterogeneity enhances clonal interference. Evolution 62:1390–1399.

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

Chao, L., B. R. Levin, and F. M. Stewart, 1977. A complex community in a simple habitat: an experimental study with bacteria and phage. Ecology 58:369–378.

Choe, J. C., 1994. Sexual selection and mating system in Zorotypus gurneyi Choe (Insecta: Zoraptera): I. Dominance Hierarchy and Mating Success. Behav. Ecol. Sociobiol. 34:87–93.

Clarke, B., 1976. The ecological relationships of host-parasite relationships. Pp. 87–103, *in* A. E. R. Taylor and R. Muller, eds. Genetic aspects of host–parasite relationships. Blackwell, Oxford.

Clay, K. and P. X. Kover, 1996. The Red Queen Hypothesis and plant/pathogen interactions. Annu. Rev. Phytopathol. 34:29–50.

Clayton, D. H., 1991. The influence of parasites on host sexual selection. Parasitol. today 7:329–334.

Comins, H. N., M. P. Hassell, and R. M. May, 1992. The Spatial Dynamics of host–Parasitoid Systems. J. Anim. Ecol. 1992:735–748.

Crews, A. E. and T. P. Yoshino, 1989. Schistosoma mansoni: effect of infection on reproduction and gonadal growth in Biomphalaria glabrata. Exp. Parasitol. 68:326–334.

Croft, D. P., R. James, A. J. W. Ward, M. S. Botham, D. Mawdsley, and J. Krause, 2005. Assortative interactions and social networks in fish. Oecologia 143:211–219.

Croft, D. P., J. Krause, S. K. Darden, I. W. Ramnarine, J. J. Faria, and R. James, 2009. Behavioural trait assortment in a social network: patterns and implications. Behav. Ecol. Sociobiol. 63:1495–1503.

Dall, S. R. X., L.-A. Giraldeau, O. Olsson, J. M. McNamara, and D. W. Stephens,

2005. Information and its use by animals in evolutionary ecology. Trends Ecol. Evol. 20:187–193.

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. J. Theor. Biol. 201:1–12.

Danchin, E., L.-A. Giraldeau, T. J. Valone, and R. H. Wagner, 2004. Public information: from nosy neighbors to cultural evolution. Science 305:487–491.

Dass, S. A. H., A. Vasudevan, D. Dutta, L. J. T. Soh, R. M. Sapolsky, and A. Vyas, 2011. Protozoan parasite Toxoplasma gondii manipulates mate choice in rats by enhancing attractiveness of males. PLoS One 6:e27229.

Davies, N. B., 1983. Polyandry, cloaca-pecking and sperm competition in dunnocks. Nature 302:334–336.

De, P., A. E. Singh, T. Wong, W. Yacoub, and A. M. Jolly, 2004. Sexual network analysis of a gonorrhoea outbreak. Sex. Transm. Infect. 80:280–285.

Débarre, F., S. Lion, M. van Baalen, and S. Gandon, 2012. Evolution of host life-history traits in a spatially structured host–parasite system. Am. Nat. 179:52–63.

Decaestecker, E., S. Gaba, J. A. M. Raeymaekers, R. Stoks, L. Van Kerckhoven, D. Ebert, and L. De Meester, 2007. host–parasite 'Red Queen' dynamics archived in pond sediment. Nature 450:870–873.

Dieckmann, U. and R. Law, 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. J. Math. Biol. 34:579–612.

Dixson, A. F., 1998. Primate Sexuality: Comparative Studies of the Prosimians, Monkeys, Apes, and Humans. Oxford University Press, Oxford, UK.

Doebeli, M., 1996. Quantitative genetics and population dynamics. Evolution 50:532–546.

Doligez, B., E. Danchin, and J. Clobert, 2002. Public information and breeding habitat selection in a wild bird population. Science 297:1168–70.

Dubois, F. and F. Cézilly, 2002. Breeding success and mate retention in birds: a meta-analysis. Behav. Ecol. Sociobiol. 52:357–364.

Dybdahl, M. F. and C. M. Lively, 1998. host–parasite coevolution: evidence for rare advantage and time-lagged selection in a natural population. Evolution 52:1057–1066.

Eames, K. T. D., 2007. Contact tracing strategies in heterogeneous populations. Epidemiol. Infect. 135:443–454.

Eames, K. T. D. and M. J. Keeling, 2002. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. Proc. Natl. Acad. Sci. U. S. A. 99:13330–13335.

———, 2003. Contact tracing and disease control. Proc. R. Soc. B 270:2565–2571.

———, 2004. Monogamous networks and the spread of sexually transmitted diseases. Math. Biosci. 189:115–130.

———, 2006. Coexistence and specialization of pathogen strains on contact networks. Am. Nat. 168:230–241.

Ebert, D. and J. J. Bull, 2003. Challenging the trade-off model for the evolution of virulence: is virulence management feasible? Trends Microbiol. 11:15–20.

Ebert, D., H. J. Carius, T. Little, and E. Decaestecker, 2004. The evolution of virulence when parasites cause host castration and gigantism. Am. Nat. 164:S19–S32.

Ebert, D., M. Lipsitch, and K. L. Mangin, 2000. The Effect of Parasites on Host Population Density and Extinction : Experimental Epidemiology with Daphnia and Six Microparasites. Am. Nat. 156:459–477.

Edmunds, W. J., G. Kafatos, J. Wallinga, and J. R. Mossong, 2006. Mixing patterns and the spread of close-contact infectious diseases. Emerg. Themes Epidemiol. 3:10.

Elgar, M. A., 1986. The establishment of foraging flocks in house sparrows: risk of predation and daily temperature. Behav. Ecol. Sociobiol. 19:433–438.

Emlen, S. T. and L. W. Oring, 1977. Ecology, sexual selection, and the evolution of mating systems. Science 197:215–223.

