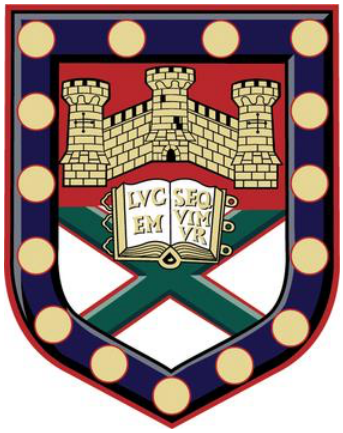
The diversity and distribution of multihost viruses in bumblebees

Submitted by David John Pascall, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Biological Sciences
In September 2017

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature……



General Abstract

The bumblebees (genus *Bombus*) are an ecologically and economically important group in decline. Their decline is driven by many factors, but parasites are believed to play a role. This thesis examines the factors that influence the diversity and distribution of multihost viruses in bumblebees using molecular and modelling techniques. In Chapter 2, I performed viral discovery to isolate new multihost viruses in bumblebees. I investigated factors that explain prevalence differences between different host species using co-phylogenetic models. I found that related hosts are infected with similar viral assemblages, related viruses infect similar host assemblages and related hosts are on average infected with related viruses. Chapter 3 investigated the ecology of four of the novel viruses in greater detail. I applied a multivariate probit regression to investigate the abiotic factors that may drive infection. I found that precipitation may have a positive or negative effect depending on the virus. Also, we observe a strong non-random association between two of the viruses. The novel viruses have considerably more diversity than the previously known viruses. Chapter 4 investigated the effect of pesticides on viral and non-viral infection. I exposed *Bombus terrestris* colonies to field realistic doses of the neoticotinoid pesticide clothianidin in the laboratory, to the mimic pulsed exposure of crop blooms. I found some evidence for a positive effect of uncertain size on the infection rate of pesticide exposed colonies relative to non-pesticide exposed colonies, a potentially important result. Chapter 5 explored the evolution of avirulent multihost digital organisms across fluctuating fitness landscapes within a discrete sequence space. Consistent with theory, I found that evolution across a fluctuating discrete landscape leads to a faster rate of adaptation, greater diversity and greater specialism or generalism, depending on the correlation between the landscapes. A large range of factors are found to be important in the distribution of infection and diversity of viruses, and we find evidence for abiotic, biotic and anthropogenic factors all playing a role.

Contents

List of Tables

List of Figures

Authors' Declaration

All work in this thesis was performed by the author, except the following:

Chapter 1:

Table 1.1 is adapted from the table of British bumblebees produced by the Natural History Museum.

Chapter 2:

All bees were caught by Lena Wilfert, Dave Goulson and Matthew Tinsley. Pooled extractions and sequencing were performed by Lena Wilfert. The 2.3.2 section of the methods was written and the work within was performed by Darren Obbard, the text was subsequently edited and expanded by the author. Figure 2.2 was generated by Darren Obbard. Lena Wilfert, Darren Obbard and Matthew Tinsley all suggested changes to the text.

Chapter 3:

All bees were caught by Lena Wilfert, Dave Goulson and Matthew Tinsley. RNA extractions and sequencing were performed by Lena Wilfert. Individual reverse transcription reactions, and the testing for and sequencing of ABPV and SBPV were performed by Joe Faulks. Sam Braine aided the author with the testing for other viruses. Lena Wilfert suggested changes to the text.

Chapter 4:

Meri Anderson and Thomas Marceau helped with the design and creation of the bespoke colony shelters, and performed all the collections after the bees were placed into the field. Meri Anderson tested the samples for *Nosema ceranae*, *Nosema bombi*, *Nosema apis* and *Crithidia bombi*. Lena Wilfert suggested changes to the text.

Chapter 5:

Lena Wilfert suggested changes to the text.

Acknowledgements

Thanks to Jarrod Hadfield for modifications to MCMCglmm to allow the analyses for Chapters 2 and 4 to be run, as well as for helpful comments.

Thanks to Beth for somehow having the fortitude to read through all my terrible drafts and give hugs when required.

Thanks to Juliet for helpful comments at many points.

And, thanks to Lena for endless patience, no matter how much I messed up over the 4 years and for somehow finding time to look at everything as I scrabbled towards the finish line.

Chapter 1 - General introduction

1.1 The causes and consequences of infection in bumblebees

The bumblebees, genus *Bombus*, are important global pollinators. Animals, and particularly bees, play a role in the pollination of plants that provide 35% of total global food production (Klein et al. 2007). The existence of these pollinators provides what economists term an ecosystem service. For this reason, we have good economic incentives for attempting to understand the diseases that afflict them. This section of the introduction will provide background both to the species and the viruses that infect them.

1.1.1 Relevant biology of the genus *Bombus*

The considerations of bumblebee disease cannot be understood without reference to their biology and ecology. Bees, like wasps and ants, are hymenoptera. Specifically, the bees (and spheciform wasps) lie within the Apoidea, sister clade to the Formicidae, the ants (Johnson et al. 2013). Like the majority of the rest of the hymenoptera, their sex determination is mediated through the haplodiploid system (Goulson 2010). In this system, unfertilised eggs develop into males, whereas fertilised eggs develop into females. Thus, the males are haploid and the females are diploid, producing the name. This has important effects on the relatedness of individuals. Assuming the mother has mated only once, a female is more closely related to her sisters than she is to her own children, a fact that has been suggested to have enabled the evolution of eusociality through kin selection (Hamilton 1972) in the hymenoptera. Most bumblebee species are indeed eusocial, the exception to this being those within the subgenus *Psithyrus*, which, being social parasites of other bumblebee nests, have no caste system (Plath 1922). The bumblebee colony is highly ordered, with a queen who lays eggs, which are then cared for by the female worker caste. Reproductives, the drones and young queens, tend to be produced later in the season when the colony size has increased, with nests led by queens who did not experience diapause having their sexuals emerging significantly later (Beekman and Van Stratum 2000). Different species of bumblebee exhibit incomplete temporal separation throughout the year, causing some degree of community partitioning even when they are spatially sympatric, with, for example, *Bombus pratorum* beginning to produce workers around a month before *Bombus lapidarius* (Goodwin 1995).

Estimates of the number of bumblebee species present in the United Kingdom vary, based on whether the species is suspected to be locally extinct. Table 1.1 shows the 27 species described in Britain, and marks the 26 believed to be extant. A comprehensive bumblebee phylogeny was generated in 2007 (Cameron, Hines, and Williams 2007), which showed that some of the previous subgenera were paraphyletic. In an attempt to solve this problem, a new subgeneric system was proposed by Williams et al. (Williams et al. 2008). In this system, the genus *Bombus* is broken into groups of which 9 have extant British examples.

TABLE 1.1 has been removed by the author of this thesis/dissertation for copyright reasons.

The nest is provisioned by foraging workers who gather pollen and nectar from flowers in the area around the nest. *Bombus terrestris* foragers have been estimated to travel as far as 1.5-2 km on their foraging journeys (Westphal, Steffan-Dewenter, and Tscharntke 2006), but lower estimates than this are the norm (Darvill, Knight, and Goulson 2004; Osborne et al. 1999, 2008; Walther-Hellwig and Frankl 2000; Wolf and Moritz 2008). Foragers of different bumblebee species exhibit preferences for different flower species. It has been long known that flower choice by bumblebee foragers within a foraging flight is not random (Thomson 1981). Beyond this, considerable interspecific differences in plant species utilisation are commonly observed (e.g. Arbulo, Santos, Salvarrey & Invernizzi 2011; Goulson and Darvill 2004; Goulson, Lye, and Darvill 2008; Harder 1985), but this is not a universal phenomenon (Lye et al. 2010), and the degree of overlap may depend on the diversity of flowers currently in bloom. Studies have shown that flower choice is correlated with species tongue length (Goulson et al. 2008; Harder 1985), which implicitly incorporates shared behavioural characteristics between closely related bee species as there is phylogenetic correlation between tongue length and relatedness (Harmon-Threatt and Ackerly 2013). Bumblebees are known to have innate aesthetic preference both for symmetry (Møller 1995) and surfaces reflecting predominantly ultraviolet, blue or green wavelengths (Lunau and

Maier 1995), but these innate preferences can be overridden by socially learned behaviours (Worden and Papaj 2005).

1.1.2 The epidemiology and diversity of the viruses of bees

Mass bee deaths have been intermittently recorded since at least the 10th century, when ‘a mortality of bees’ was noted in the *Annála Uladh* (Fleming 1871). This recording has led to the discovery of a large number of bee pathogens and parasites and since the 1960s, a large number of viruses. Most viruses of bees were originally described in the honeybee, as historically, that species has been the focus of study. Table 1.2 below summarises the viral pathogens of bees.

Table 1.2: The viral pathogens of bees. Information from a wide variety of sources, references in the table

Pathogen	Acronym	Classification	Multi-host?	Known to infect bumblebees?	Notes
Deformed wing virus - type A	DWV-A	Iflavirus (Oers 2010)	Y (Zhang et al. 2012)	Y (Genersch et al. 2006; Singh et al. 2010)	Was considered a Japanese strain of EBV, but then renamed DWV-A. Reviewed in de Miranda et al. (2010)
Kakugo virus	KV	Iflavirus (Oers 2010)	Y (Fujiyuki et al. 2006)	?	Discovered by Fujiyuki et al. (Fujiyuki et al. 2004)
Sacbrood virus	SBV	Iflavirus (Oers 2010)	Y (Levitt et al. 2013)	Y (Levitt et al. 2013)	Positively identified by Bailey et al. (1964), but may have previously been discovered by Brcak et al. (1963)
Deformed wing virus - type B (*Varroa destructor virus-1*)	DWV-B (VDV-1)	Iflavirus (Oers 2010)	Y (replicates in *V. destructor*)	Y (Schoonvaere et al. 2016)	Discovered by Ongus et al. (2004). Has been repeatedly shown to recombine with DWV-A (Moore et al. 2011; Wang et al. 2013; Zioni, Soroker, and Chejanovsky 2011).
Israeli acute paralysis virus	IAPV	Dicistroviridae (Maori et al. 2007)	Y (Levitt et al. 2013)	Y (Levitt et al. 2013)	Characterised by Maori et al. (2007)
Black queen cell virus	BQCV	Dicistroviridae (Bonning and Miller 2010)	Y (Zhang et al. 2012)	Y (Peng et al. 2011)	Discovered by Bailey and Woods (1977)
Kashmir bee virus	KBV	Dicistroviridae (Bonning and Miller 2010)	Y (Singh et al. 2010)	Y (Singh et al. 2010)	Discovered by Bailey and Woods (Bailey and Woods 1977)
Acute bee paralysis virus	ABPV	Dicistroviridae (Bonning and Miller 2010)	Y (Bailey and Gibbs 1964)	Y (Bailey and Gibbs 1964)	Discovered by Bailey et al. (1963)
Slow bee paralysis virus	SBPV	Iflavirus (Oers 2010)	Y (replicates in *V. destructor*) (de Miranda and Genersch 2010)	Y (McMahon et al. 2015)	Discovered by Bailey and Woods (1974)

Chronic bee paralysis virus	CBPV	Unassigned (+ve RNA) (Olivier et al. 2008)	Y (Yang et al. 2013)	?	Discovered by Bailey et al. (1963)
Lake Sinai virus-1	LSV-1	Unassigned (+ve RNA) with homology to Nodaviruses (Runckel et al. 2011)	?	?	Discovered by Runckel et al. (2011)
Lake Sinai virus-2	LSV-2	Unassigned (+ve RNA) with homology to Nodaviruses (Runckel et al. 2011)	?	?	Discovered by (2011)
Lake Sinai virus-3	LSV-3	Unassigned (+ve RNA) with homology to Nodaviruses (Runckel et al. 2011)	?	?	Discovered by Cornman et al. (2012)
Lake Sinai virus-4	LSV-4	Unassigned (+ve RNA) with homology to Nodaviruses (Runckel et al. 2011)	?	?	Discovered by Ravoet et al. (2013)
Apis mellifera filamentous virus	*Am*FV	Putative Nudivirus (Bailey, Carpenter, and Woods 1981; Wang and Jehle 2009)	?	?	Discovered by Clark (1978)
Arkansas bee virus	ABV	?	?	?	Discovered by Bailey and Woods (1974)
Aphid lethal paralysis virus - strain Bookings	ALPV	Dicistroviridae (Van Munster et al. 2002)	Y	?	Found in bees by Runckel et al. (2011)
Big Sioux River virus	BSRV	Dicistroviridae (Runckel et al. 2011)	?	?	Discovered by Runckel et al. (2011)
Berkeley bee picornavirus	BBPV	?	?	?	Discovered by Lommel et al., (1985)
Bee Virus X	BVX	Possibly similar to Nudaurelia *β* virus (Bailey, Carpenter, et al. 1980), so putative member of the Tetraviridae	?	?	Discovered by Bailey and Woods (Bailey and Woods 1974)
Egypt bee virus	EBV	Iflavirus (DWV is a strain of this) (de Miranda and Genersch 2010)	?	?	Discovered by Bailey et al. (Bailey, Carpenter, and Woods 1979)
Thai sacbrood virus	TSBV	Iflavirus (as a strain sacbrood virus) (Bailey, Carpenter, and Woods 1982)	?	?	Discovered by Bailey et al. (Bailey et al. 1982)
Chronic bee paralysis satellite virus	CBPSV	Unassigned (+ve RNA) (Ball, Overton, and Buck 1985; Olivier et al. 2008)	?	?	Discovered by Bailey et al. (Bailey, Ball, et al. 1980)

Cloudy wing virus	CWV	?	?	?	Discovered by Bailey et al. (1980)
Bee virus Y	BVY	Possibly similar to Nudaurelia *β* virus (Bailey, Carpenter, et al. 1980) so putative member of the Tetraviridae	?	?	Discovered by Bailey et al. (1980)
Apis iridescent virus	AIV/ IIV-24	Iridovirus (Bailey, Ball, and Woods 1976)	?	?	Discovered by Bailey et al. (Bailey et al. 1976)
Cricket paralysis virus	$CrPV_{BEE}$	Dicistroviridae (Oers 2010)	Y	?	Detected in *A. mellifera* by Anderson and Gibbs (Anderson and Gibbs 1988)
Unnamed Entomopoxvirina e	-	New Entomopoxvirina e Family (Clark 1982)	?	Y	Discovered by Clark (Clark 1982)
Bee macula-like virus	BeeMLV	Tymoviridae (de Miranda et al. 2015)	Y	Y (Parmentier et al. 2016)	Discovered by de Miranda et al. (de Miranda et al. 2015)
Moku virus	MV	Iflavirus (Mordecai et al. 2016)	Y	?	Discovered by Mordecai et al. (Mordecai et al. 2016)
DWV-C	DWV-C	Iflavirus (Mordecai et al. 2015)	Y	?	Discovered by Mordecai et al. (Mordecai et al. 2015)
Scaldis River bee virus	SRBV	Chuvirus (Schoonvaere et al. 2016)	Y	?	Discovered by Schoonvaere et al. (Schoonvaere et al. 2016)
Ganda bee virus	GBV	Bunyaviridae (Schoonvaere et al. 2016)	Y	Y	Discovered by Schoonvaere et al. (Schoonvaere et al. 2016)
Apis mellifera rhabdovirus 1	ARV-1	Rhabdoviridae (Remnant et al. 2017)	Y	?	Discovered by Remnant et al. (Remnant et al. 2017)
Apis mellifera rhabdovirus 2	ARV-2	Rhabdoviridae (Remnant et al. 2017)	Y	?	Discovered by Remnant et al. (Remnant et al. 2017)
Apis mellifera bunyavirus 1	ABV-1	Bunyaviridae (Remnant et al. 2017)	Y	?	Discovered by Remnant et al. (Remnant et al. 2017)
Apis mellifera flavivirus	AFV	Flaviviridae (Remnant et al. 2017)	Y	?	Discovered by Remnant et al. (Remnant et al. 2017)
Apis mellifera bunyavirus 2	ABV-2	Bunyaviridae (Remnant et al. 2017)	Y	?	Discovered by Remnant et al. (Remnant et al. 2017)
Apis mellifera dicistrovirus	ADV	Dicistroviridae (Remnant et al. 2017)	Y	?	Discovered by Remnant et al. (Remnant et al. 2017)
Apis mellifera Nora virus	ANV	Picornaviridae (Remnant et al. 2017)	Y	?	Discovered by Remnant et al. (2017)
Tobacco ringspot virus	TRSV	Secoviridae (Roberts 1988)	Y	?	Found in bees by Li et al. (2014)

There are currently 42 reported bee-associated viruses. There is also evidence that bumblebees have been historically infected with plant viruses, as there are sequences with high homology to plant viruses integrated into the *B. terrestris* genome (Cui and Holmes 2012), but it is unclear as to whether these still actively spread in bumblebee populations. *Apis mellifera*, the honeybee, is however known to be a viable host for Tobacco ringspot virus (Li et al. 2014).