Farine, D. R., 2014. Measuring phenotypic assortment in animal social networks: weighted associations are more robust than binary edges. Anim. Behav. 89:141–153.

Farine, D. R., C. J. Garroway, and B. C. Sheldon, 2012. Social network analysis of mixed-species flocks: exploring the structure and evolution of interspecific social behaviour. Anim. Behav. 84:1271–1277.

Faruque, S. M., K. Biswas, S. M. N. Udden, Q. S. Ahmad, D. A. Sack, G. B. Nair, and J. J. Mekalanos, 2006. Transmissibility of cholera: in vivo-formed biofilms and their relationship to infectivity and persistence in the environment. Proc. Natl. Acad. Sci. U. S. A. 103:6350–6355.

Fenton, A., J. Antonovics, and M. A. Brockhurst, 2009. Inverse-gene-for-gene infection genetics and coevolutionary dynamics. Am. Nat. 174:E230–E242.

———, 2012. Two-Step Infection Processes Can Lead To Coevolution Between Functionally Independent Infection and Resistance Pathways. Evolution 66:2030–2041.

Fenton, A. and M. A. Brockhurst, 2007. Epistatic interactions alter dynamics of multilocus gene-for-gene coevolution. PLoS One 2:e1156.

Flor, H. H., 1956. The complementary genetic systems in flax and flax rust. Adv. Genet. 8:29–54.

Flores, C. O., J. R. Meyer, S. Valverde, L. Farr, and J. S. Weitz, 2011. Statistical structure of host–phage interactions. Proc. Natl. Acad. Sci. U. S. A. 108:E288–E297.

Forde, S. E., R. E. Beardmore, I. Gudelj, S. S. Arkin, J. N. Thompson, and L. D. Hurst, 2008. Understanding the limits to generalizability of experimental evolutionary models. Nature 455:220–223.

Forde, S. E., J. N. Thompson, and B. J. M. Bohannan, 2004. Adaptation varies through space and time in a coevolving host–parasitoid interaction. Nature 431:841–844.

———, 2007. Gene flow reverses an adaptive cline in a coevolving host–parasitoid interaction. Am. Nat. 169:794–801.

Frank, S. A., 1993a. Coevolutionary genetics of plants and pathogens. Evol. Ecol. 7:45–75.

———, 1993b. Evolution of host–parasite diversity. Evolution 47:1721–1732.

———, 1993c. Specificity versus detectable polymorphism in host–parasite genetics. Proc. R. Soc. B 254:191–197.

———, 1994. Recognition and polymorphism in host–parasite genetics. Philos. Trans. R. Soc. London. B, Biol. Sci. 346:283–293.

———, 1996. Problems inferring the specificity of plant-pathogen genetics. Evol. Ecol. 10:323–325.

———, 2000. Specific and non-specific defense against parasitic attack. J. Theor. Biol. 202:283–304.

Freeland, W. J., 1976. Pathogens and the Evolution of Primate Sociality. Biotropica 8:12–24.

Funk, S., M. Salathé, and V. A. A. Jansen, 2010. Modelling the influence of human behaviour on the spread of infectious diseases: a review. J. R. Soc. Interface 7:1247–1256.

Galef, B. G. and L.-A. Giraldeau, 2001. Social influences on foraging in vertebrates: causal mechanisms and adaptive functions. Anim. Behav. 61:3–15.

Gallet, R., Y. Shao, and I.-N. Wang, 2009. High adsorption rate is detrimental to bacteriophage fitness in a biofilm-like environment. BMC Evol. Biol. 9:241.

Gandon, S., 2002. Local adaptation and the geometry of host–parasite coevolution. Ecol. Lett. 5:246–256.

Gandon, S., P. Agnew, and Y. Michalakis, 2002. Coevolution between parasite virulence and host life-history traits. Am. Nat. 160:374–388.

Gandon, S., A. Buckling, E. Decaestecker, and T. Day, 2008. host–parasite coevolution and patterns of adaptation across time and space. J. Evol. Biol. 21:1861–1866.

Gandon, S., Y. Capowiez, Y. Dubois, and Y. Michalakis, 1996. Local adaptation and gene-for-gene coevolution in a metapopulation model. Proc. R. Soc. B 263:1003–1009.

Gandon, S. and T. Day, 2009. Evolutionary epidemiology and the dynamics of adaptation. Evolution 63:826–838.

Gandon, S. and Y. Michalakis, 2002. Local adaptation, evolutionary potential and host–parasite coevolution: interactions between migration, mutation, population size and generation time. J. Evol. Biol. 15:451–462.

Gandon, S. and S. L. Nuismer, 2009. Interactions between genetic drift, gene flow, and selection mosaics drive parasite local adaptation. Am. Nat. 173:212–24.

Gandon, S. and S. P. Otto, 2007. The evolution of sex and recombination in response to abiotic or coevolutionary fluctuations in epistasis. Genetics 175:1835–53.

Garber, P. A., 1997. One for all and breeding for one: cooperation and competition as a tamarin reproductive strategy. Evol. Anthropol. 5:187–199.

Garnett, G. P. and R. M. Anderson, 1996. Sexually transmitted diseases and sexual behavior: insights from mathematical models. J. Infect. Dis. 174:S150–S161.

Ghani, A. C., C. Donnelly, and G. P. Garnett, 1998. Sampling biases and missing data in explorations of sexual partner networks for the spread of sexually transmitted diseases. Stat. Med. 17:2079–2097.

Ghani, A. C. and G. P. Garnett, 1998. Measuring sexual partner networks for transmission of sexually transmitted diseases. J. R. Stat. Soc. Ser. A (Statistics Soc. 161:227–238.

Gillespie, D. T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 93555:2340–2361.

———, 2001. Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. 115:1716–1733.

Gokhale, C. S., A. Papkou, A. Traulsen, and H. Schulenburg, 2013. Lotka-Volterra dynamics kills the Red Queen: population size fluctuations and associated stochasticity dramatically change host–parasite coevolution. BMC Evol. Biol. 13:254.