As Table 1.2 indicates, many known bee viruses were initially discovered in honeybees, but have since been identified as multihost parasites, also infecting bumblebees and a wide variety of other sympatric insect species (Manley, Boots, and Wilfert 2015). This section of the introduction will explore some of unique issues that relate to the multihost pathogens of bees. The majority of the viruses of bees are multihost viruses, having been reported in a wide variety of pollinator species. Any single parasite will spend different amounts of time in different host species, but all multihost parasites spend some degree of time in multiple hosts. A rich theoretical literature has developed around the factors that lead to persistence and high levels of infection in these systems. The ability of a parasite to maintain infection in a host species can be described by its basic reproductive ratio, or R_0. Precisely, the R_0 is the measure of the expected number of secondary infections in from a single infected individual in a completely naïve population (Anderson and May 1992) If the R_0 of a parasite in a host species is greater than 1, that parasite will spread and an outbreak will occur. If the R_0 is less than 1, the infection will be lost from the population, potentially after a long stochastic stuttering transmission chain. The R_0s of parasites in the different species give a useful way of classifying multihost parasites (Fenton et al. 2015). If the R_0 of a parasite in any one host species is greater than 1, then the multihost parasite can be considered a facultative multihost parasite, with the other hosts being unnecessary for the transmission of that parasite. If the R_0 in no species exceeds 1 but the total R_0 in the community exceeds 1, then the parasite is an obligate multihost parasite and persistence would not be possible if hosts were lost from the community. Cases where there is a host where the R_0 is well below 1, but in other hosts the R_0 is high represent spillover hosts for the parasite.

A reasonable definition of multihost parasite is therefore simply a parasite which exhibits has a positive R_0 in multiple species, even if this R_0 is close to 0 in some of those hosts. This definition has the advantage of being conceptually simple, but is practically difficult to assess, especially in understudied or uncommon hosts, where the rare transmission driven by those minor hosts may be undetectable. Given this practical issue, we use a working definition of a parasite detectable at prevalences greater than 5% in multiple species this thesis, with the caveats that some of the hosts infected at high levels may not provide any onward transmission. The question of whether or not many of these bumblebee species represent “true” hosts of the viruses here, and thus “true" multihost parasites would require further experimentation.

The best studied multihost bee virus is Deformed wing virus (DWV), a virus predominantly of honeybees, but that also infects bumblebees (McMahon et al. 2015). DWV exists as three master variants (Mordecai et al. 2015). The transmission modes of DWV have been intensely studied, and like many pathogens (Antonovics et al. 2017), multiple modes have been recorded. A major driver of transmission (and pathology) in European honeybees is the virus’ vectoring by the mite, *Varroa destructor*, with *V. destructor* mediated transmission leading to systemic infection (Ryabov et al. 2014). *V. destructor* vectoring is only known to occur at high frequencies in *Apis mellifera,* as *Apis cerana* the Asian honeybee, exhibits much higher resistance to *V. destructor* infestation (Peng et al. 1987), and *V. destructor* does not parasitise bumblebees. DWV is also horizontally transmitted. de Miranda and Fries (de Miranda and Fries 2008) conclusively showed that artificial insemination with sperm from DWV infected males can cause infection in previously uninfected virgin females, with the virus actively replicating within the ovaries, and possibly the spermathecal tissues. Infected wild caught drones have a large variance in the viral titre of DWV in their semen, but drones with a high sperm titre do not appear to suffer adverse effects relative to those without (Yañez et al. 2012), and likely cases of natural transmission by mating have been recorded (Amiri, Meixner, and Kryger 2016). Artificial insemination experiments using both DWV-positive and DWV-negative semen have demonstrated that vertical transmission can occur if either the queen or male is infected with the virus (Yue et al. 2007). Infection can also occur experimentally on consumption of contaminated food,

but infection generally remains localised to the gut in these cases (Möckel, Gisder, and Genersch 2011).

Many of the dynamics of bee viruses in bumblebees are believed to represent spillover dynamics. Furst and McMahon showed that the distribution of DWV in bumblebees is best explained by the higher prevalence of the virus in sympatric honeybees (Fürst et al. 2014; McMahon et al. 2015), potentially exhibiting both superabundance and superinfection superspreading dynamics (Streicker, Fenton, and Pedersen 2013).

While the example of DWV provides a good illustration of how a generic bee virus could transmit between nest-mates when it enters a bumblebee colony, given that bumblebees freely defecate inside the nest (Free 1955), it does not explain inter-colony transmission. Bee viruses have the added challenge of ensuring that between-colony transmission can occur. It is generally thought that the majority of cross-colony transmission takes place at flowers, however the direct evidence for this is limited to two papers. In the first, *Bombus terrestris* workers were shown to become infected with *Crithidia bombi,* a trypansomatid bumblebee parasite, at experimentally contaminated flowers (Durrer and Schmid-Hempel 1994) and in the second, *B. terrestris* workers were shown to be able to vector *A. mellifera* parasites between flowers and *vice versa* (Graystock, Goulson, and Hughes 2015). Viral particles on collected pollen are known to often be infective (Mazzei et al. 2014; Singh et al. 2010). Taken together, these results indirectly imply the possibility of cross-species transmission at flowers.

Other possibilities for cross-colony transmission in bumblebees exist. Sexual transmission, as described above, could potentially allow enable infection between colonies, though this may be more important in allowing diseases to persist across seasons than for transmission between colonies within a season, as a source of infection for the overwintering queens (Schmid-Hempel 1998). Another possibility is the cuckoo bumblebees in the sub-genus *Psithyrus*. As inquiline parasites, they move into existing nests (Franks 1987), providing an obvious route for contamination, especially due to the propensity for viruses of *Bombus* species to infect multiple hosts (Manley et al. 2015). If a *Psithyrus*

female is infected with a parasite, her invasion will expose the nest to that parasite. This would be a particular risk to the nest of the host species if *Psithyrus* individuals are systematically exposed to different parasites to their hosts, through, for instance, differing preferences for diapause locations. Despite the fact that a single *Psithyrus* species (*Bombus vestalis*) may parasitise up to 35% of nests *B. terrestris* in some areas (Erler and Lattorff 2010), there is a general paucity of knowledge on the biology of species in the sub-genus *Psithyrus* and almost nothing is known about their viral status.

1.1.3 The impacts of infection in bumblebees

The large-scale impacts of viral infection in wild pollinators such as bumblebees is unclear. Over the last century, there have been consistent declines in both the abundance and range of wild bumblebees in Europe, North America and Asia (Williams and Osborne 2009). A wide range of factors are implicated in these declines (e.g. land use changes (Williams, Araújo, and Rasmont 2007) and pesticides (Bryden et al. 2013)), but pathogens are also believed to play a role (Brown et al. 2016).

Our understanding of the pathology of bee viruses in bumblebees remains limited, as most studies investigating the effect of infection are performed on honeybees. But in honeybees, many bee viruses cause no obvious symptoms in standard infections (e.g. SBPV (Bailey and Ball 1991), DWV (Benaets et al. 2017), ABPV (Bailey, Ball, et al. 1980)). However, at least in the case of DWV, even in covert infections, there are increases in mortality, in both honeybees (Benaets et al. 2017) and bumblebees (Graystock et al. 2016). Similarly, in many cases, pathology is found to be context-dependent. Manley et al. (2017) showed that, in *B. terrestris* workers, SBPV only causes increased mortality under starvation, which is likely to be a proxy for general resource limitation. DWV, likewise, generally only causes the classical deformed wing phenotype, when injected directly into the bee bypassing the gut (Möckel et al. 2011), as occurs in honeybees during *V. destructor* mediated transmission. This may explain the lack of a common deformed wing phenotype in infected bumblebees, even though infection is relatively common (Fürst et al. 2014) and the phenotype is known to be able to occur (Genersch et al. 2006). Bee viruses also interact with pesticides, another important stressor, and these effects are

generally synergistic in nature. For instance, Doublet et al. (2015) found that Black queen cell virus, a virus that infects honeybees and bumblebees (Peng et al. 2011), interacts synergistically with the pesticide thioclopid when they are provided simultaneously, leading to an increase in mortality beyond what would be expected from both treatments individually.

Both the lack of specific studies on the pathology of viruses in bumblebees, and the context dependence of pathology in honeybees makes it difficult to determine the overall impact of bumblebee viruses on their currently declines.

1.2 Aims of the thesis

The aim of this thesis is to explore the underlying factors that explain the distribution of infection in wild bumblebees and examine how that relates to the diversity both the species and genetic level of the viral parasites infecting them.

The distribution of infection in bumblebees is affected by within-species and between-species factors. Within a species, hosts differ in their susceptibility to infectious agents. This can be caused by genetic factors, either in the host (Whitehorn et al. 2011) or the parasite (Nagata et al. 2006). It can also be caused by environmental and ecological factors, such as malnutrition (Schaible and Kaufmann 2007), exposure to poisonous chemicals (Di Prisco et al. 2013) and exposure to a different infectious agent (Telfer et al. 2010). Pesticides are known to have immunological effects in bees (e.g. Pettis et al. 2012; Di Prisco et al. 2013), so they may have developed an important role in determining the individuals in a population that become infected. Chapter 4 therefore set out to quantify this risk through experimental manipulation, by exposing *B. terrestris* bumblebees to pesticides and monitoring their subsequent uptake of pathogens from the environment. The aim of this chapter was to attempt confirm the presence or absence of a field realistic dose of commonly used pesticides on the distribution of infection in the population. We used an experimental procedure based on the one Whitehorn et al. (2012) used in the study of the ecological endpoints of pesticide exposure on bumblebee colonies.

A different set of factors would be expected to determine the distribution of infection between hosts of different species. Hosts differ in systematic ways, having different ecologies and underlying genetics. The genetic relatedness between hosts is known to have an impact on parasite sharing between host species (de Vienne, Hood, and Giraud 2009) though mechanisms exist for the emergence of pathogens into highly divergent hosts (Araujo et al. 2015). The aim of Chapter 2 was therefore to quantify the range of viruses infecting bumblebee communities, and determine the effect that the evolutionary history of hosts and parasites has on the distribution of infection in a natural multi-host multi-parasite system. We aimed to perform this analysis using the co-phylogenetic methods as developed by Hadfield et al. (2014) and a suite of novel viruses found as part of the study. Novel viruses were used both due the inherent biological interest in better understanding of the viral biodiversity of an economically important species, and the fact that most viruses described in bees were originally described in honeybees. If the known honeybee viruses other than DWV represent spillover infections into bumblebees (Fürst et al. 2014), then the pattern observed may be biased, by the non-detection of existing “true” bumblebee viruses. The co-phylogenetic methods used jointly determine the amount of variation explained by two phylogenies simultaneously in a system, and allow partitioning variation explained by both the hosts evolutionary history and the pathogens evolutionary history and their interactions.

Abiotic factors are also known to influence disease risk and generally come in two forms. Firstly, factors that change disease risk by directly modifying the susceptibility of the host, such as temperature increasing crop susceptibility to a fungal pathogen (Sharma, Duveiller, and Ortiz-Ferrara 2007). And secondly, those that interact with the transmission route of the pathogen, a classic example being the epidemics of waterborne pathogens that occur after flooding caused by heavy rain (Baqir et al. 2012). No abiotic predictors of the risk of bumblebee infection are currently known. As abiotic factors are important in the distribution of infection in other systems, an aim of Chapter 3 was to perform exploratory research to see if infection with a series of viruses correlates strongly with any reasonable abiotic predictor in a large cross-host sample, after accounting for the host species. Simultaneously, we sought to explore the

diversity of the viruses infecting the bumblebees and observe whether diversity was similar or different across infecting viruses.

It is increasingly understood that different hosts provide very different environments for multihost parasites (Duffy, Turner, and Burch 2006; Remold, Rambaut, and Turner 2008), and that measurable incongruities between the fitness landscapes in different hosts exist (Cervera, Lalić and Elena 2016). The aim of Chapter 5 was to investigate the effects of these incongruities on the evolution of digital multihost parasites and determine whether, in concordance with theory and prior simulations, evolution across multiple hosts increased diversity (Kassen 2002), and the rate of adaptation (Cheetham 1993).

Chapter 2 - Evolution of host range: The prevalence of novel bumblebee viruses is explained by both host and pathogen phylogenies

David J. Pascall containing additional work from Matthew Tinsley, Darren J. Obbard and Lena Wilfert

Note: This chapter features a box containing an aside, it is labelled 2.2.1

2.1 Abstract

The question of what drives the association of an infectious disease with one host species but not with another is one of the most important in disease ecology. Both the ecology of the host and the evolutionary histories of the hosts and pathogens have previously been implicated. Here we use transcriptome sequencing of 13 species of wild caught bumblebee to discover a new set of viruses, and to quantify the impact of phylogeny and range overlap on the associations between viruses and their pollinator hosts. We present 18 novel bumblebee-associated viruses, with the possible existence of 19 more, and use strand-specific PCR and small RNA sequencing to demonstrate that they form active infections in wild bumblebees. Using a phylogenetic mixed model approach, we show that the evolutionary histories of both the host and virus have impacted the distribution of infection of 15 of these novel viruses, as well as three viruses previously described in honeybees, with related hosts sharing viral assemblages, related viruses sharing host assemblages and related hosts being infected to similar degrees by related viruses.

2.2 Introduction

2.2.1 Introduction Part 1

Pathogens that naturally infect more than one host species have a particularly high risk of disease emergence (Woolhouse and Gowtage-Sequeria 2005). One particularly important group of pathogens are the viruses, whose ubiquity leads to them having a disproportionate role in the regulation of natural populations, especially in the oceans (Suttle 2007). The determinants of the strength of an association between a virus and a potential host species are numerous and complex, shaped by genetic and environmental factors. Infection is a function both of the rate of contact between the virus and the host, and the probability of infection on contact. The contact rate between the virus and the host is primarily determined by ecological and behavioural factors. Conversely, the probability of infection on contact is a function of the host immune system, the viral infectivity in that host species, and the interaction between their cellular and reproductive machineries.