Gomez, P. and A. Buckling, 2011. Bacteria-Phage Antagonistic Coevolution in Soil. Science 332:106–109.

Gomulkiewicz, R., J. N. Thompson, R. Holt, S. L. Nuismer, and M. E. Hochberg, 2000. Hot Spots, Cold Spots, and the Geographic Mosaic Theory of Coevolution. Am. Nat. 156:156–174.

Gouveia-Oliveira, R. and A. G. Pedersen, 2009. Higher variability in the number of sexual partners in males can contribute to a higher prevalence of sexually transmitted diseases in females. J. Theor. Biol. 261:100–106.

Graves, B. M. and D. Duvall, 1995. Effects of sexually transmitted diseases on heritable variation in sexually selected systems. Anim. Behav. 50:1129–1131.

Grosberg, R. K. and M. W. Hart, 2000. Mate Selection and the Evolution of Highly Polymorphic Self/Nonself Recognition Genes. Science 289:2111–2114.

Gupta, S., R. M. Anderson, and R. M. May, 1989. Networks of sexual contacts: implications for the pattern of spread of HIV. AIDS 3:807–817.

Gururani, M. A., J. Venkatesh, C. P. Upadhyaya, A. Nookaraju, S. K. Pandey, and S. W. Park, 2012. Plant disease resistance genes: Current status and future directions. Physiol. Mol. Plant Pathol. 78:51–65.

Guttridge, T. l., S. H. Gruber, J. D. DiBattista, K. A. Feldheim, D. P. Croft, S. Krause, and J. Krause, 2011. Assortative interactions and leadership in a free-ranging population of juvenile lemon shark Negaprion brevirostris. Mar. Ecol. Prog. Ser. 423:235–245.

Haerter, J. O., A. Trusina, and K. Sneppen, 2011. Targeted Bacterial Immunity Buffers Phage Diversity. J. Virol. 85:10554–10560.

Hall, A. R., P. D. Scanlan, and A. Buckling, 2011a. Bacteria-phage coevolution and the emergence of generalist pathogens. Am. Nat. 177:44–53.

Hall, A. R., P. D. Scanlan, A. D. Morgan, and A. Buckling, 2011b. host–parasite coevolutionary arms races give way to fluctuating selection. Ecol. Lett. 14:635–642.

Hall, M. D. and D. Ebert, 2013. The genetics of infectious disease susceptibility: has the evidence for epistasis been overestimated? BMC Biol. 11:79.

Hamilton, W. D., 1975. Gamblers since life began: barnacles, aphids, elms. Quart Rev Biol 50:175–180.

———, 1980. Sex versus Non-Sex versus Parasite. Oikos 35:282–290.

———, 1990. Mate choice near or far. Am. Zool. 30:341–352.

Hamilton, W. D., R. Axelrod, and R. Tanese, 1990. Sexual reproduction as an adaptation to resist parasites (a review). Proc. Natl. Acad. Sci. U. S. A. 87:3566–3573.

Hamilton, W. D. and M. Zuk, 1982. Heritable true fitness and bright birds: a role for parasites? Science 218:384–387.

Hamilton, W. J. and R. Poulin, 1997. The Hamilton and Zuk hypothesis revisited: a meta-analytical approach. Behaviour 134:299–320.

Haraguchi, Y. and A. Sasaki, 2000. The evolution of parasite virulence and transmission rate in a spatially structured population. J. Theor. Biol. 203:85–96.

Haraldsdottir, S., S. Gupta, and R. M. Anderson, 1992. Preliminary studies of sexual networks in a male homosexual community in Iceland.

Harrison, E., A.-L. Laine, M. Hietala, and M. A. Brockhurst, 2013. Rapidly fluctuating environments constrain coevolutionary arms races by impeding selective sweeps. Proc. R. Soc. B 280:20130937.

Hart, B. L., 1990. Behavioral adaptations to pathogens and parasites: five strategies. Neurosci. Biobehav. Rev. 14:273–294.

Hartvigsen, G. and S. Levin, 1997. Evolution and spatial structure interact to influence plant-herbivore population and community dynamics. Proc. R. Soc. B 264:1677–1685.

Hassell, M. P., R. M. May, S. W. Pacala, and P. L. Chesson, 1991. The persistence of host–parasitoid associations in patchy environments. I. A general criterion. Am. Nat. 138:568–583.

Heath, K. D. and S. L. Nuismer, 2014. Connecting functional and statistical definitions of genotype by genotype interactions in coevolutionary studies. Front. Genet. 5:1–7.

Heilmann, S., K. Sneppen, and S. Krishna, 2010. Sustainability of virulence in a phage-bacterial ecosystem. J. Virol. 84:3016–3022.

———, 2012. Coexistence of phage and bacteria on the boundary of self-organized refuges. Proc. Natl. Acad. Sci. U. S. A. 109:12828–12833.

Hellweger, F. L. and V. Bucci, 2009. A bunch of tiny individualsIndividual-based modeling for microbes. Ecol. Modell. 220:8–22.

Herfst, S., E. J. A. Schrauwen, M. Linster, S. Chutinimitkul, E. de Wit, V. J. Munster, E. M. Sorrell, T. M. Bestebroer, D. F. Burke, D. J. Smith, G. F. Rimmelzwaan, A. D. M. E. Osterhaus, and R. A. M. Fouchier, 2012. Airborne transmission of influenza A/H5N1 virus between ferrets. Science 336:1534–1541.

Hethcote, H. W. and J. A. Yorke, 1984. Gonorrhea: transmission dynamics and control. *in* Lect. Notes Biomath. no. 56. Springer-Verlag, Berlin, Germany.

Hill, A. V., C. E. Allsopp, D. Kwiatkowski, N. M. Anstey, P. Twumasi, P. A. Rowe, S. Bennett, D. Brewster, A. J. McMichael, and B. M. Greenwood, 1991. Common west African HLA antigens are associated with protection from severe malaria. Nature 352:595–600.