At the broadest level, a multihost parasite can arise in two ways: a parasite infecting a single host may gain the ability to infect other host species or, alternatively, infecting multiple host species is the ancestral state, and is inherited as a trait in the daughter species. These different origins would be expected to lead to detectably different phylogenetic signals of shared history, see Figure 1. Dissimilarity between hosts increases with evolutionary distance, and the probability of a successful infection in a novel host is dependent on the similarity between the potential new host and the current host (Longdon et al. 2011; Perlman and Jaenike 2003). If multihost parasites are predominantly generated by viruses gaining the ability to infect new host species, then related host species should share virus assemblages, as host switch events should occur more often between closely related species (Longdon et al. 2014). Contrastingly, if infecting multiple hosts were the ancestral state, it would lead to related groups of viruses having shared host assemblages. Both ancestral infection of multiple host species followed by host range expansion to related host species, and host-virus cospeciation, should lead to a statistical coevolutionary interaction where related hosts share related viruses. As well as interactions between the hosts and parasites, both realised general susceptibility and realised general infectivity (see box) have been observed to vary systematically across host and parasite phylogenies (Hadfield et al. 2014; Longdon et al. 2011; Waxman et al. 2014) and a growing empirical literature indicates that the evolutionary history of hosts and their pathogens may be important in explaining how these associations vary between related species (Hadfield et al. 2014; Hafner and Nadler 1990; Liu et al. 2016; Longdon et al. 2011, 2015; Waxman et al. 2014).

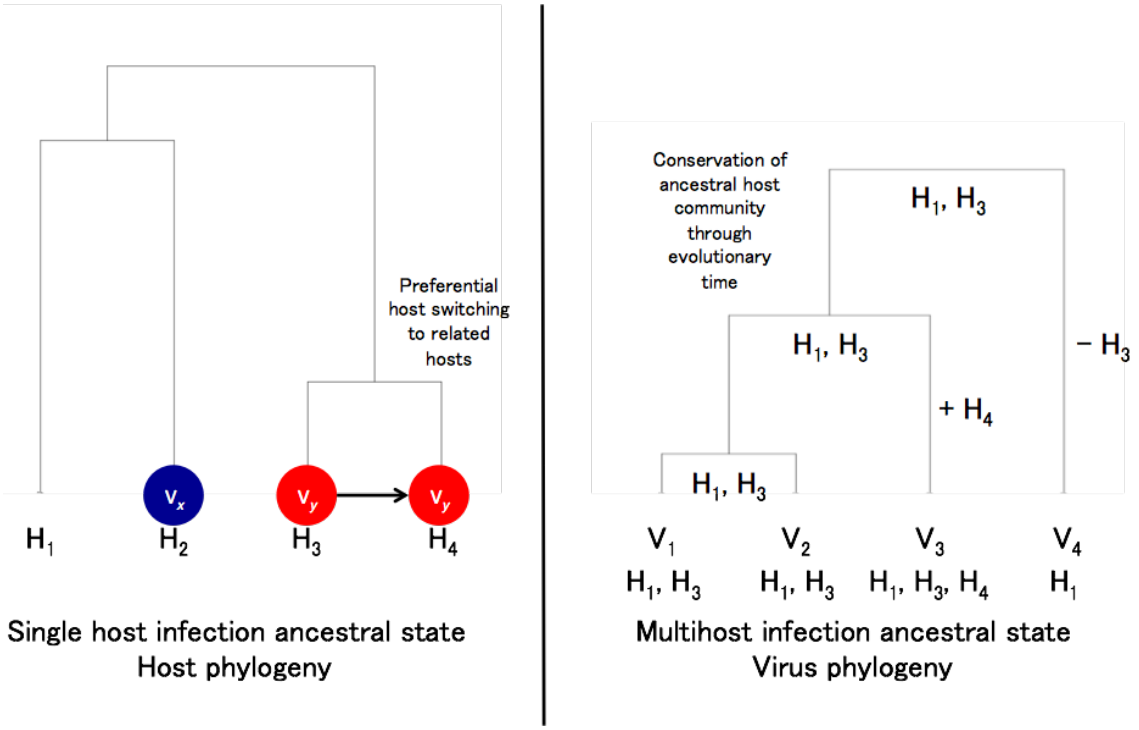


Figure 2.1 Mechanisms for the emergence of novel multihost viruses. The generation of novel multihost viruses through host switching (left panel) leads to a host evolutionary assemblage effect, as the consistent host switching of viruses to hosts near their focal host will causes to related hosts having correlated viral assemblages. The generation of novel multihost viruses through speciation (right panel) of can lead to a viral assemblage effect through the inheritance of the ancestral host range, leading to the daughter species having correlated host assemblages.

The prevalence of pathogens such as viruses across host species is, in principal, structured in two levels. Firstly, the virus may be present or entirely absent in a potential host species, and secondly, if a virus does infect a species, a potentially different set of factors may then influence how prevalent a pathogen is within that species. While this distinction is ultimately often difficult to make in empirical studies, given the uncertainty in distinguishing between a true absence and a rare infection in most field and experimental studies, considering both levels allows us to consider the potentially different factors driving these infection patterns and allows us to go beyond studying infectivity as a mere binary trait.

At the binary level, a complete lack of infection in nature can occur firstly if an a host and virus exist in allopatry, preventing transmission irrespective of the host's susceptibility. Secondly, a physiological mismatch between a host and a virus can prevent infection. For example, Feline panleukopenia virus (FPV) is incapable of infecting dog cells due to the inability of FPV to bind to the canine transferrin receptors, a form of resistance that is removed if the feline transferrin

receptors are expressed on the dog cells in cell culture (Hueffer et al. 2003). Neither mechanism represents an immutable barrier. Spatially separated hosts and parasites may come into contact through migrations or human facilitated invasions, allowing new associations to emerge. For example, the arrival of *Plasmodium relictum,* a causative agent of the previously absent avian malaria, to the Hawaiian islands, led to population declines and potentially contributed to extinctions in the naturally susceptible but naïve populations (van Riper et al. 1986). Equally, incompatibility can break down if evolution in the pathogen or host result removes the physiological barriers to infection. For instance, Canine parvovirus type 2 (CPV2) emerged as a novel disease from FPV when it evolved to the ability to bind to the canine transferrin receptor (Hueffer et al. 2003).

At the quantitative level, differences in prevalence between species can be driven by variation in transmission rates, e.g. through varying direct and indirect contact rates and through physiological variation in viral replication. Examples of this include the propensity for group living (Johnson et al. 2011), population densities (Arneberg et al. 1998), the biodiversity of the community (Civitello et al. 2015) and host avoidance behaviours (Curtis 2014). Variation in prevalence among host species can also be driven by physiological factors, with hosts having varying suitability for the replication of a given parasite, through differential intracellular processes and anatomical factors. For instance, the low rate of human H5N1 avian influenza A virus infection is thought to be due to the cells with the strongest binding affinities being localised deeper in the respiratory tract than those bearing receptors for the common human influenza strains, leading to inefficient aerosolisation (Shinya et al. 2006). In the extreme case, a host species may exhibit only condition dependent susceptibility, where under normal conditions, with a fully functioning immune system, that host species would only very rarely become infected. However, when the immune system is suppressed to some degree, either directly, through an immunosuppressant disease or chemical agent, or indirectly, through trade-offs in resource allocation brought about by malnutrition, infection can occur with higher frequency (Chandra 1983). In humans, for example, individuals suffering from AIDS experience bacteremic and fungeremic episodes at around a 300 times higher rate than the background population (Meyer, Skinhøj, and Prag

1994) and the causative agents are compositionally different to those infecting HIV-negative individuals (Bouza and Rodriguez-Créixems 1999).

2.2.2 Box 1

As historical events, such as spatial segregation, can determine the presence and strength of a host-parasite association, we here define the general relationship between a host's degree of infection averaged over all the pathogens that infect it as the realised general susceptibility (RGS). Realised general susceptibility is observed in field surveys, whereas the true general physiological susceptibility of a host can only be estimated by exposing a host species to a random selection of pathogens under controlled lab conditions. RGS is a function of both the physiology and natural history of a host. A virus may also be considered to have a realised general infectivity (RGI), analogous to a host's realised general susceptibility.

In a phylogenetic regression, either RGS or RGI can be correlated between host or viral species across an estimated phylogeny to test for statistical associations driven by relatedness. These estimated correlations are termed phylogenetic effects. In the case of RGS, if there were a large phylogenetic effect, related hosts would share similar average susceptibilities. Phylogenetic regressions can also be used to analyse association data using two phylogenies, allowing the estimation of phylogenetic effects for RGS and RGI simultaneously.

These techniques were extended by Hadfield et al. (2014) to allow more complicated phylogenetic models, including interactions between phylogenies, to be fit. An evolutionary assemblage effect, termed an evolutionary interaction by Hadfield et al., occurs when the structure of the assemblage of interaction partners is correlated between related species. In the host-virus case, this would represent either related hosts sharing association strengths for sets of viruses (a host evolutionary assemblage effect) or related viruses share association strengths for sets of hosts (a virus evolutionary assemblage effect). A co-evolutionary interaction occurs when related species share interaction strengths for related interaction partners. In the host-virus case, this would represent related pathogens sharing association strengths for related hosts.

2.2.3 Introduction Part 2

We tested for the role of evolutionary interactions in shaping differences in infection level within a set of species from an ecologically and economically important group, the bumblebees. We performed viral discovery by RNA-seq, finding a suite of new viruses and then tested 11 species of bumblebee, wild caught from sites across Scotland, for a subset of these novel viruses, as well as three previously reported honeybee viruses: Slow bee paralysis virus (Bailey and Woods 1974), Acute bee paralysis virus (Bailey et al. 1963) and Hubei partiti-like virus 34 (Cornman et al. 2012; Shi et al. 2016). We analysed the prevalence using co-phylogenetic models to determine the presence or absence and relative strengths of the evolutionary signals expected to shape the host/parasite assemblage in this system.

2.3 Methods

2.3.1 Sampling Regime and Post Collection Treatment

A total of 926 individual bumblebees of 13 species were collected on the wing from 9 sites across Scotland in 2009 and 2011, and frozen in liquid nitrogen or at -80°C. In 2009, we sampled the Ochil Hills, Glenmore, Dalwhinnie, Stirling, Iona, Staffa, and the Pentlands were sampled and in 2011 we sampled Edinburgh and Gorebridge. Precise locations of the sites are available in Chapter 2 Appendix Table A2.1. All individuals were cut in half longitudinally. The cryptic species complex of *Bombus terrestris*, *Bombus lucorum, Bombus cryptarum* and *Bombus magnus* was resolved using RFLP analysis following Murray et al. (2008). One half of each bee was used in grouped RNA extractions of 2-11 individuals per species (median 10; see Chapter 2 Appendix Table A2.2 for pool composition). Two of these groups (DIV and P11) were included in the RNAseq, but excluded from prevalence testing, as the DIV group contained multiple species and the P11 group contained bees that were also in other *B. pascuorum* pools. The groups of bees were ground in liquid nitrogen and added to TRIzol reagent (Life Technologies) for RNA extractions following the manufacturers' standard protocol. The RNA concentrations in the pooled samples were equalized to 200 ng/ul/individual based on Nanodrop measurements.

2.3.2 RNA Sequencing and Bioinformatics

The RNA was combined by species for *B. terrestris* (239 individuals), *Bombus pascuorum* (212 individuals), *B. lucorum* (182 individuals) and other *Bombus* (293 individuals) into four RNA pools. Pools were sequenced using the Illumina HiSeq platform with 100bp paired end reads (Beijing Genomics Institute) after poly-A selection to enrich for polyadenylated mRNAs, and positive sense single stranded RNA (+ssRNA) virus genomes. The single-species bumblebee pools were also re-sequenced following duplex specific nuclease normalization, to enrich for rare transcripts and reduce rRNA representation while retaining non-polyadenylated viruses and virus products. The small RNAs of the same RNA pools of *B. terrestris*, *B. lucorum* and *B. pascuorum* were also sequenced to test for the replication of viruses identified via the transcriptome sequencing.

For each pool, paired end RNAseq data were mapped to the published *Bombus terrestris* and *B. impatiens* genomes using bowtie2 (Langmead and Salzberg 2012) to reduce the representation of conserved bee sequences. Read pairs that did not map concordantly, including divergent bee sequences and other associated microbiota, were assembled *de novo* using Trinity 2.2.0 (Grabherr et al. 2011) as paired end libraries, following automated trimming ('--trimmomatic') and digital read normalisation ('--normalize_reads'). Where two RNAseq libraries (Poly-A and DSN) had been sequenced, these were combined for assembly.

To identify putative viruses, all long open reading frames from each contig were identified and concatenated to provide a 'bait' sequence for similarity searches using Diamond (Buchfink, Xie, and Huson 2015) and BLASTp (Altschul et al. 1990). Contigs shorter than 500 base pairs were discarded. These contig translations were used to search against a Diamond database comprising all of the virus protein sequences available in NCBI database 'nr', and all of the Dipteran, Hymenopteran, Nematode, Fungal, Protist, and prokaryotic proteins available in NCBI database 'refseq_protein' (mode 'blastp'; e-value 0.001; maximum of one match). Matches to phage and short matches to large DNA viruses were excluded. Remaining contigs were manually curated to identify and annotate high-confidence virus-like sequences. To quantify approximate

fold-coverage, and to assess viRNA properties, the raw RNAseq and trimmed small RNA reads were mapped against the putative viral contigs using bowtie2's '--very-sensitive' setting and retaining only the top map (Langmead and Salzberg 2012), and reads per kilobase of transcript per million mapped reads (RPKMs) were generated from this output. Following Fauquet and Stanley (2005), we defined contigs exhibiting greater than 10% difference in nucleotide identity as separate viruses and those exhibiting less than 10% difference as strains of known viruses.

2.3.3 PCR Validation and Testing

To select contigs for further validation, the putative virus sequences were manually filtered on two conditions: the presence of mapping reads in the bumblebee small RNAs (for the *B. terrestris*, *B. pascuorum* and *B. lucorum* pools) or the transcriptomic RNAs (for the other *Bombus* pool where small RNAs were not generated), and the closest blast match being viral RNA-dependent RNA-polymerase. This approach will necessarily exclude some viral contigs, but as RNA viruses contain a single RdRp, this should ensure that a single contig per novel virus was generated and unconnected contigs from the same virus were not retained. Internal primers for these contigs were generated using primer3 (Untergasser et al. 2012) and amplification of the target was verified via Sanger sequencing. See Chapter 2 Appendix Table A2.3 for PCR conditions. Mayfield virus 1 and 2 were Sanger sequence validated over the entirety of the contig. Loch Morlich and River Liunaeg viruses both had further contigs linked to their initial contig by between contig PCRs, and the complete region was Sanger sequenced. All groups excepting DIV and P11 were then tested for the presence of validated viruses via PCR. Black Hill virus was excluded from further analysis after post-hoc sequencing found that the PCR also amplified a host sequence of equal length to the viral target.

2.3.4 Phylogenetic Analysis

We inferred a host phylogeny using cytochrome oxidase I, elongation factor 1-alpha, opsin, phosphoenolpyruvate carboxykinase, the mitochondrial 16S rRNA gene and arginine kinase genes, based on the bumblebee phylogeny reported in Cameron et al. (2007). Where possible the sequences used were the same

as those used by Cameron et al. with the only new additions being for species not included in their initial phylogeny. To break up long branches and as outgroups, additional species were added. When a specific gene could not be acquired for a species, the gene was entered as missing data after alignment had been performed on the other sequences. The sequences taken from genbank with their accession numbers are shown in Chapter 2 Appendix Table A2.4. The DNA sequences were aligned with MAFFT (Katoh et al. 2002; Katoh and Standley 2013) using the default settings. The alignments for each region were then submitted to the GUIDANCE2 server (Sela et al. 2015), and any columns with a score of less than 1 were removed to reduce the effects of alignment uncertainty on the downstream analysis. The 6 gene alignments were then used to generate the phylogeny in BEAST v2.4.5 (Bouckaert et al. 2014), treating each file as a separate partition, using bModelTest (Bouckaert and Drummond 2015) and the "transitionTransversion split" setting. Partitions were allowed to vary in mutation rate and a lognormal relaxed clock shared between partitions was used to model rate variation across the tree (Drummond et al. 2006). Strong priors were placed on the topology of the tree to represent the strong prior information we have about the branching order of the Hymenoptera and Diptera. The bumblebees in the tree were constrained to be monophyletic, then a second monophylly prior was placed around the bumblebee-honeybee (Apinae) group, above this, a third monophylly prior was then placed around the Apinae members and the other members of the hymenoptera (*Allantus luctifer* and *Philotrypesis pilosa*). The dipterans (*Drosophila melanogaster*, *Anopheles dirus*, *Episyrphus balteatus* and *Episyrphus pertinax*) and lepidopterans (*Bombyx mori* and *Heliconius melpomene*) were both separately given monophylly priors. These priors follow evidence from the phylogenies of Hines (2008), Hedtke et al. (2013) and Wiegmann et al. (2009). This allowed the branching within the bumblebees, within the flies, between *Allantus luctifer* and *Philotrypesis pilosa* and the position of the root between the Hymenoptera, Diptera and Lepidoptera to vary. Four separate runs of the phylogeny were performed for 550,000,000 generations with the first 50,000,000 generations being discarded as burn in. Convergence was assessed in Tracer v1.6 (Rambaut et al. 2013). The posterior distribution of trees was then downsampled to 1000 trees.