Hochberg, M. E. and M. van Baalen, 1998. Antagonistic coevolution over productivity gradients. Am. Nat. 152:620–634.

Hock, K. and N. H. Fefferman, 2012. Social organization patterns can lower disease risk without associated disease avoidance or immunity. Ecol. Complex. 12:34–42.

Hofbauer, J. and K. Sigmund, 1990. Adaptive dynamics and evolutionary stability. Appl. Math. Lett. 3:75–79.

Hoffman, J. I., J. Forcada, P. N. Trathan, and W. Amos, 2007. Female fur seals show active choice for males that are heterozygous and unrelated. Nature 445:912–914.

House, T., 2011. Epidemic prediction and control in clustered populations. J. Theor. Biol. 272:1–7.

Howard, R. S. and C. M. Lively, 1994. Parasitism, mutation accumulation and the maintenance of sex. Nature 367:554–557.

———, 2003. Opposites attract? Mate choice for parasite evasion and the evolutionary stability of sex. J. Evol. Biol. 16:681–689.

Hudson, P. J., A. P. Dobson, and D. Newborn, 1998. Prevention of Population Cycles by Parasite Removal. Science 282:2256–2258.

Hunt, C. W., 1989. Migrant labor and sexually transmitted disease: AIDS in Africa. J. Health Soc. Behav. 30:353–373.

Hyman, J. and J. Li, 1997. Behavior Changes in SIS STD Models with Selective Mixing. SIAM J. Appl. Math. 57:1082–1094.

Jacot, A., H. Scheuber, and M. W. G. Brinkhof, 2004. Costs of an induced immune response on sexual display and longevity in field crickets. Evolution 58:2280–2286.

Jacquez, J. A., C. P. Simon, J. Koopman, L. Sattenspiel, and T. Perry, 1988. Modeling and analyzing HIV transmission: the effect of contact patterns. Math. Biosci. 92:119–199.

Jaenike, J., 1978. An hypothesis to account for the maintenance of sex within populations. Evol. Theory 3:191–194.

Jansen, V. A. A. and M. van Baalen, 2006. Altruism through beard chromodynamics. Nature 440:663–666.

Jennions, M. D. and M. Petrie, 1997. Variation in mate choice and mating preferences: a review of causes and consequences. Biol. Rev. Camb. Philos. Soc. 72:283–327.

———, 2000. Why do females mate multiply? A review of the genetic benefits. Biol. Rev. Camb. Philos. Soc. 75:21–64.

Jonah, G., A. Rainey, A. Natonson, L. F. Maxfield, and J. M. Coffin, 2003. Mechanisms of avian retroviral host range extension. J. Virol. 77:6709.

Jones, J. D. G. and J. L. Dangl, 2006. The plant immune system. Nature 444:323–329.

Kavaliers, M., E. Choleris, A. Agmo, and D. W. Pfaff, 2004. Olfactory-mediated parasite recognition and avoidance: linking genes to behavior. Horm. Behav. 46:272–283.

Kavaliers, M., D. D. Colwell, and E. Choleris, 1999. Parasites and behavior: an ethopharmacological analysis and biomedical implications. Neurosci. Biobehav. Rev. 23:1037–1045.

Keeling, M. J., 1999. The effects of local spatial structure on epidemiological invasions. Proc. R. Soc. B 266:859–867.

———, 2005. The implications of network structure for epidemic dynamics. Theor. Popul. Biol. 67:1–8.

Keeling, M. J. and K. T. D. Eames, 2005. Networks and epidemic models. J. R. Soc. Interface 2:295–307.

Keeling, M. J. and D. A. Rand, 1995. A Spatial Mechanism for the Evolution and Maintenance of Sexual Reproduction. Oikos 74:414–424.

Keeling, M. J., M. E. J. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. Wilesmith, and B. T. Grenfell, 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science 294:813–817.

Kerr, B., C. Neuhauser, B. J. M. Bohannan, and A. M. Dean, 2006. Local migration promotes competitive restraint in a host–pathogen 'tragedy of the commons'. Nature 442:75–78.

King, K. C., L. F. Delph, J. Jokela, and C. M. Lively, 2009. The geographic mosaic of sex and the Red Queen. Curr. Biol. 19:1438–1441.

King, K. C., J. Jokela, and C. M. Lively, 2011. Trematode parasites infect or die in snail hosts. Biol. Lett. 7:265–268.

Kirkpatrick, M. and M. J. Ryan, 1991. The evolution of mating preferences and the paradox of the lek. Nature 350:33–38.

Kisdi, E., 2006. Trade–off geometries and the adaptive dynamics of two coevolving species. Evol. Ecol. Res. 8:33–38.

Knell, R. J., 1999. Sexually transmitted disease and parasite-mediated sexual selection. Evolution 53:957–961.

———, 2004. Syphilis in renaissance Europe: rapid evolution of an introduced sexually transmitted disease? Proc. R. Soc. B 271:S174–S176.

Knell, R. J. and K. M. Webberley, 2004. Sexually transmitted diseases of insects: distribution, evolution, ecology and host behaviour. Biol. Rev. Camb. Philos. Soc. 79:557–581.

Kokko, H., E. Ranta, G. Ruxton, and P. Lundberg, 2002. Sexually transmitted disease and the evolution of mating systems. Evolution 56:1091–1100.

Koskella, B. and M. A. Brockhurst, 2014. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. FEMS Microbiol. Rev. (in press).

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

Koskella, B., J. N. Thompson, G. M. Preston, and A. Buckling, 2011. Local Biotic Environment Shapes the Spatial Scale of Bacteriophage Adaptation to Bacteria. Am. Nat. 177:440–451.

Kouyos, R. D., M. Salathé, and S. Bonhoeffer, 2007. The Red Queen and the persistence of linkage-disequilibrium oscillations in finite and infinite populations. BMC Evol. Biol. 7:211.