For the viral phylogeny, amino acid sequences were predicted based on the translated ORFs for regions predicted to contain RdRp motifs using the GenomeNet MOTIF search function (Kanehisa et al. 2002) against the Pfam database (Finn et al. 2014) with an expectation cut-off of 0.00001. If a virus had no annotated motifs, the canonical GDD RdRp amino acid motif (Kamer and Argos 1984) was searched for manually. The sequences were then cut to a trailing region either side of this motif. Additional viral species were added to the phylogeny to anchor species with incomplete sequence and to break up long branches. See Chapter 2 Appendix Table A2.5 for genbank accession numbers of the additional viruses used. Given the long evolutionary distance between the viruses, PROMALS3D (Pei, Kim, and Grishin 2008) was used to align viral sequences. This is a multistage alignment tool that first aligns closely related amino acid sequences then performs structural prediction and additionally uses the structural information to align more divergent amino acid chains. Two of the novel viruses (Agassiz Rock virus and Cnoc Mor virus) were not included in this phylogeny because they lacked the section of the RdRp gene required. Whether the negative sense RNA viruses are phylogenetically related to the positive sense and double stranded RNA viruses or represent a separate origin is currently under scientific debate (reviewed in Koonin et al. 2015). We therefore generated the phylogeny twice, with and without the negative sense RNA viruses (see Chapter 2 Appendix Table A2.5). The trees exist to quantify expected variance (under a Brownian motion model of evolution) between closely related viruses. While the deep splits in the phylogeny are likely to be poorly resolved, due to the fast evolutionary rates of RNA viruses and the considerable time since divergence (if indeed a common ancestor existed), this should not overly bias the conclusions as beyond a certain evolutionary distance, the viruses would be expected to become essentially uncorrelated. The alignment including all the RNA viruses was trimmed to the first conserved secondary structural element at both ends, then any internal columns containing amino acid sequence from only one virus were removed to speed computation. For the positive sense and double stranded RNA viruses, the alignment was trimmed to the first predicted conserved secondary structural motif on the left hand side and the second on the right hand side, removing any internal columns with containing amino acid sequence from only one virus.

We performed model selection using MODELGENERATOR (Keane et al. 2006) with 8 categories for the gamma distributed rate variation, selecting the model with the best AIC_2 performance (Crandall 2001). In both cases, the two best fitting rate matrices included the rtREV rate matrix (Dimmic et al. 2002) with empirical amino acid frequencies, a combination that is not implemented in BEAST, so instead the third best fitting model was used; the BLOSUM62 rate matrix (Henikoff and Henikoff 1992) with gamma distributed rate variation using 8 gamma categories, an estimated proportion of invariant sites, and a lognormal relaxed clock (Drummond et al. 2006). This model was run over 30 separate chains for 13,000,000 generations each for the tree containing the negative-sense viruses, and 40 separate chains with the same conditions for the positive-sense only tree. Both trees were run on a cluster in BEAST v1.8.4 (Drummond et al. 2012), with 10,000,000 generations discarded as burn-in for the negative-sense viruses containing tree and 10,750,000 discarded as burn-in for the positive-sense only tree. Convergence was assessed in Tracer v1.6 (Rambaut et al. 2013). The posterior distributions of trees were downsampled to 1000 trees.

2.3.5 Prevalence Estimation

Raw prevalence was estimated for each host virus combination with more than one pool using the JAGS model associated with the ‘truePrevPools’ function in the prevalence package (Devleesschauwer et al. 2014) in R 3.3.2 (R Core Development Team 2016), with a uniform Beta (1, 1) prior, with 100000 iterations for both burnin and sampling. If no pools are negative for a species, the posterior is nearly flat over a large proportion of the parameter space for the underlying probability, making estimates of shortest posterior intervals highly dependent on the stochasticity of the MCMC chains. Given this, 90% shortest posterior intervals are presented in these cases and point estimates were not generated, as the MLE/posterior mode (with a flat prior) of 1 is not meaningful given the extensive bias. Highest posterior density intervals were calculated by the SPIn method (Liu, Gelman, and Zheng 2015). For simplicity, a common likelihood was assumed for each virus between sampling locations.

2.3.6 Co-phylogenetic Mixed Model Analysis

The presence/absence data for each of the viruses in each pool in the phylogenies and the posterior distributions of the host and viral trees were combined to model the effect of the evolutionary histories of the hosts and viruses on the distribution of infection by adapting the model and code of Hadfield et al. (2014) estimated in MCMCglmm (Hadfield 2010). The modelling procedure ignored the fact that sites were sampled in different years. The pools were originally generated by combination of hosts by species rather than by location, which means that a minority of pools had individuals from multiple locations. As such, we treated each location and realised combination of locations as a level of the random effect, terming this the spatial composition effect. In addition to the phylogenetic terms, non-phylogenetic host and virus terms were fitted, as well as a pool ID term and the spatial composition term as random effects. Four versions of the model were fitted. Model 1 included all the viruses, Model 2 excluded the negative sense RNA viruses, Model 3 fitted a pseudo-taxonomic model, where a polytomic viral tree with arbitrary branch lengths (summing to 1, with an equal length between each split) was generated with the viruses being split first by their genomic type (+ve sense RNA, -ve sense RNA and dsRNA) implying a covariance of 0 between genome structures, followed by splitting by the putative viral clades identified by Shi et al. (2016) and Model 4 was run without any of the terms involving the virus phylogeny to assess the effect of uncertainty in the deep splits on the other parameters in the model. An additional model using all the viruses and aggregating the pools per species is shown in Appendix A2.6.

As the samples were small pooled groups of individuals, such that a PCR ‘positive’ represents one or more infections, the data do not conform to a standard binomial model. Instead we modelled the data using a likelihood based on the binomial cumulative density function, as has been previously described (Ebert, Brlansky, and Rogers 2010; Gibbs and Gower 1960), using a logit link. To account for uncertainty in inferred phylogeny/phylogenies, we fitted the models 200 times using random phylogenies drawn from the downsampled joint independent posterior distribution of trees for both the host and virus phylogenies. The posterior distributions from each run of the models were combined manually, with the posterior distributions of all chains downsampled

to below the autocorrelation time of the slowest chain in order to equalise information between runs. Summary statistics were derived from this combined posterior.

The form of Models 1, 2 and 3 is shown below, where *y'* is the value of the latent variable, *i* is the index of the data point, μ is the global mean of the latent variable, ε is a fixed unidentified error term and all other effects are estimated by partial pooling:

y'_i = μ + $host_i$ + $virus_i$ + host phylogenetic $effect_i$ + virus phylogenetic $effect_i$ + host evolutionary assemblage $effect_i$ + virus evolutionary assemblage $effect_i$ + coevolutionary $interaction_i$ + pool ID_i + species $composition_i$ + ε

The form of Model 4 is:

y'_i = μ + $host_i$ + $virus_i$ + host phylogenetic $effect_i$ + host evolutionary assemblage $effect_i$ + pool ID_i + species $composition_i$ + ε

All variance-covariance matrices were generated as described in Hadfield et al (2014). The MCMCglmm default normal fixed effect prior with mean 0 and variance 1^{10} was retained for the global intercept, all variance components were given improper, close to uniform, priors (parameters: V=2x10^{-16}, nu=-2). While in pooled binomial models, the residual variance can in theory be estimated, data is often weakly informative for this parameter, so it was fixed at 1 (Hadfield, *pers comm*). Intraclass correlations were calculated from the model outputs and reported. Highest posterior density intervals were calculated by the SPIn method (Liu, Gelman, and Zheng 2015).

The total phylogenetic variance was calculated as (with the appropriate terms removed for Model 4):

$(\sigma_{hostphylogenetic} + \sigma_{hostassemblage} + \sigma_{virusphylogenetic} + \sigma_{virusassemblage} + \sigma_{coevolutionaryinteraction})/(\sigma_{total} + 1 + \pi^2/3)$

2.4 Results

2.4.1 Read and Assembly Statistics

A total of 134,026,056 sequencing reads were generated for *Bombus lucorum*, 135,590,922 for *Bombus terrestris*, 128,670,194 for *Bombus pascuorum* and

26,838,390 for the other *Bombus* species with 0.37, 0.38, 3.36 and 15.12 percent of reads mapping to the known viruses included in our reference set or the novel bee viruses found in the study. The correlation of the number of reads mapping to the previously described viruses and the novel putative viral contigs between the poly-A selected and nuclease normalized sequencing runs was universally high. Post-sequence processing, the correlation was 1.000 for *Bombus terrestris*, 0.999 for *Bombus pascuorum* and 0.998 for *Bombus lucorum*. This is surprising as in theory these two methods should be sampling from different RNA populations.

2.4.2 Virus-like Sequences

2.4.2.1 Previously Described Viruses Detected

RNA-seq reads mapped to three previously described bee viruses. The majority of these reads mapped either to the Acute bee paralysis virus/Kashmir bee virus complex (henceforth ABPV) or to Slow bee paralysis virus (SBPV). Additionally, in the mixed *Bombus* pool, reads were found mapping to Hubei partiti-like virus 34 (HPLV34), a virus initially detected, though not named, in honeybees by Cornman et al. (2012), then subsequently rediscovered in Chinese landsnails by Shi et al. (2016). Reads in the sequenced small RNAs were found mapping to a considerably more diverse viral community potentially because of the higher comparable depth of sequencing in the small RNAs; Deformed wing virus – type A (Bailey and Ball 1991), Chronic bee paralysis virus (Bailey et al. 1963) , Bee macula-like virus (de Miranda et al. 2015), Ganda bee virus (Schoonvaere et al. 2016), Scaldis River bee virus (Schoonvaere et al. 2016), Black queen cell virus (Bailey and Woods 1977), Apis rhabdovirus 1 (Remnant et al. 2017), Apis rhabdovirus 2 (Remnant et al. 2017), Apis bunyavirus 1 (Remnant et al. 2017), Apis bunyavirus 2 (Remnant et al. 2017), Apis flavivirus (Remnant et al. 2017), Apis dicistrovirus (Remnant et al. 2017), Apis Nora virus (Remnant et al. 2017) and members of the Lake Sinai virus complex (Runckel et al. 2011). Two of the viral contigs generated by the *de novo* assembly had high similarity to previously described plant viruses; both RNAs of White clover cryptic virus 2 (Boccardo et al. 1985) (96% identity), and both RNAs of Arabis mosaic virus (Smith and Markham 1944) (91% identity).

2.4.2.2 Putative Novel Viral-like Sequences

We identified 37 putative novel viral contigs, four assumed to be from DNA viruses and 33 from RNA viruses. See Table 2.1 for more details.

Table 2.1 The putative viral contigs found by the *de novo* assembly of raw RNAseq reads after bioinformatic checking

Putative viral contig	Clade	Genome structure	Sequence length	Closest blastx match	Query cover/ Percent identity	Validated by PCR?	Prevalence data acquired?	Notes
Densovirus_cfminiambidensovirus	Parvoviridae?	positive sense ssDNA?	811	Neodiprion lecontei nucleopolyhedrovirus (YP_025282.1)	39%/ 42%			One incomplete ORF; predicted protein contains the Parvovirus coat protein VP1 (PF08398) motif.
Densovirus_cfminiambidensovirus2	Parvoviridae	positive sense ssDNA	1071	Densovirus SC1065 (AFH02754.1)	29%/ 43%			One incomplete ORF; predicted protein contains the Parvovirus coat protein VP1 (PF08398) motif.
Densovirus_cfDiaphorinaCitriDensovirus	Parvoviridae	positive sense ssDNA	2009	Diaphorina citri densovirus (ALV85426.1)	64%/ 36%			One incomplete ORF; predicted protein contains the Parvovirus non-structural protein NS1 (PF01057) motif.
Densovirus_cfViltainVirus	Parvoviridae	positive sense ssDNA	1345	Diaphorina citri densovirus (ALV85426.1)	67%/ 28%			One incomplete ORF; no predicted motifs. Blastx hits against RNA helicases.
Agassiz Rock virus	Reo	dsRNA	1385	Hubei odonate virus 14 (APG79163.1)	72%/ 40%	x	x	One incomplete ORF; no predicted motifs. Blastx hits against RdRps.
Cnoc Mor virus	Reo	dsRNA	1116	Grange virus (AMO03252.1)	80%/ 38%	x	x	One incomplete ORF; no predicted motifs. Blastx hits against RdRps.
CnocMor-PossibleFragment	Reo	dsRNA	1048	Hubei odonate virus 14 (APG79163.1)	96%/ 47%			One incomplete ORF; no predicted motifs. Blastx hits against RdRps. Aligns with Elf Loch virus and Agassiz Rock virus, potentially part of Cnoc Mor virus.
Elf Loch virus	Reo	dsRNA	2996	Hubei odonate virus 14 (APG79163.1)	87%/ 44%	x	x	One incomplete ORF; no predicted motifs, but manual search found GDD motif of RdRp. Blastx hits against RdRps.
Dumyat virus	Toti-Chryso	dsRNA	4981	Leptopilina boulardi Toti-like virus (YP_009072448.1)	46%/ 48%	x	x	One incomplete ORF; predicted protein contains the RdRp 4 (PF02123) motif.

Sheriffmuir virus	Toti-Chryso	dsRNA	549	Leptopilina boulardi Toti-like virus (YP_009072448.1)	100%/63%	x	x	One incomplete ORF; no predicted motifs, but manual search found GDD motif of RdRp. Blastx hits against RdRps.
Clamshell Cave virus	Bunya-Arena	negative sense ssRNA	505	Ganda bee virus (APT68154.1)	99%/86%	x	x	One incomplete ORF; predicted protein contains the Bunyavirus RdRp (PF04196) motif.
ClamshellCave_cfGandaBeeVirus-possibleFragment1	Bunya-Arena	negative sense ssRNA	736	Ganda bee virus (APT68155.1)	81%/60%			One incomplete ORF; no predicted motifs. Hit against Bunyaviridae glycoprotein. Potentially Clamshell Cave virus M segment.
ClamshellCave_cfGandaBeeVirus-possibleFragment2	Bunya-Arena	negative sense ssRNA	747	Ganda bee virus (APT68156.1)	99%/84%			One incomplete ORF; no predicted motifs. Hit against Bunyaviridae nucleoprotein. Potentially Clamshell Cave virus S segment.
Phlebovirus_cfSalangaVirusGlycoProtein	Bunya-Arena	negative sense ssRNA	2188	EgAN 1825-61 virus (AEL29653.1)	61%/25%			One incomplete ORF; predicted protein contains Phlebovirus glycoprotein G2 (PF07245).
Orthomyxovirus_cfAransasBayVirus	Orthomyxovi-dae	negative sense ssRNA	1449	Hubei orthoptera virus 6 (APG77906.1)	90%/26%			One incomplete ORF; predicted protein contains the Influenza RdRp subunit PB1 (PF00602) motif.
Allermuir Hill virus 1	Hepe-Virga	positive sense ssRNA	7586	Xinzhou nematode virus 1 (YP_009345041.1)	39%/34%	x	x	Two putative ORFs; one containing the FtsJ-like methyltransferase (PF01728), RdRp 1 (PF00978) and viral (Superfamily 1) RNA helicase (PF01443) motifs and one with no predicted motifs and no blast hits.
Allermuir Hill virus 2	Hepe-Virga	positive sense ssRNA	9078	Xinzhou nematode virus 1 (YP_009345041.1)	49%/34%	x	x	Two putative ORFs; one containing the FtsJ-like methyltransferase (PF01728), viral methyltransferase (PF01660), RdRp 1 (PF00978) and viral (Superfamily 1) RNA helicase (PF01443) motifs and one with no one with no predicted motifs and no blast hits.
Allermuir Hill virus 3	Hepe-Virga	positive sense ssRNA	6339	Xinzhou nematode virus 1 (YP_009345041.1)	47%/34%	x	x	Three putative ORFs; one containing the RdRp 1 (PF00978) and viral (Superfamily 1) RNA helicase (PF01443) motifs, one with no one with no predicted motifs and no blast hits, and one with no predicted motifs and blast hits against hypothetical viral proteins.