Krause, J. and G. D. Ruxton, 2002. Living in groups. Oxford University Press, Oxford, UK.

Labrie, S. J., J. E. Samson, and S. Moineau, 2010. Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. 8:317–327.

Laidre, M. E., 2010. How rugged individualists enable one another to find food and shelter: field experiments with tropical hermit crabs. Proc. R. Soc. B 277:1361–1369.

LeBoeuf, B. J., 1974. Male-male competition and reproductive success in elephant seals. Am. Zool. 14:163–176.

Lenski, R. E., 1984a. Coevolution of bacteria and phage: are there endless cycles of bacterial defenses and phage counterdefenses? J. Theor. Biol. 108:319–325.

———, 1984b. Two-step resistance by Escherichia coli B to bacteriophage T2. Genetics 107:1–7.

Lenski, R. E. and B. R. Levin, 1985. Constraints on the Coevolution of Bacteria and Virulent Phage: A Model , Some Experiments , and Predictions for Natural Communities. Am. Nat. 125:585–602.

Levin, B. R. and J. J. Bull, 2004. Population and evolutionary dynamics of phage therapy. Nat. Rev. Microbiol. 2:166–173.

Levin, D. A., 1975. Pest Pressure and Recombination Systems in Plants. Am. Nat. 109:437–451.

Levin, S. A. and D. Pimentel, 1981. Selection of intermediate rates of increase in parasite-host systems. Am. Nat. 117:308–315.

Levri, E. P., 1995. Parasite-induced change in host behavior of a freshwater snail: parasitic manipulation or byproduct of infection? Behav. Ecol. 10:234–241.

Liljeros, F., C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, 2001. The web of human sexual contacts. Nature 411:907–908.

Lion, S. and M. Boots, 2010. Are parasites "prudent" in space? Ecol. Lett. 13:1245–1255.

Lipsitch, M., E. A. Herre, and M. A. Nowak, 1995. Host population structure and the evolution of virulence: A "law of diminishing returns". Evolution 49:743–748.

Lipsitch, M. and M. A. Nowak, 1995. The evolution of virulence in sexually transmitted HIV/AIDS. J. Theor. Biol. 174:427–440.

Lipsitch, M., S. Siller, and M. A. Nowak, 1996. The evolution of virulence in pathogens with vertical and horizontal transmission. Evolution 50:1729–1741.

Little, T. J., K. Watt, and D. Ebert, 2006. Parasite-host specificity: experimental studies on the basis of parasite adaptation. Evolution 60:31–38.

Lively, C. M., 1987. Evidence from a New Zealand snail for the maintenance of sex by parasitism. Nature 328:519–521.

———, 2006. The ecology of virulence. Ecol. Lett. 9:1089–1095.

———, 2010a. A review of Red Queen models for the persistence of obligate sexual reproduction. J. Hered. 101:S13–S20.

———, 2010b. An epidemiological model of host–parasite coevolution and sex. J. Evol. Biol. 23:1490–1497.

Lively, C. M. and M. F. Dybdahl, 2000. Parasite adaptation to locally common host genotypes. Nature 405:679–681.

Lloyd-Smith, J. O., W. M. Getz, and H. V. Westerhoff, 2004. Frequency-dependent incidence in models of sexually transmitted diseases: portrayal of pair-based transmission and effects of illness on contact behaviour. Proc. R. Soc. B 271:625–634.

Lloyd-Smith, J. O., S. J. Schreiber, P. E. Kopp, and W. M. Getz, 2005. Superspreading and the effect of individual variation on disease emergence. Nature 438:355–339.

Lockhart, A. B., P. H. Thrall, and J. Antonovics, 1996. Sexually transmitted diseases in animals: Ecological and evolutionary implications. Biol. Rev. Camb. Philos. Soc. 71:415–471.

Loehle, C., 1995. Social Barriers to Pathogen Transmission in Wild Animal Populations. Ecology 76:326–335.

———, 1997. The pathogen transmission avoidance theory of sexual selection. Ecol. Modell. 103:231–250.

Lombardo, M. P., 1998. On the evolution of sexually transmitted diseases in birds. J. Avian Biol. 29:314–321.

Lopez-Pascua, L. D. C. and A. Buckling, 2008. Increasing productivity accelerates host–parasite coevolution. J. Evol. Biol. 21:853–860.

Lopez-Pascua, L. D. C., S. Gandon, and A. Buckling, 2012. Abiotic heterogeneity drives parasite local adaptation in coevolving bacteria and phages. J. Evol. Biol. 25:187–195.

Luijckx, P., H. Fienberg, D. Duneau, and D. Ebert, 2013. A matching-allele model explains host resistance to parasites. Curr. Biol. 23:1085–1088.

Luong, L. T. and H. K. Kaya, 2005. Sexually transmitted parasites and host mating behavior in the decorated cricket. Behav. Ecol. 16:794–799.

Lusseau, D., B. Wilson, P. S. Hammond, K. Grellier, J. W. Durban, K. M. Parsons, T. R. Barton, and P. M. Thompson, 2006. Quantifying the influence of sociality on population structure in bottlenose dolphins. J. Anim. Ecol. 75:14–24.

Matz, C., D. McDougald, A. M. Moreno, P. Y. Yung, F. H. Yildiz, and S. Kjelleberg, 2005. Biofilm formation and phenotypic variation enhance predation-driven persistence of Vibrio cholerae. Proc. Natl. Acad. Sci. U. S. A. 102:16819–16824.

May, R. M., 1973. On relationships among various types of population models. Am. Nat. 107:46–57.

May, R. M. and R. M. Anderson, 1978. Regulation and Stability of host–Parasite Population Interactions: II. Destabilizing Processes. J. Anim. Ecol. 47:249–267.

———, 1979. Population biology of infectious diseases: Part II. Nature 280:455–461.