Mill Lade virus	Hepe-Virga	positive sense ssRNA	3152	Aedes camptorhynchus negev-like virus (YP_009388601.1)	44%/64%	x	x	One incomplete ORF; predicted protein contains the RdRp 2 (PF00978) and viral (Superfamily 1) RNA helicase (PF01443) motifs.
Negevirus_cfBlackford Virus	Hepe-Virga	positive sense ssRNA	1681	Blackford virus (AMO03220.1)	63%/63%			One incomplete ORF; predicted protein contains the RdRp 2 (PF00978) motif.
Negevirus_cfLoretovirus	Hepe-Virga	positive sense ssRNA	962	Hubei virga-like virus 2 (APG77663.1)	92%/31%			One incomplete ORF; predicted protein contains the viral (Superfamily 1) RNA helicase (PF01443) and AAA (PF13245) motifs.
Virga-like_virus_cfHubeiVirgalike15	Hepe-Virga	positive sense ssRNA	2265	Aedes camptorhynchus negev-like virus (YP_009388603.1)	28%/39%			Multiple short ORFs; no predicted motifs, blast hits against hypothetical viral proteins.
Virga-like_virus_cfXingshan Nematode Virus1	Hepe-Virga	positive sense ssRNA	1840	Hubei virga-like virus 17 (YP_009337715.1)	52%/34%			One incomplete ORF; predicted protein contains the viral methyltransferase (PF01660) motif.
Black Hill virus	Picorna-Calici	positive sense ssRNA	1332	Sacbrood virus (AID58097.1)	92%/55%	x		One incomplete ORF; predicted protein contains the RdRp 1 (PF00680) motif.
Boghill Burn virus	Picorna-Calici	positive sense ssRNA	9873	Hubei picorna-like virus 57 (APG78030.1)	60%/98%	x	x	Two putative ORFs; one containing predicted the RNA helicase (PF00680) and RdRp 1 (PF00910) motifs, one with no predicted motifs. The closest well-studied virus is Acyrthosiphon pisum virus, which contains a second ORF at the 3' end of its first with no start codon. This ORF is translated by -1 translational frameshift during protein synthesis (van der Wilk, Dullemans, Verbeek, & Van den Heuvel, 1997). This contig contains similar shift from the protein coding sequence lying in the +3 frame (as defined from the beginning) into the +2 frame, with partial overlap of the ORFs.

Gorebridge virus	Picorna-Calici	positive sense ssRNA	5639	Hubei picorna-like virus 15 (APG77985.1)	45%/ 97%	x	x	Two putative ORFs; one containing the RdRp 1 (PF00680) motif and one containing the CRPV capsid protein like (PF08762) and picornavirus capsid protein (PF00073) motifs. Apparent standard Dicistroviridae organisation.
Gorebridge -like	Picorna-Calici	positive sense ssRNA	1819	Hubei picorna-like virus 15 (APG77986.1)	75%/ 90%			One incomplete ORF; predicted protein contains the CRPV capsid protein like (PF08762) and picornavirus capsid protein (PF00073) motifs.
Loch Morlich virus	Picorna-Calici	positive sense ssRNA	3817	Sacbrood virus (AJA38040.1)	60%/ 40%	x	x	One incomplete ORF; predicted protein contains the RdRp 1 (PF00680) motif.
Mayfield virus 1	Picorna-Calici	positive sense ssRNA	8948	Wenzhou picorna-like virus 47 (APG78496.1)	53%/ 29%	x	x	One complete ORF; predicted protein contains the RdRp 1 (PF00680) and RNA helicase (PF00910) motif.
Mayfield virus 2	Picorna-Calici	positive sense ssRNA	9019	Wenzhou picorna-like virus 47 (APG78496.1)	53%/ 29%	x	x	One complete ORF; predicted protein contains the RdRp 1 (PF00680) and RNA helicase (PF00910) motif.
Nepovirus_ cfBeetRing spot_RNA2	Picorna-Calici	positive sense ssRNA	2426	Beet ringspot virus (NP_620113. 1)	87%/ 95%			One incomplete ORF; predicted protein contains the Nepovirus coat protein, N-terminal domain (PF03689), Nepovirus coat protein, central domain (PF03391) and Nepovirus coat protein, C-terminal domain (PF03688) motifs.
Nepovirus_ cfSoybean LatentSphe ricalVirus	Picorna-Calici	positive sense ssRNA	1282	Soybean latent spherical virus (YP_009330 271.1)	58%/ 47%			One incomplete ORF; no predicted motifs. Blastx hits against RdRps.
Nepovirus_ cfTomatoBl ackRing_R NA2	Picorna-Calici	positive sense ssRNA	3816	Tomato black ring virus (CAA56792.1)	91%/ 60%			One incomplete ORF; predicted protein contains the Nepovirus coat protein, N-terminal domain (PF03689), Nepovirus coat protein, central domain (PF03391) and Nepovirus coat protein, C-terminal domain (PF03688) motifs.
Picornavira les_cfHube iArthropod Virus3	Picorna-Calici	positive sense ssRNA	1798	Kilifi virus (YP_009140 560.1)	19%/ 30%			One incomplete ORF and one complete ORF; no predicted motifs. No blastx hits against identifed proteins.

River Liunaeg virus	Picorna-Calici	positive sense ssRNA	1955	Hubei tick virus 2 (YP_009336542.1)	91%/ 40%	x	x	One incomplete ORF; predicted protein contains the RdRp 1 (PF00680) motif.
Castleton Burn virus	Tombus-Noda	positive sense ssRNA	2714	Shangao tombus-like virus 1 (APG76298.1)	44%/ 43%	x	x	One incomplete ORF and one complete ORF; predicted protein contains the RdRp 3 (PF00998) motif and one ORF contains no predicted motifs. No blastx hits against the ORF with no predicted motifs.
Nodavirus_cfWuhanNodavirus64	Tombus-Noda	positive sense ssRNA	1011	Wuhan nodavirus (ABB71128.1)	78%/ 66%			One incomplete ORF; predicted protein contains the nodavirus capsid protein (PF11729) motif.

2.4.3 Evidence for Infection

2.4.3.1 Titre-based Evidence for Infection

Viruses with large numbers of mapping RNAseq reads in a sample are likely to be infective, as it is the most parsimonious explanation for the large number of mapping reads. If we consider an RPKM of 5 to a novel contig as inconclusive evidence of the virus replicating, the replicating sequence can be seen in Table 2.2. It is worth noting that the titre-based results from the mixed *Bombus* pool is less directly interpretable, as viruses that are common in one species in the pool may be diluted out by the addition of species where they are less common, or completely absent, so a threshold of an RPKM of 5 is more conservative in this case.

We also detected a large number of mapping reads from White clover cryptic virus 2 in *Bombus pascuorum* and Arabis mosaic virus in *Bombus terrestris*, but these do not represent evidence for infection, as the source of the virus being in the collected pollen is a better explanation for the presence of the reads given that these viruses are known to infect plants.

Table 2.2 The reads per kilobase of transcript per million mapped reads mapping to the putative viral contigs from the RNASeq in the *Bombus terrestris*, *Bombus lucorum, Bombus* pascuorum and mixed *Bombus* pools and whether each contig met the criteria of having at least 50 mapping small RNA reads with a size distribution centred at 22nt in *Bombus terrestris, Bombus lucorum* and *Bombus pascuorum*.

Putative viral contig	*Bombus terrestris*		*Bombus pascuorum*		*Bombus lucorum*		mixed *Bombus*
	RNASeq	siRNA	RNASeq	siRNA	RNASeq	siRNA	RNASeq
Densovirus_cfminiambidensovirus							12.3

Densovirus_cfminiambidensovirus2	5.7				221.7	x	71.7
Densovirus_cfDiaphorinaCitriDensovirus		x			51.2	x	13.7
Densovirus_cfViltainVirus							14.2
Agassiz Rock virus							
Cnoc Mor virus							20.4
CnocMor-PossibleFragment							
Elf Loch virus				x			
Dumyat virus							10.3
Sheriffmuir virus							
Clamshell Cave virus							
ClamshellCave_cfGandaBeeVirus-possibleFragment1							5.6
ClamshellCave_cfGandaBeeVirus-possibleFragment2							12.1
Phlebovirus_cfSalangaVirusGlycoProtein							
Orthomyxovirus_cfAransasBayVirus							
Allermuir Hill virus 1	16.0	x		x		x	
Allermuir Hill virus 2		x	22.3	x			
Allermuir Hill virus 3		x		x		x	
Mill Lade virus		x					7.3
Negevirus_cfBlackfordVirus							
Negevirus_cfLoretovirus							
Virga-like_virus_cfHubeiVirgalike15	20.1				5.5		90.4
Virga-like_virus_cfXingshanNematodeVirus1							
Black Hill virus							
Boghill Burn virus					13.9		
Gorebridge virus							
Gorebridge-like	7.3						
Loch Morlich virus							7.8
Mayfield virus 1	403.0	x			255.4	x	7.7
Mayfield virus 2	5.5	x	339.0	x			558.8
Nepovirus_cfBeetRingspot_RNA2							
Nepovirus_cfSoybeanLatentSphericalVirus							
Nepovirus_cfTomatoBlackRing_RNA2							6.3
Picornavirales_cfHubeiArthropodVirus3							
River Liunaeg virus							9.6
Castleton Burn virus		x	11.0	x	12.7	x	123.0
Nodavirus_cfWuhanNodavirus64							

2.4.3.2 siRNA-based Evidence for Infection

RNA interference is the predominant arthropod defence against RNA viruses, and may be important for protection against DNA viruses (Bronkhorst et al. 2012). The pathway cleaves cytoplasmic dsRNA, leading to the production of short RNA fragments. The presence of these fragments is strongly indicative of a replicating virus. The RNAs produced by the bumblebee RNA interference system are 22nt. For positive sense RNA viruses, the presence of a strong peak of mapping 22nt fragments against the negative strand is the best indictor of infection, as the negative strand is only produced during replication, the reverse is true of negative sense RNA viruses, as the positive strand is only produced during replication. As such, we decided at least 50 reads mapping with at 22nt as the criteria for confirmation of replication. The putative viral contigs that met this threshold are marked in Table 2.2. The distribution of the mapped small RNA reads across the viral contigs is shown in Figure 2.2.

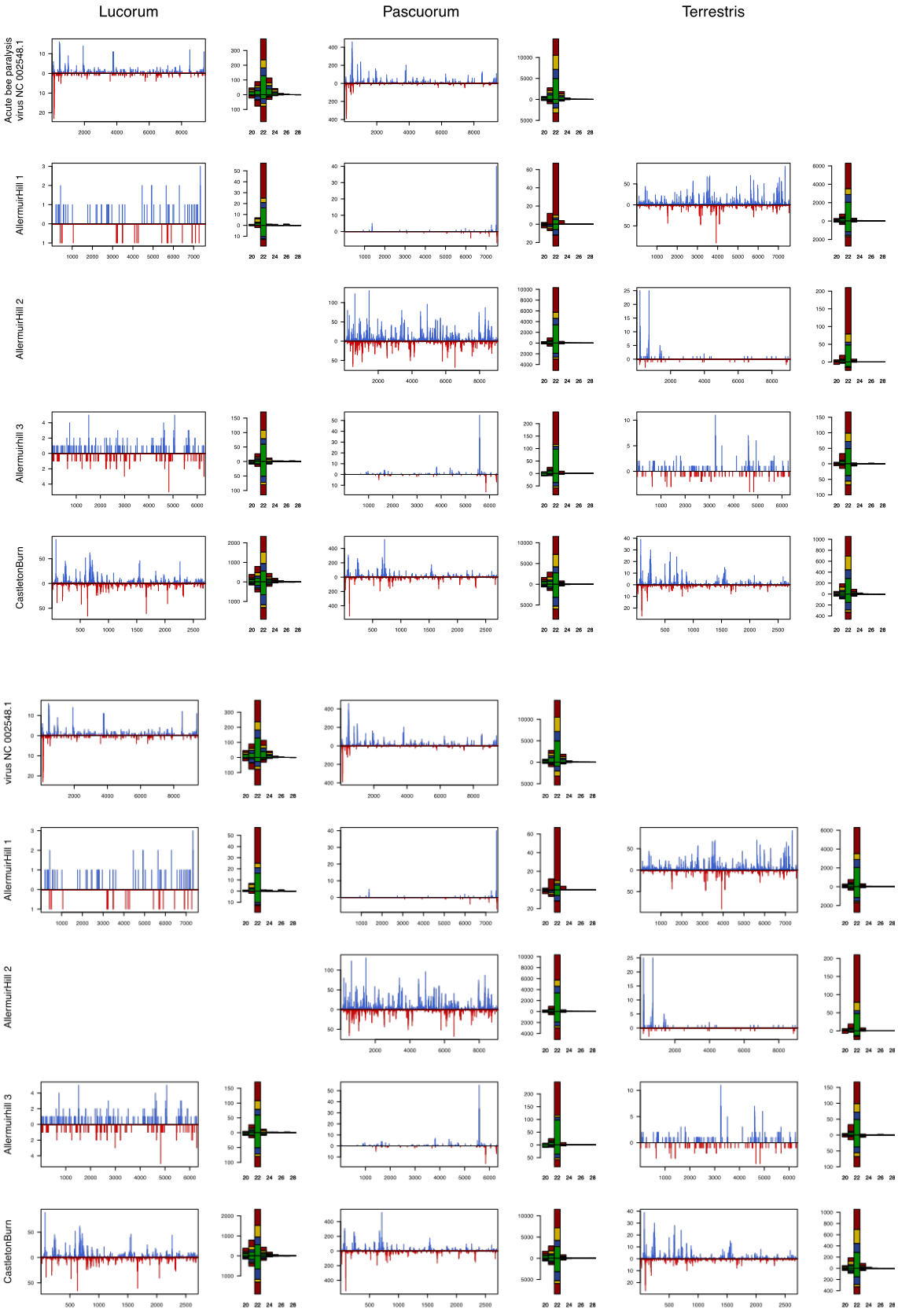
Small RNA reads mapping to the putative viruses provide more information about whether the virus is really replicating in the bee species than general RNA-seq reads. This is true for two reasons. Firstly, small RNA reads are only generated if the virus is replicating in the sample. This is not true with RNA-seq reads, which may be present purely through contact with the virus rather than its replication. Secondly, as reads of a specific size would be expected, a read spectra not centred at 22nt implies that the virus may be replicating within a species associated with the bumblebee rather than the bee itself. RNA-seq reads provide no information about this question.

In all three species with siRNA data, Phlebovirus_cfSalangaVirusGlycoProtein has a number of reads exceeding the threshold mapping to it. However, the size spectra of these reads is centred on 24nt with a strong bias for a terminal uracil and the antisense mapping orientation being more prevalent. This U-bias is consistent with insect piRNAs (Brennecke et al. 2007), and the predominant antisense orientation is consistent with the piRNA mapping pattern to endogenous viral elements (EVEs) in mosquitoes (Suzuki et al. 2017). However, the size of piRNAs in bumlebees is generally larger than this, potentially providing evidence against this hypothesis (Sam Lewis, *pers comm*). This pattern indicates that Phlebovirus_cfSalangaVirusGlycoProtein is

potentially an EVE that has either been gained multiple times or has been maintained in the bumblebee genome since at least the *B. pascuorum*-*B. terrestris*/*B. lucorum* split or has been independently gained multiple times.

The read length spectra of Mayfield virus 1, Mayfield virus 2 and Slow bee paralysis virus, while still peaking at 22nt (excepting for Mayfield virus 1 in *B. lucorum*), are noticeably flatter than that for the other viruses observed. This is consistent with the existence of an RNAi suppressor in these viruses, given both the assumption that this behaviour is both aberrant and virus driven, and the similarity in spectra to the Drosophila C virus in *Drosophila* (Webster et al. 2015), which contains a suppressor (van Rij et al. 2006).

B. pascuorum also had siRNA reads mapping to Nepovirus_cfTomatoBlackRing RNA2. However, the read length spectra were centred on 21nt, rather than the 22nt of bumblebee viRNAs. A 21nt centring is consistent with siRNA's produced from DLC4, the main protein acting against positive sense RNA viruses in plants (Ding and Voinnet 2007). Given that this contig appears to be related to the nepoviruses, a classically plant infecting group, the acquisition of the small RNAs through the nectar or pollen seems likely in this case. Though to our knowledge, high concentrations of viRNAs in the nectar and pollen of plants have never been reported.


Lucorum
Pascuorum
Terrestris
Acute bee paralysis virus NC 002548.1
AllermuirHill 1
AllermuirHill 2
AllermuirHill 3
CastletonBurn
virus NC 002548.1
AllermuirHill 1
AllermuirHill 2
AllermuirHill 3
CastletonBurn

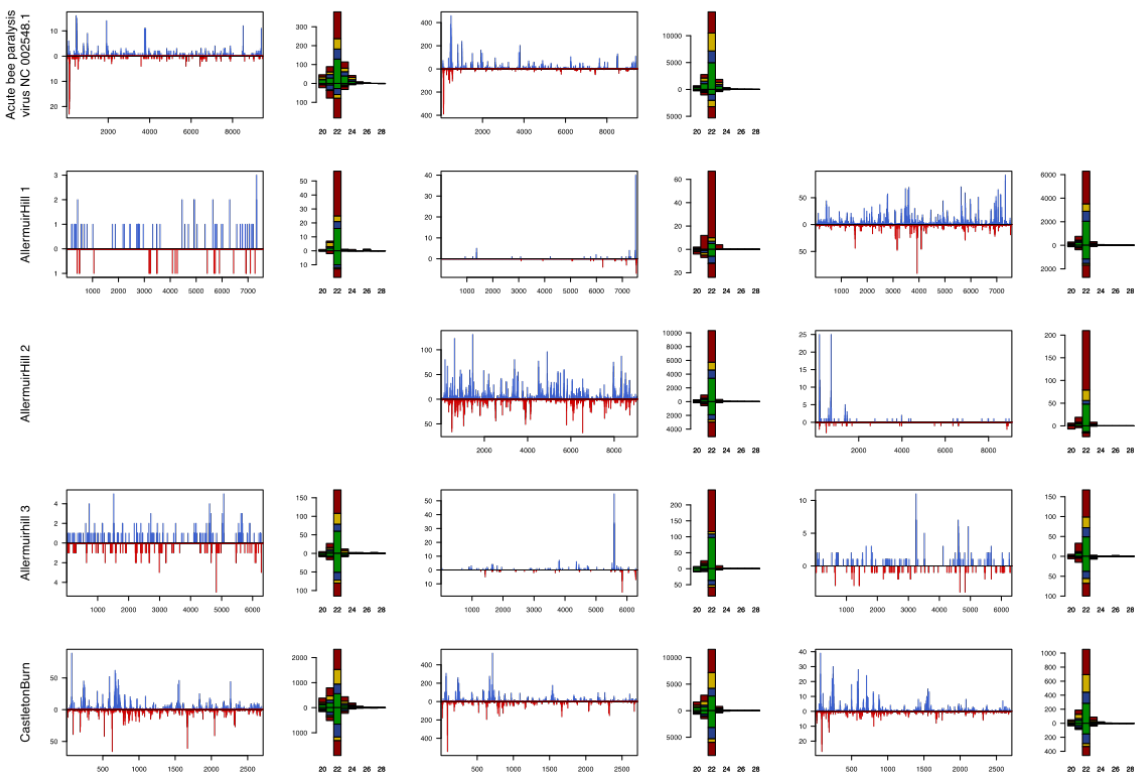


Figure 2.2 The mapping small RNA reads to each virus with at least 50 mapping reads, and their read size spectra. For positive sense RNA viruses, reads mapping to the negative sense strand are indicative of infection. For negative sense RNA viruses, reads mapping to the positive sense strand are indicative of infection. In both cases, the read size spectra should be centred at 22 nucleotides to provide evidence of infection. Blue lines represent reads mapping to the forward strand at that genomic position, red lines represent reads mapping to the reverse strand. The histogram of read size spectra shows the count of reads of each length mapping in the forward (above) and reverse (below) directions. The colouring of each bar shows the counts of the reads beginning with each 5' base (red-U, blue-C, green-A, yellow-G).

2.4.4 Prevalence

Species level prevalences differed dramatically between the different viruses, see Figure 2.3. Prevalences were generally low, with modal viral prevalences for most host-virus combinations being below 15%, with many cases of viruses being either completely or near completely absent from hosts in our sample. Slow bee paralysis virus was by far the most common virus in the sample, with estimated prevalences of greater than 25% in multiple species. Our ability to locate the prevalences of common viruses where all pools were positive was limited by the pooling, leading us to only be able to assign lower bounds to prevalences in these cases (as can be seen by the broad 90% highest posterior density intervals in Figure 2.3), but in 7 of 11 species, all pools were positive for SBPV. Acute bee paralysis virus, Hubei partiti-like virus 34, Castleton Burn virus, Gorebridge virus, Mayfield virus 1 and Mayfield virus 2 all reached 15-25% prevalences in multiple species. Therefore, all previously reported

honeybee viruses for which prevalence was estimated are generalists in bumblebees. Several viruses showed strong signals of species specificity, having very low to zero prevalences in multiple host species but high prevalences in others. Examples of this pattern include Allermuir Hill virus 1 in *B. terrestris,* Allermuir Hill virus 2 in *B. pascuorum*, Allermuir Hill virus 3 in *B. magnus* and *B. monticola*, as well as Loch Morlich virus and River Luinaeg virus in *B. jonellus*.

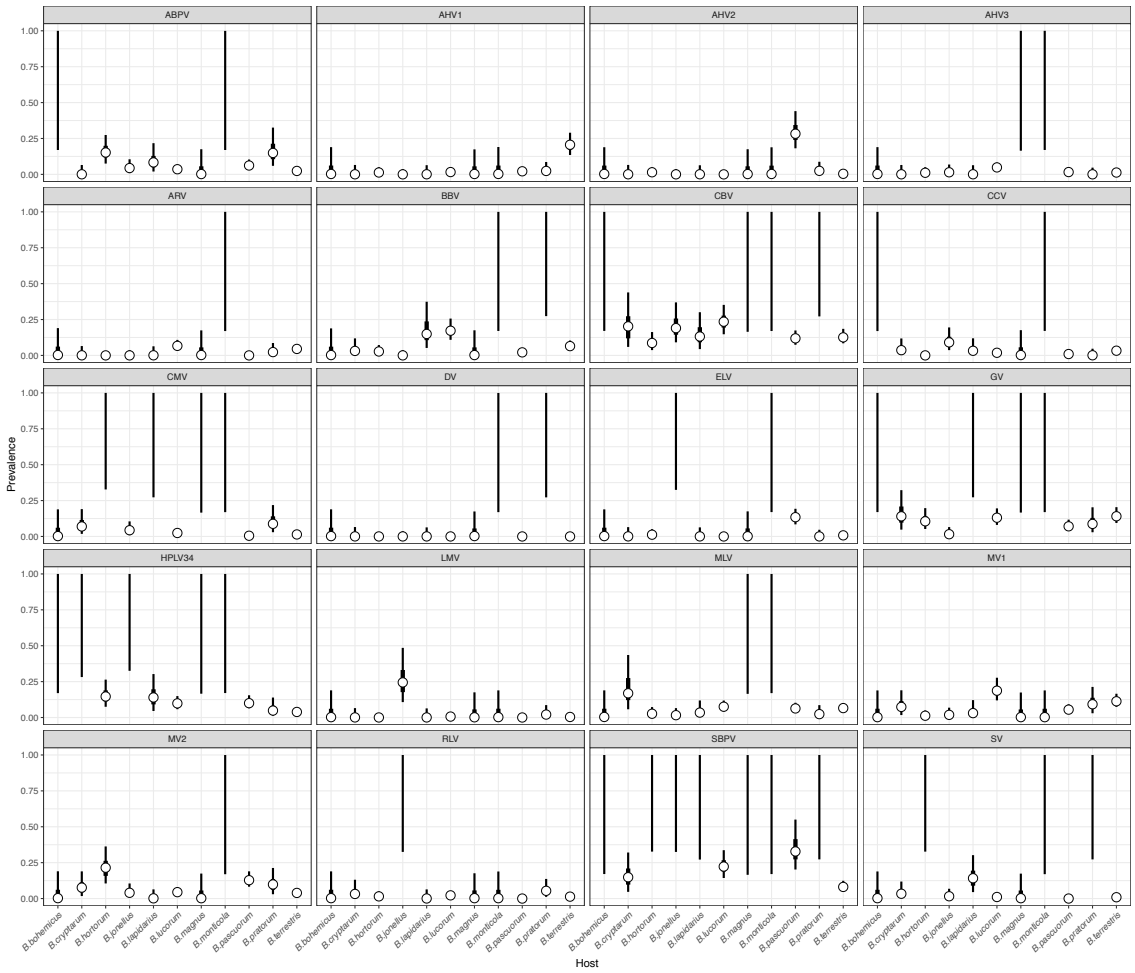


Figure 2.3 The estimated prevalence of a subset of the novel viruses in every host. The circle represents the posterior mode, the thick lines represent the 50% highest posterior density interval and the thin lines show the 90% highest posterior density interval. For host-virus combinations where every pool was positive, only the 90% highest posterior density interval is shown, to illustrate the lower bound of the estimated prevalence.

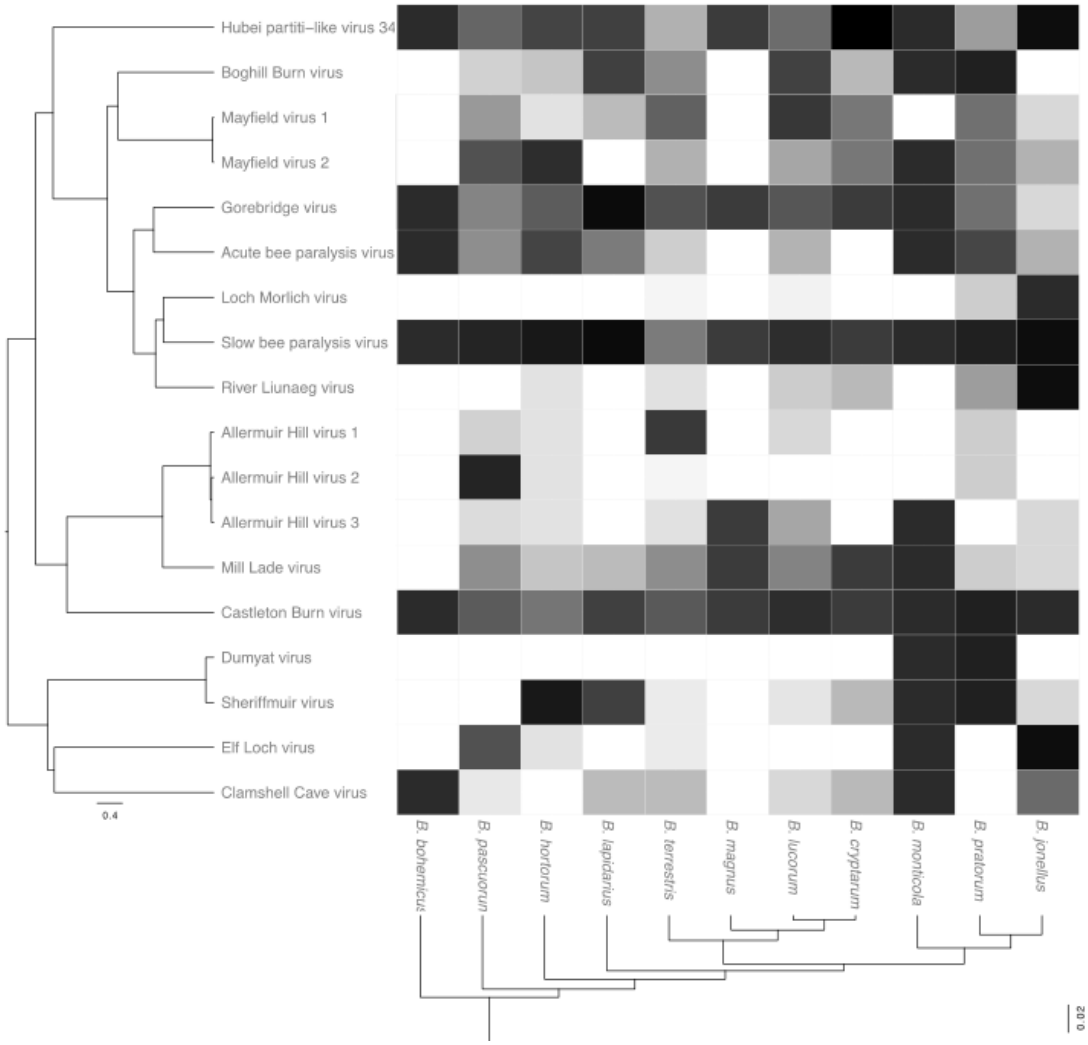


Figure 2.4 A heatmap of the posterior modal estimates of the pooled prevalence estimates of the pooled prevalence on the data scale for each virus/host combination with a beta (1,1) prior. Hosts and viruses are ordered by phylogenetic relatedness, the trees represent the maximum clade credibility trees.

2.4.5 Host-Pathogen Co-phylogenetic Models

The model results fall into two clear groups. The first group is comprised of the model including the estimated phylogeny with all the viruses (M1), the model including the estimated phylogeny excluding the negative sense RNA viruses (M2) and the pseudotaxonomic model (M3), the second is the model excluding any kind of viral relationships (M4). The fact that the pseudotaxonomic model gave qualitatively similar results to the models including the estimated viral phylogenies is relatively surprising, despite previous reports of similar results, and means that the results are robust both to phylogenetic uncertainty and uncertainty in common ancestry. See Figure 2.5 and Table 2.3 for model comparisons.

Four results were common to all models. The proportion of variance explained by the pool ID random effect, representing the within-species variance in prevalence explained by different pools of the same species having different underlying prevalences, was precisely estimated and small, implying that each pool was roughly representative of the species it was comprised of. The proportion of variance explained by the spatial composition random effect was also precisely estimated and small, indirectly implying that different sampling locations did not systematically differ in the average background viral prevalence at the time of sampling. The final shared result between the models was that both the host non-phylogenetic and host phylogenetic effects were estimated to explain a small proportion of the total variance in prevalence between samples with posterior modes near zero. The host non-phylogenetic intraclass correlation was estimated with moderate precision and the host phylogenetic intraclass correlation was estimated with moderate to poor precision. These results imply that there is neither large amounts of non-phylogenetic variation nor a large phylogenetic effect in realised general susceptibility.