———, 1983. Epidemiology and genetics in the coevolution of parasites and hosts. Proc. R. Soc. B 219:281–313.

May, R. M., S. Gupta, and A. R. McLean, 2001. Infectious disease dynamics: What characterizes a successful invader? Philos. Trans. R. Soc. London. B, Biol. Sci. 356:901–910.

Maynard Smith, J., 1978. The evolution of sex. Cambridge University Press, Cambridge, UK.

Messinger, S. M. and A. Ostling, 2009. The consequences of spatial structure for the evolution of pathogen transmission rate and virulence. Am. Nat. 174:441–454.

Metz, A. L., L. Hatcher, and J. A. Newman, 1985. Venereal pox in breeder turkeys in Minnesota. Avian Dis. 29:850–853.

Meyer, J. R., D. T. Dobias, J. S. Weitz, J. E. Barrick, R. T. Quick, and R. E. Lenski, 2012. Repeatability and contingency in the evolution of a key innovation in phage lambda. Science 335:428–432.

Mink, G. I., 1993. Pollen and seed-transmitted viruses and viroids. Annu. Rev. Phytopathol. 31:375–402.

Mitchell, M., M. D. Thomure, and N. L. Williams, 2006. The role of space in the success of coevolutionary learning. *in* Artif. Life X Proc. Tenth Int. Conf. Simul. Synth. Living Syst., Pp. 118–124.

Mizoguchi, K., M. Morita, C. R. Fischer, M. Yoichi, Y. Tanji, and H. Unno, 2003. Coevolution of bacteriophage PP01 and Escherichia coli O157: H7 in continuous culture. Appl. Environ. Microbiol. 69:170–176.

Moller, A. P., 1987. House sparrow, Passer domesticus, communal displays. Anim. Behav. 35:203–210.

Morgan, A. D., M. A. Brockhurst, L. D. C. Lopez-Pascua, C. Pal, and A. Buckling, 2007. Differential impact of simultaneous migration on coevolving hosts and parasites. BMC Evol. Biol. 7:1.

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

Morran, L. T., O. G. Schmidt, I. A. Gelarden, R. C. Parrish, and C. M. Lively, 2011. Running with the Red Queen: host–Parasite Coevolution Selects for Biparental Sex. Science 333:216–218.

Morris, M. and M. Kretzschmar, 1997. Concurrent partnerships and the spread of HIV. AIDS 11:641–648.

Moshkin, M., N. Litvinova, E. A. Litvinova, A. Bedareva, A. Lutsyuk, and L. Gerlinskaya, 2012. Scent recognition of infected status in humans. J. Sex. Med. 9:3211–3218.

Mossong, J., N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W. J. Edmunds, 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 5:e74.

Nakamura, M., 1990. Cloacal protuberance and copulatory behavior of the alpine accentor (Prunella collaris). Auk 107:284–295.

Nowak, M. A. and R. M. May, 1994. Superinfection and the evolution of parasite virulence. Proc. R. Soc. B 255:81–89.

Nuismer, S. L., S. P. Otto, and F. Blanquart, 2008. When do host–parasite interactions drive the evolution of non-random mating? Ecol. Lett. 11:937–946.

Nunn, C. L., 2003. Behavioural defences against sexually transmitted diseases in primates. Anim. Behav. 66:37–48.

O'Keefe, K. J., 2005. The evolution of virulence in pathogens with frequency-dependent transmission. J. Theor. Biol. 233:55–64.

O'Keefe, K. J. and J. Antonovics, 2002. Playing by different rules: the evolution of virulence in sterilizing pathogens. Am. Nat. 159:597–605.

Otto, S. P., 2003. The advantages of segregation and the evolution of sex. Genetics 164:1099–1118.

Otto, S. P. and S. L. Nuismer, 2004. Species interactions and the evolution of sex. Science 304:1018–1020.

Parejo, D., J. White, J. Clobert, A. Dreiss, and E. Danchin, 2007. Blue Tits use fledgling quantity and quality as public information in breeding site choice. Ecology 88:2373–2382.

Parker, M. A., 1994. Pathogens and sex in plants. Evol. Ecol. 8:560–584.

Pascual, M., 1993. Diffusion-Induced Chaos in a Spatial Predator–Prey System. Proc. R. Soc. B 251:1–7.

Paterson, S., T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M. Quail, F. Smith, D. Walker, B. Libberton, A. Fenton, N. Hall, and M. A. Brockhurst, 2010. Antagonistic coevolution accelerates molecular evolution. Nature 464:275–278.

Penman, B. S., B. Ashby, C. O. Buckee, and S. Gupta, 2013. Pathogen selection drives nonoverlapping associations between HLA loci. Proc. Natl. Acad. Sci. U. S. A. 110:19645–19650.

Penn, D. and W. Potts, 1998. Chemical signals and parasite-mediated sexual selection. Trends Ecol. Evol. 5347:391–396.

Penn, D. J. and W. K. Potts, 1999. The Evolution of Mating Preferences and Major Histocompatibility Complex Genes. Am. Nat. 153:145–164.

Peters, A. D. and C. M. Lively, 1999. The Red Queen and fluctuating epistasis: a population genetic analysis of antagonistic coevolution. Am. Nat. 154:393–405.

Pike, T. W., M. Samanta, J. Lindström, and N. J. Royle, 2008. Behavioural phenotype affects social interactions in an animal network. Proc. R. Soc. B 275:2515–2520.

Potterat, J. J., L. Phillips-Plummer, S. Q. Muth, R. B. Rothenberg, D. E. Woodhouse, T. S. Maldonado-Long, H. P. Zimmerman, and J. B. Muth, 2002. Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs. Sex. Transm. Infect. 78:i159–i163.

Poullain, V., S. Gandon, M. A. Brockhurst, and A. Buckling, 2008. The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage. Evolution 62:1–11.