The remaining results were qualitatively similar within the groups of models, but differed between them. As no effects involving the viral phylogeny were estimated in M4, the viral phylogenetic effect, viral evolutionary assemblage effect and coevolutionary interaction terms were undefined. Therefore, any phylogenetic variation in realised general infectivity falls into non-phylogenetic viral random effect, this lead to a moderately sized and imprecisely estimated non-phylogenetic viral intraclass correlation in M4. The results in the other group of models were consistent with this, with a modally small but highly imprecisely estimated viral phylogenetic effect intraclass correlation and a modally slightly larger but still imprecisely estimated non-phylogenetic viral intraclass correlation. This implies the variation between viruses in RGI explains a large amount of the total variation in prevalence, but that we cannot accurately split it into variation consistent with evolution by Brownian motion along a phylogeny and unexplained residual between virus variation.

In M3, the largest proportion of the total variation in prevalence between samples was due to a host evolutionary assemblage effect. This accounts for

the variation explained by related hosts being infected to similar degrees with the same sets of viruses. The fact that this effect became considerably smaller when the virus phylogenies were added in the other models implies that some of this variation was better explained either by one or more of the three additional terms. When the virus phylogenies are added, a small host evolutionary interaction is still estimated, implying that the effect is indeed real and not simply due to the lack of ability to partition viral phylogenetic variation in M4. In M1 and M2, a viral evolutionary assemblage effect and coevolutionary interaction of a similar size to the host evolutionary assemblage effect were also estimated, in M3, these effects could not be resolved from 0. The three effects were all estimated with moderate precision. This implies that related viruses do share host assemblage to a detectable level, and likewise that related hosts are likely to have similar prevalences for related viruses.

M4 estimated that approximately 40% of the total variation in prevalence can be explained by effects involving the phylogeny of the host. The addition of the viral phylogeny and the effects associated with it in the other three models increased the proportion of variance explained phylogenetic random effects to 55-60%.

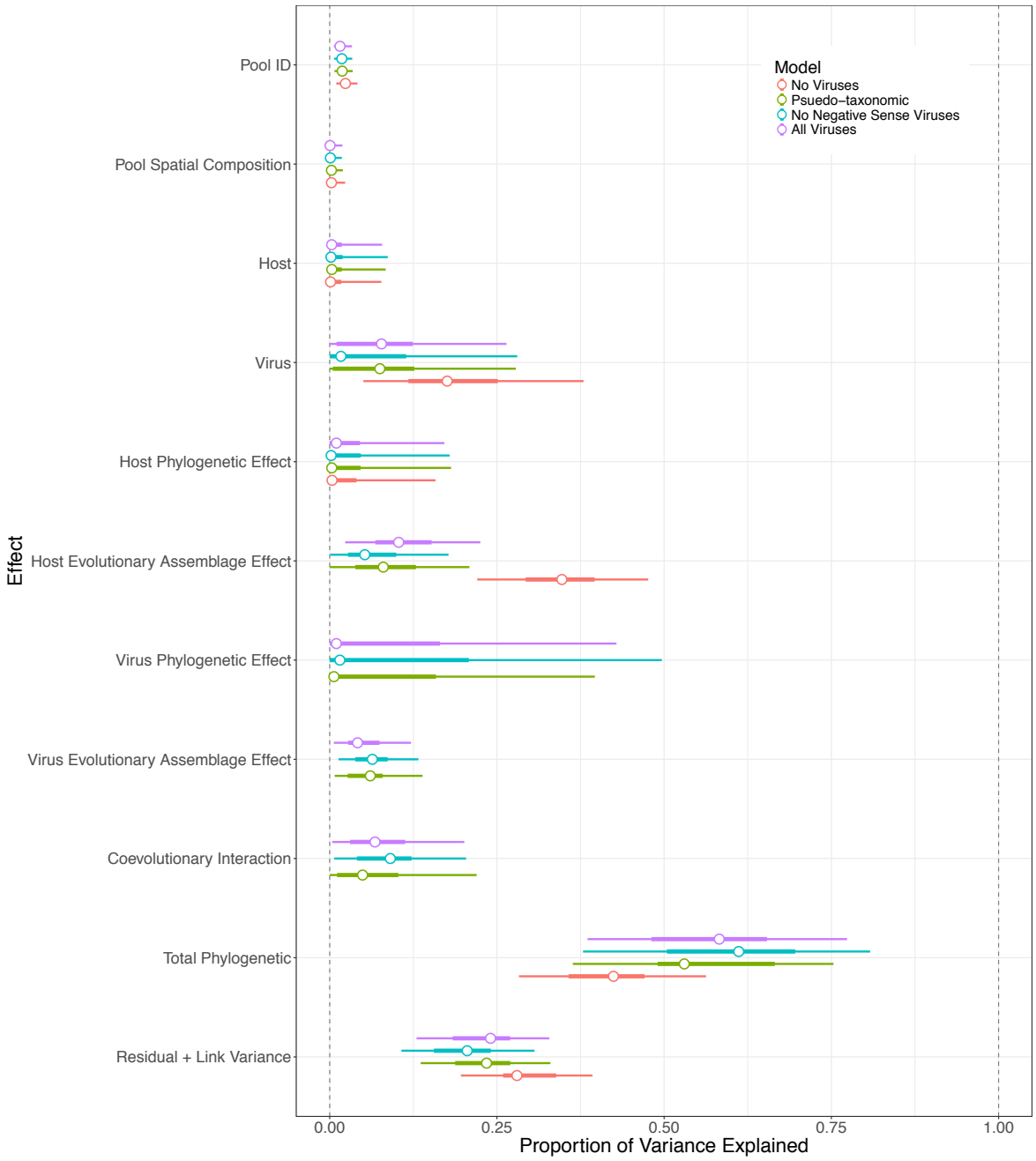


Figure 2.5 Comparison of estimated proportion of variance in prevalence explained by different parameters between models. For each parameter, the circle represents the modal estimate, the thick bars represent the 50% highest posterior density interval and the thin bars represent the 90% highest posterior density interval.

Table 2.3 Model estimates for the intra-class correlations of each variance component. The point estimate is the posterior mode, the numbers in brackets represent the 90% highest posterior density interval.

Effect	Model 1 All Viruses	Model 2 No Negative Sense Viruses	Model 3 Psuedo-taxonomic	Model 4 No Viruses
Pool ID	0.015 (0.006, 0.033)	0.018 (0.007, 0.034)	0.019 (0.007, 0.034)	0.024 (0.010, 0.034)
Pool Spatial Composition	0.001 (0.000, 0.019)	0.001 (0.000, 0.018)	0.003 (0.000, 0.020)	0.003 (0.000, 0.023)
Host	0.003 (0.000, 0.078)	0.002 (0.000, 0.087)	0.003 (0.000, 0.084)	0.001 (0.000, 0.077)
Virus	0.077 (0.000, 0.264)	0.017 (0.000, 0.280)	0.075 (0.000, 0.278)	0.176 (0.050, 0.380)
Host Phylogenetic Effect	0.010 (0.000, 0.171)	0.002 (0.000, 0.179)	0.003 (0.000, 0.181)	0.004 (0.000, 0.158)
Virus Phylogenetic Effect	0.010 (0.000, 0.429)	0.015 (0.000, 0.496)	0.006 (0.000, 0.396)	-
Host Evolutionary Assemblage Effect	0.103 (0.023, 0.225)	0.053 (0.001, 0.178)	0.080 (0.000, 0.209)	0.347 (0.221, 0.476)
Virus Evolutionary Assemblage Effect	0.042 (0.006, 0.122)	0.064 (0.013, 0.133)	0.061 (0.007, 0.139)	-
Coevolutionary Interaction	0.068 (0.004, 0.201)	0.091 (0.007, 0.204)	0.049 (0.000, 0.220)	-
Total Phylogenetic	0.582 (0.385, 0.773)	0.611 (0.379, 0.808)	0.530 (0.364, 0.753)	0.424 (0.283, 0.563)
Residual + Link Variance	0.241 (0.130, 0.328)	0.206 (0.107, 0.306)	0.235 (0.136, 0.330)	0.280 (0.196, 0.393)

2.5 Discussion

We have found 33 novel putative viral contigs in the transcriptomes of wild caught bumblebees from across Scotland. Using 15 newly discovered and 3 previously known multihost viruses found in 11 wild bumblebee species that share transmission opportunities, we have found that variation in the prevalence of infection in the wild is explained by related hosts having similar viral assemblages, related viruses having similar host assemblages and related hosts having similar prevalences for related viruses.

2.5.1 Virus Discovery

There is now an extensive diversity of viruses known in bees, with nearly every new transcriptomic study looking at virus diversity finding novel viruses (Cornman et al. 2012; Mordecai et al. 2015; Remnant et al. 2017; Runckel et al. 2011; Schoonvaere et al. 2016). The fact that we have found up to 33 more viruses indicates that we are not yet near saturation. The size distribution of the small RNA reads, with a peak at 22nt for most viruses characteristic of the honeybee anti-viral RNAi pathway (Chejanovsky et al. 2014), indicates that these are likely infecting the bees themselves. However, it cannot categorically be ruled out that some of the described viruses are infecting pollinator-associated organisms, such as their microbiota or parasites. While mite viRNAs appear to be centred at 24nt (Remnant et al. 2017), nematode viRNAs are also centred at 22nt (Félix et al. 2011). Very little is known of the nematode parasites of bumblebees excepting for *Sphaerularia bombi* which is only regularly known to infect queens. However, other nematodes have been rarely reported in members of the *Bombus* genus not used in this study (Rao, Poinar, and Henley 2017). It seems unlike that they are common enough to explain the observed prevalences of viruses here however. Therefore, as for all sequence-based viral discovery studies, ultimate confirmation that these truly represent viruses of bumblebees could only be obtained by controlled lab experiments, or growth in bumblebee cell culture.

2.5.2 Phylogenetic Effects

We found no evidence for a host phylogenetic effect, and equivocal evidence for a viral phylogenetic effect, as phylogenetic viral and non-phylogenetic viral variation in prevalence could not be partitioned. The potential viral phylogenetic effect, with realised general infectivity being correlated across the phylogeny, is potentially the largest effect detected in the study and potentially non-existent. This is likely due to the fact that the number viruses in the study was small given that the predominant interest was on the assemblage and coevolutionary effects rather than the phylogenetic effects. Given this, we can make no strong conclusions about whether variation in realised general infectivity is phylogenetically associated, only that viruses strongly differ in it.

2.5.3 Host Evolutionary Assemblage Effect

As hypothesised, a host assemblage effect, where phylogenetically related hosts share viral assemblages, termed a host evolutionary effect in Hadfield *et al* (2014), was found, showing that related hosts share prevalences for similar sets of viruses. At the level of presence or absence in a specific host, this pattern can be explained by biased host switching, where parasites successfully establish infections on hosts closely related to ones that they are already capable of infecting (assuming that the initial pool of infected hosts species was random). Biased host switching is known to be a general phenomenon, and has been observed in macroparasites, viruses and protozoans (reviewed in Longdon (2014)). Given that most of the viruses found here are true multihost viruses, with most being detected in over half the host species, and the dataset is small, biased host switching as traditionally envisioned is unlikely to be the cause, as there would be little power to detect it. Instead, it appears to be that related hosts have similar prevalences for each of the viruses that infect them. Two non-exclusive hypotheses could explain this: phylogenetically-biased cross-species transmission and phylogenetically-biased exposure.

Phylogenetically-biased cross-species transmission, i.e multihost viruses being more frequently transmitted among close relatives, could cause their prevalences to be phylogenetically correlated. Mechanistically, this would be expected to occur simply because related hosts present correlated environments from the perspective of the virus, so there should be cross-adaptation. This cross-adaptation should cause the probability of successful infection on contact to be similar between related hosts for leading to a then a phylogenetically-biased transmission network mediated through the underlying correlation between host environments.

Phylogenetically-biased exposure represents an evolutionarily-driven ecological phenomenon that biases transmission probabilities. This can be mediated by niche overlap. In this case, flower preference might be a driver of phylogenetically-biased exposure, as contaminated flowers are likely to be important source of intra- and interspecies pathogen transmission (Durrer and Schmid-Hempel 1994; Graystock et al. 2015; McArt et al. 2014). In the single pathogen case, the flower visitation network has been shown to be associated

with the partitioning of genetic diversity of *Crithidia bombi* between bumblebee hosts (Ehlenfeldt and Martin 2009). Different bee species show tongue length differences, which are phylogenetically associated (Harmon-Threatt and Ackerly 2013), and the differences in tongue length correlate with differential flower usage between bee species (Goulson et al. 2008; Inouye 1978). If infection occurs at contaminated flowers, then even if infection were occurring at random with respect to the host phylogeny, the structuring of the flower usage network would cause flowers to build up their viral communities through chance sorting. Related bee species would then become more heavily infected with whichever viral communities the flowers that bee species of their tongue lengths happen to visit. This would drive consistent phylogenetically-correlated differences in viral infection rates through differential exposure. Once a structured exposure network has been generated, it is expected that this would be reinforced by parasite adaptation, as the (fast evolving) parasite would then be spending relatively more or less time in some host species due to the partitioning of hosts between floral species.

2.5.4 Virus Evolutionary Assemblage Effect

A viral evolutionary assemblage effect was estimated to explain about 5% of the total variance in prevalence within the samples, implying that related virus species share similar host assemblages to a small degree. Given that most of the viruses used in this study were not closely related, it appears implausible that the host community infected by the viruses was conserved over evolutionary time, especially as the deep splits in the viral families are have been argued to predate the origin of the host species in this case by many millions of years (Koonin et al. 2008). A possible alternate explanation is that groups of related viruses share broad replication strategies, by which we mean aspects of their physiology that relate to their replication, rather than any specific trait. These broad differences lead to different potential targets for the immune response. In mammals, some restriction factors, part of the innate immune system, are known to target specific viral families while others have a broader range (Duggal and Emerman 2012). If there were variation between host species in the expression of immune genes that had targeted responses to specific viral families, this could lead to related viruses having similar levels of efficiency of replication in those species. Assuming that these differences in

expression of the immune genes were not themselves correlated between host species, this could lead to related viruses having similar prevalences in sets of hosts.

2.5.5 Coevolutionary Interaction

A coevolutionary interaction of a similar size as the virus and host evolutionary assemblage effects was estimated, as has previously been observed in other species (Hadfield et al. 2014; de Vienne et al. 2009). The simultaneous action of the factors leading to a viral evolutionary assemblage effect and a host evolutionary assemblage effect can lead to a coevolutionary interaction, where related hosts are infected to similar degrees by related viruses. If related viruses share their assemblages due to effects that are predominantly acting at evolutionary time such as the gain or loss of host immune components, then either phylogenetically-biased exposure or phylogenetically-biased transmission allow transmission from these hosts to related hosts, then a coevolutionary interaction can emerge. Similarly, cospeciation could lead to this signature, but due to both the multihost nature of the viruses involved and the difference in timescales between the viral and host evolution, this is highly unlikely to be the explanation in this system.

2.5.6 Conclusion

While it is clear that viruses are abundant in pollinators, it has remained uncertain what determines the distribution of pollinator viruses, outside of a few well studied cases (Fürst et al. 2014; McMahon et al. 2015). With the series of novel viruses discovered in this study, we have investigated predictors of these virus-host associations and found that both the virus and host evolutionary histories contribute to the variation in prevalence between samples. This supports both theory and prior empirical evidence that related species are more at risk of infection from each other's diseases than the diseases of distantly related species. On an applied level, this suggests that the introduction of a novel virus into a community, for instance through poor biosecurity in bumblebee breeding facilities leading to the spread of infected colonies, may not only put at risk closely related bumblebees, but the entire bumblebee community, as the viral host assemblage is predicted by factors other than just host relatedness.