Poullain, V. and S. L. Nuismer, 2012. Infection genetics and the likelihood of host shifts in coevolving host–parasite interactions. Am. Nat. 180:618–628.

Pöysä, H., 1992. Group foraging in patchy environments: the importance of coarse-level local enhancement. Ornis Scand. 23:159–166.

Prado, F., A. Sheih, J. D. West, and B. Kerr, 2009. Coevolutionary cycling of host sociality and pathogen virulence in contact networks. J. Theor. Biol. 261:561–569.

Qvarnström, A. and E. Forsgren, 1998. Should females prefer dominant males? Trends Ecol. Evol. 13:498–501.

Rand, D. A., M. J. Keeling, and H. B. Wilson, 1995. Invasion, stability and evolution to criticality in spatially extended, artificial host–pathogen ecologies. Proc. R. Soc. B 259:55–63.

Read, J. M. and M. J. Keeling, 2003. Disease evolution on networks: the role of contact structure. Proc. R. Soc. B 270:699–708.

Reader, S. M., J. R. Kendal, and K. N. Laland, 2003. Social learning of foraging sites and escape routes in wild Trinidadian guppies. Anim. Behav. 66:729–739.

Reid, J., P. Arcese, A. E. V. Cassidy, A. Marr, J. M. Smith, and L. Keller, 2005. Hamilton and Zuk meet heterozygosity? Song repertoire size indicates inbreeding and immunity in song sparrows (Melospiza melodia). Proc. R. Soc. B 272:481–487.

Rey, F., O. Schwartz, and S. Wain-Hobson, 2013. Gain-of-function research: unknown risks. Science 342:311.

Rhodes, C. J. and R. M. Anderson, 1996. Persistence and dynamics in lattice models of epidemic spread. J. Theor. Biol. 180:125–133.

Rohani, P., D. J. D. Earn, and B. T. Grenfell, 1999. Opposite Patterns of Synchrony in Sympatric Disease Metapopulations. Science 286:968–971.

de Roode, J. C. and T. Lefevre, 2012. Behavioral Immunity in Insects. Insects 3:789–820.

Russell, C. A., J. M. Fonville, A. E. Brown, D. F. Burke, D. L. Smith, S. L. James, S. Herfst, S. van Boheemen, M. Linster, E. J. Schrauwen, L. Katzelnick, A. Mosterín, T. Kuiken, E. Maher, G. Neumann, A. D. Osterhaus, Y. Kawaoka,

R. A. Fouchier, and D. J. Smith, 2012. The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. Science 336:1541–1547.

Sarasa, M., E. Serrano, R. C. Soriguer, J.-E. Granados, P. Fandos, G. Gonzalez, J. Joachim, and J. M. Pérez, 2011. Negative effect of the arthropod parasite, Sarcoptes scabiei, on testes mass in Iberian ibex, Capra pyrenaica. Vet. Parasitol. 175:306–312.

Sasaki, A., 2000. host–parasite coevolution in a multilocus gene-for-gene system. Proc. R. Soc. B 267:2183–2188.

Scanlan, P. D., A. R. Hall, L. D. C. Lopez-Pascua, and A. Buckling, 2011. Genetic basis of infectivity evolution in a bacteriophage. Mol. Ecol. 25:1–9.

Schulte, R. D., C. Makus, B. Hasert, N. K. Michiels, and H. Schulenburg, 2010. Multiple reciprocal adaptations and rapid genetic change upon experimental co-evolution of an animal host and its microbial parasite. Proc. Natl. Acad. Sci. U. S. A. 107:7359–7364.

Shannon, C. E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27:379–423.

Sheldon, B. C., 1993. Sexually transmitted disease in birds: occurrence and evolutionary significance. Philos. Trans. R. Soc. London. B, Biol. Sci. 339:491–497.

Sherman, P. T., 1995. Social organization of cooperatively polyandrous white-winged trumpeters (Psophia leucoptera). Auk 112:296–309.

Sherman, P. W., T. D. Seeley, and H. K. Reeve, 1988. Parasites, pathogens, and polyandry in social Hymenoptera. Am. Nat. 131:602–610.

Simms, E. L. and J. Triplett, 1994. Costs and benefits of plant responses to disease: resistance and tolerance. Evolution 48:1973–1985.

Smith, D. J., A. S. Lapedes, J. C. de Jong, T. M. Bestebroer, G. F. Rimmelzwaan, A. D. M. E. Osterhaus, and R. A. M. Fouchier, 2004. Mapping the antigenic and genetic evolution of influenza virus. Science 305:371–376.

Smith, M. J., A. White, J. A. Sherratt, S. Telfer, M. Begon, and X. Lambin, 2008. Disease effects on reproduction can cause population cycles in seasonal environments. J. Anim. Ecol. 77:378–389.

Soper, D. M., L. F. Delph, and C. M. Lively, 2012. Multiple paternity in the freshwater snail, Potamopyrgus antipodarum. Ecol. Evol. 2:3179–3185.

Suttle, C. A., 2007. Marine viruses - major players in the global ecosystem. Nat. Rev. Microbiol. 5:801–812.

Tait, K., L. C. Skillman, and I. W. Sutherland, 2002. The efficacy of bacteriophage as a method of biofilm eradication. Biofouling 18:305–311.

Tanner, C. J. and A. L. Jackson, 2012. Social structure emerges via the interaction between local ecology and individual behaviour. J. Anim. Ecol. 81:260–277.

Thompson, J. N., 1994. The Coevolutionary Process. University of Chicago Press, Chicago, IL.

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

Thrall, P. H., J. Antonovics, and J. D. Bever, 1997. Sexual transmission of disease and host mating systems: within-season reproductive success. Am. Nat. 149:485–506.

Thrall, P. H., J. Antonovics, and A. P. Dobson, 2000. Sexually transmitted diseases in polygynous mating systems: prevalence and impact on reproductive success. Proc. R. Soc. B 267:1555–1563.

Thrall, P. H. and J. J. Burdon, 2002. Local Adaptation in the Linum Marginale-Melampsora Lini host–Pathogen Interaction. Evolution 56:1340–1351.

———, 2003. Evolution of virulence in a plant host–pathogen metapopulation. Science 299:1735–1737.

Tybur, J. M. and S. W. Gangestad, 2011. Mate preferences and infectious disease: theoretical considerations and evidence in humans. Philos. Trans. R. Soc. London. B, Biol. Sci. 366:3375–3388.

Valone, T. J., 2007. From eavesdropping on performance to copying the behavior of others: a review of public information use. Behav. Ecol. Sociobiol. 62:1–14.

Van Valen, L., 1973. A new evolutionary law. Evol. Theory 1:1–30.

Vanderwaal, K. L., E. R. Atwill, L. A. Isbell, and B. McCowan, 2014. Linking social and pathogen transmission networks using microbial genetics in giraffe (Giraffa camelopardalis). J. Anim. Ecol. 83:406–414.

Vogwill, T., A. Fenton, and M. A. Brockhurst, 2008. The impact of parasite dispersal on antagonistic host–parasite coevolution. J. Evol. Biol. 21:1252–1258.

Vogwill, T., A. Fenton, A. Buckling, M. E. Hochberg, and M. A. Brockhurst, 2009. Source populations act as coevolutionary pacemakers in experimental selection mosaics containing hotspots and coldspots. Am. Nat. 173:E171–176.

Volz, E. and L. A. Meyers, 2007. Susceptible-infected-recovered epidemics in dynamic contact networks. Proc. R. Soc. B 274:2925–2933.

Vos, M., P. J. Birkett, E. Birch, R. I. Griffiths, and A. Buckling, 2009. Local adaptation of bacteriophages to their bacterial hosts in soil. Science 325:833.

Watters, J. and A. Sih, 2005. The mix matters: behavioural types and group dynamics in water striders. Behaviour 142:1417–1431.

Watts, C. H. and R. M. May, 1992. The influence of concurrent partnerships on the dynamics of HIV/AIDS. Math. Biosci. 108:89–104.

Webberley, K. M., G. D. Hurst, J. Buszko, and M. E. Majerus, 2002. Lack of parasite-mediated sexual selection in a ladybird/sexually transmitted disease system. Anim. Behav. 63:131–141.

Weber, N., S. P. Carter, S. R. X. Dall, R. J. Delahay, J. L. McDonald, S. Bearhop, and R. A. McDonald, 2013. Badger social networks correlate with tuberculosis infection. Curr. Biol. 23:R915–R916.

Webster, J. P., J. I. Hoffman, and M. Berdoy, 2003. Parasite infection, host resistance and mate choice: battle of the genders in a simultaneous hermaphrodite. Proc. R. Soc. B 270:1481–1485.

Webster, J. P. and M. E. J. Woolhouse, 1999. Cost of resistance: relationship between reduced fertility and increased resistance in a snail-schistosome host–parasite system. Proc. R. Soc. B 266:391–396.

Weigler, B. J., A. A. Girjes, N. A. White, N. D. Kunst, F. N. Carrick, and M. F. Lavin, 1988. Aspects of the epidemiology of Chlamydia psittaci infection in a population of koalas (Phascolarctos cinereus) in southeastern Queensland, Australia. J. Wildl. Dis. 24:282–291.

Weitz, J. S., H. Hartman, and S. A. Levin, 2005. Coevolutionary arms races between bacteria and bacteriophage. Proc. Natl. Acad. Sci. U. S. A. 102:9535–9540.

West, S. A., C. M. Lively, and A. F. Read, 1999. A pluralist approach to sex and recombination. Science 12:1003–1012.

White, K. A. J. and B. T. Grenfell, 1997. Regulation of complex host dynamics by a macroparasite. J. Theor. Biol. 186:81–91.

Wilfert, L. and P. Schmid-Hempel, 2008. The genetic architecture of susceptibility to parasites. BMC Evol. Biol. 8:187.

Wolinska, J. and K. C. King, 2009. Environment can alter selection in host–parasite interactions. Trends Parasitol. 25:236–244.

Woodhouse, D. E., R. B. Rothenberg, J. J. Potterat, W. W. Darrow, S. Q. Muth, A. S. Klovdahl, H. P. Zimmerman, H. L. Rogers, T. S. Maldonado, J. B. Muth, and J. U. Reynolds, 1994. Mapping a social network of heterosexuals at high risk for HIV infection. AIDS 8:1331–1336.

Woolhouse, M. E. J., C. Dye, J. F. Etard, T. Smith, J. D. Charlwood, G. P. Garnett, P. Hagan, J. L. K. Hii, P. D. Ndhlovu, R. J. Quinnell, C. H. Watts, S. K. Chandiwana, and R. M. Anderson, 1997. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. Proc. Natl. Acad. Sci. U. S. A. 94:338–342.

Woolhouse, M. E. J. and S. Gowtage-Sequeria, 2005. Host range and emerging and reemerging pathogens. Emerg. Infect. Dis. 11:1842–1847.

Wylie, J. L. and A. Jolly, 2001. Patterns of chlamydia and gonorrhea infection in sexual networks in Manitoba, Canada. Sex. Transm. Dis. 28:14–24.

Zaman, L., S. Devangam, and C. Ofria, 2011. Rapid host–parasite coevolution drives the production and maintenance of diversity in digital organisms. *in* Proc. 13th Annu. Conf. Genet. Evol. Comput., Pp. 219–226. ACM.

Zhang, Q.-G. and A. Buckling, 2011. Antagonistic coevolution limits population persistence of a virus in a thermally deteriorating environment. Ecol. Lett. 14:282–288.

Zumla, A., M. Raviglione, R. Hafner, and C. F. von Reyn, 2013. Tuberculosis. N. Engl. J. Med. 368:745–755.