Chapter 3 - The ecological predictors of viral infection in wild bumblebees

David J. Pascall containing additional work from Matthew Tinsley, Sam Braine, Joseph Faulks, Darren J. Obbard and Lena Wilfert

3.1 Abstract

Viruses are a key regulator of natural populations. Despite this, we have limited knowledge of the diversity and ecology of viruses without large fitness effects on their host. Surprisingly, this is even the case in wild host populations that provide direct economic benefits, where even small fitness effects may have major financial impacts in aggregate. One such group of hosts are the bumblebees (genus *Bombus*), which have a major role in the pollination of food crops and have suffered population declines and species losses in recent decades. In this study, we investigated the ecological factors that determine the prevalences of four recently discovered bumblebee viruses without known fitness effects (Mayfield virus 1, Mayfield virus 2, River Liunaeg virus and Loch Morlich virus), along with two previously known viruses (Acute bee paralysis virus and Slow bee paralysis virus). We found evidence for a positive effect of precipitation on the prevalence of River Luinaeg virus. The sequencing of Mayfield virus 1 isolates also indicated that cryptic viral specialisation may underlie differences in infection between host species. There is a strong association between the presence of Loch Morlich virus and River Luinaeg virus, which remains after controlling for host species, location and other relevant ecological variables. This study represents one of the first steps in the description of predictors of bumblebee infection in the wild independently of the presence of honeybees.

3.2 Introduction

Viruses are amongst the most abundant and diverse groups of organisms on Earth (Wommack et al. 2015); they affect almost all species as obligate pathogens. Despite this, viral ecology in natural populations remains understudied. In the wild, infection is generally only detectable when clear symptoms of the underlying disease are present, such as discolouration, aberrant tissue structures, or a noticeable increase in mortality. These symptoms are only rarely present in natural infections (Mackenzie and Jeggo 2013). By focusing only on those viruses that cause obvious symptoms in well-studied host species, we are likely to be underestimating both the diversity of viruses and their ecological importance in the regulation of natural populations.

For example, it is only in the last 30 years, that we have come to understand that the dynamics of algal blooms are strongly driven by density dependent regulation of the algae through viral infections (Bratbak, Heldal, and Egge 1991; Brussaard et al. 1996). This observability problem has recently been ameliorated by the development of relatively cheap and easily applicable molecular techniques allowing the detection and identification of pathogenic organisms within both the host, and the environment, enabling the systematic study of viral ecology in wild populations. This is especially important for threatened host species, where understanding the viral burden may have direct conservation consequences.

Given the importance of pollinators economically, their viruses are comparatively well-studied. Over 50 viruses have now been described in bees, and their importance to survival is well recognised (Dainat et al. 2012; Doublet et al. 2015; Genersch et al. 2010; Natsopoulou et al. 2017). However, the majority of this work has been performed in honeybees, and the knowledge of the viral ecology of bumblebees is less full. Viruses known from honeybees have been shown to have pathogenic effects in *Bombus terrestris* (Graystock et al. 2016; Manley et al. 2017) and their prevalences have been assayed across the UK (Fürst et al. 2014; McMahon et al. 2015), but outside of the presence of sympatric honeybees, no strong predictors of infection have been described.

In the wild, differences in prevalence between hosts or locations can be explained by a variety of ecological factors. If a virus is spread by environmental contamination or aerosolisation, then abiotic factors can become important. In bumblebees, infection is often thought to take place at flowers (Durrer and Schmid-Hempel 1994; Graystock et al. 2015; McArt et al. 2014) and so any factors that reduce contamination of the flower heads may be predicted to correspondingly reduce the rate of infection in the general bumblebee population. As flower heads are exposed to the elements, obvious candidates are UV exposure, rainfall, temperature and humidity. Areas with high UV levels may deactivate most virus particles, an effect thought to be highly important in the regulation of viral populations in oceanic waters (Suttle and Chen 1992). Rainfall may physically clean the flowers themselves if it is frequent enough. Aerosolised viruses are known to have different rates of deactivation in different

relative humidities depending on whether or not they are encapsulated (Yang and Marr 2012). Increasing temperature is also known to increase the rate of viral deactivation, both independently and through an interaction with relative humidity (Mbithi, Springthorpe, and Sattar 1991). Furthermore, bees must physically reach the flowers where infection can occur, so factors that reduce the contact of workers with heavily contaminated flowers may also reduce viral prevalence. Changes in wind speed are known to change the relative rates of collection of pollen and nectar (Peat and Goulson 2005), which may cause bees to visit different flowers in the community, while simultaneously altering the energetic costs of foraging (Wolf, Ellington, and Begley 1999), and consequently the susceptibility to infection.

Environmental effects could lead to systematic spatial partitioning in prevalence between areas. This could cause interspecific differences in prevalence if the species distributions differ spatially, as bumblebees do in the United Kingdom (Sladen 1912). However, environmental conditions are not expected to lead to interspecific prevalence differences locally. Incidence is also impacted by the background level of infection in the community, as this would change the contact rate between a host and other infected individuals. If an infection is uncommon, naïve hosts would rarely become infected, as their rate of contact with the virus itself would be low all else being equal. This is an especially important driver in species that act primarily as spillover hosts for an infection, where, by definition, the intraspecific transmission rate is small relative to the interspecific transmission rate. This pattern is shown by the prevalence of Deformed wing virus in bumblebee species, which is largely predicted by the corresponding prevalence in sympatric honeybees (Fürst et al. 2014; McMahon et al. 2015).

Systematic differences in infection between areas can also be generated by low levels of interaction between populations, termed weak coupling in the epidemiological literature (reviewed in Keeling, Bjørnstad, and Grenfell 2004). These potential spatial differences in infection may have important knock-on effects on the amount of genetic diversity in the viral populations. When the linkage between host populations is weak, viral evolution occurs more independently as variants cannot mix globally. This may lead to spatially

structured genetic diversity, with viral types being shared unevenly between locations. The degree of linkage between bumblebee populations is only poorly understood (Goulson 2010), and as the linkage of the viral populations depends strongly on the linkage of their hosts, the amount of diversity that would be expected in multihost populations is unclear. This is particularly important in multihost pathogen populations, where incongruent fitness landscapes between hosts may favour cryptic specialisation with different genetic variants of the same pathogen preferentially associating with some host species (Le Gac et al. 2007; Withenshaw et al. 2016). It is worth noting that this is not the same as cryptic pathogen species infecting the same hosts, which has also been commonly observed (Martínez-Aquino et al. 2009; Sehgal et al. 2006). This may lead to considerably within pathogen diversity even in sympatric host populations.

Here we investigate the impact of environmental factors on the prevalences of viruses in natural bumblebee populations from 9 sites across Scotland. We hypothesise that differences in UV radiation, precipitation, humidity, temperature and wind speed will result in differences in the prevalences of four recently discovered bumblebee viruses without known fitness effects (Mayfield virus 1, Mayfield virus 2, River Liunaeg virus and Loch Morlich virus). We also consider the diversity in these viruses and contrast this to two known honeybee viruses with fitness effects on the host found in these populations (Acute bee paralysis virus and Slow bee paralysis virus).

3.3 Methods

3.3.1 Sampling Regime and Post Collection Treatment

We collected a total of 759 bumblebees of 13 species from across 9 sites across Scotland, UK (see Table 3.1), using sampling methods described in Chapter 2. We performed individual RNA extractions using TRIzol (Life Technologies) following the manufacturers' standard protocol. RNA was transcribed into cDNA using random hexamers and goScript MMLV reverse transcriptase (Promega) following the manufacturers' instructions.

Table 3.1 The species-site breakdown of the bumblebees used in the study

	Locations									
	Dalwhinnie	Edinburgh	Glenmore	Gorebridge	Iona	Ochils	Pentlands	Staffa	Stirling	Total
Bombus bohemicus	0	0	1	6	0	1	2	0	0	10
Bombus campestris	0	0	0	1	0	0	0	0	0	1
Bombus cryptarum	0	3	0	5	2	19	1	0	0	30
Bombus hortorum	4	0	1	59	11	0	0	0	0	75
Bombus jonellus	3	0	31	0	1	0	0	21	0	56
Bombus lapidarius	0	8	0	17	0	4	0	1	2	32
Bombus lucorum	0	30	1	75	3	35	3	0	5	152
Bombus magnus	0	0	0	0	1	7	0	0	0	8
Bombus monticola	0	0	2	0	0	4	3	3	0	12
Bombus pascuorum	0	43	0	47	3	44	2	1	14	154
Bombus pratorum	0	29	0	13	0	1	0	0	3	46
Bombus sylvestris	0	0	0	1	0	1	1	0	0	3
Bombus terrestris	0	50	0	104	0	12	1	0	13	180
Total	7	163	36	328	21	128	13	26	37	759

3.3.2 Viral Prevalence

The prevalence of Mayfield virus 1, Mayfield virus 2, River Luinaeg virus, Loch Morlich virus was assayed per individual via PCR, see Chapter 3 Appendix Table A3.1 for primers and PCR conditions. All Mayfield virus 1 and Mayfield virus 2 samples were sequenced using an ABI Genetic Analyzer, and then aligned using the geneious de novo assembler (Kearse, Moir, and Cheung 2012), to ID them to the species level. A subset of the samples was additionally tested for Slow bee paralysis virus and Acute bee paralysis virus.

3.3.3 Diversity Analysis

To analyse sequence diversity, we used the raw reads from the RNA sequencing as discussed in detail in Chapter 2; briefly, these consist of 100bp-

paired end RNA-Seq data from pools of *B. terrestris, Bombus pascuorum* and *Bombus lucorum*, each sequenced twice, once using duplex specific normalisation and once using poly-A selection, and a pool of mixed *Bombus* species, sequenced only with poly-A selection. Mayfield virus 1, Mayfield virus 2, River Luineag virus, Lock Morlich virus, SBPV Rothamsted (EU035616.1) and ABPV (AF486072.2) sequences were aligned on the TranslatorX server (Abascal, Zardoya, and Telford 2010), using its MAFFT setting (Katoh and Standley 2013). Post-alignment, sequences were manually trimmed to the conserved region of the RdRp gene, minus eight codons, owing to the shortness of the River Luinaeg sequence. Trailing regions of 200 base pairs at both ends were retained so that reads were not prevented from mapping due to an overhang. This gave final sequence lengths of 1483, 1483, 1434, 1501, 1519 and 1522 base pairs for Mayfield virus 1, Mayfield virus 2, River Luineag virus, Lock Morlich virus, SBPV and ABPV respectively. Raw bioinformatic reads were trimmed in sickle version 1.33 using the default parameters (Joshi and Fass 2011). Overlapping mate reads were combined by FLASH version 1.2.11 using the default settings (Magoč and Salzberg 2011). Reads were aligned to the RdRp sequences generated above using MOSAIK version 2.1.73 (Lee et al. 2014). Both merged read and singletons from the sickle run were aligned together in the single end setting. Unmerged paired end reads were separately aligned using the paired end setting. In both cases a quality threshold of 30 was used to remove ambiguously mapping reads. SAM files were recombined after the fact using SAMtools version 1.5 (Li et al. 2009). Given the high coverage of SBPV, Mayfield virus 1 and Mayfield virus 2, duplicate sequences were not marked. Variants were then called using the default settings in LoFreq version 1.2.1 (Wilm et al. 2012). Base quality scores were then recalibrated using the outputted vcf file in GATK (DePristo et al. 2011). Variant calling and recalibration were repeatedly performed until the base quality scores converged to a stable distribution (total of four recalibrations), as there was no variant database to recalibrate quality scores off initially. Once the score distribution stabilised, variant calling was then performed again a last time to generate a set of variants for the entire sample. These variants were used to recalibrate the scores of each species-specific mapping, and generate species level variant calls. Given low numbers of mapping reads, several species-virus combinations were removed from the variant analysis. *B. lucorum* was analysed for SBPV

and ABPV. *B. terrestris* was analysed for SBPV, Mayfield virus 1 and Mayfield virus 2. *B. pascuorum* was analysed for ABPV, SBPV and Mayfield virus 2. The mixed *Bombus* pool was analysed for all viruses.

The number of polymorphic sites were calculated for each virus. Calculations were performed with and without the removal of variants with allele frequencies less than 5% to test the effect of the non-detection of very low frequency variants in lower coverage viruses. Variants with allele frequencies greater than 99% were removed as these represent fixed or nearly fixed differences from the underlying reference sequence.

A selection of the Mayfield virus 1 and Mayfield virus 2 Sanger sequences used to assign species identity for these viruses were aligned. Due to sequencing issues, a large proportion of the sequences, while sufficient for species confirmation, were of too low quality for haplotype analysis and were discarded, a threshold of greater than 5% sequence quality and less than 5% ambiguous bases was used. ABPV and SBPV sequences were amplified by PCR and sequenced using Sanger sequencing to explore between host diversity. The sequences were aligned using the geneious de novo assembler (Kearse et al. 2012), and error correction was performed manually. Haplotype networks using a Median Joining Network with an epsilon parameter of 0 were then estimated in POPART (Leigh and Bryant 2015).

3.3.4 Phylogenetics

A phylogeny was generated from the viral samples extracted from individual bees for Mayfield virus 1 in BEAST version 1.8.2 (Drummond et al. 2012), using 8 gamma categories both with and without a strict molecular clock, and with ambiguous regions being used in the calculation of the likelihood. Convergence was assessed by the comparison of the posterior distributions of the parameters of four separate runs in Tracer version 1.6 (Rambaut et al. 2013) starting from random starting trees. As the tree indicated several well-resolved clades within the virus, we tested for an association between viral type and host species post-hoc using BaTS (Parker, Rambaut, and Pybus 2008) to account for the phylogenetic uncertainty within the clades, using 1000 state randomizations for

the null distribution, having downsampled the posterior distribution of trees from BEAST to 1000 trees.

3.3.5 Community Similarity

We estimated the underlying sampling probability of each species of bumblebee at each site by treating the observed samples as being drawn from a multinomial distribution with 13 categories, corresponding to the 13 host species, using Bayesian estimation of the underlying probabilities with a Dirichlet prior with these 13 categories and a concentration parameter of 1 for each category, implying complete uncertainty about the underlying probability. Probabilities were estimated independently for each site. Ten thousand simulations were taken from the posterior distributions generated for each site to generate possible values of the underlying sampling probabilities of each bee species at each site, which we assume to be equivalent to the frequency of that bumblebee species at that site. For each of the 10000 simulations from the posteriors at the sites, we generated estimates of the community dissimilarity using the Morisita-Horn index (Horn 1966), implemented in the R package vegan (Oksanen et al. 2017). The posterior mode and 90% shortest probability intervals for the dissimilarity index were then reported.

3.3.6 Prevalence Estimation

The prevalence of each virus in each host at each location was estimated using the basic methods for the Bayesian estimation of a proportion, with a Beta (1,1) prior over the underlying probability. Posterior models and 90% SPIn intervals (Liu, Gelman, and Zheng 2015) were then calculated direct from simulations from the Beta distributed posterior.

3.3.7 Factors Influencing Infection

Climatic data for each of the 9 sites at which bees were collected was taken from the WorldClim database (Fick and Hijmans 2017). The variables were provided at 1km resolution. Data for July and August were extracted for mean daily maximum temperature, mean precipitation, mean solar radiation, mean vapour pressure and mean wind speed at the grip reference for the sites with a buffer area of 2km to account for the fact that bumblebees are known to forage over approximately that distance (Wolf and Moritz 2008). All values were

averaged to generate a consensus value for that site, and were then mean centred and scaled to unit standard deviation.

The individual prevalence data for Mayfield virus 1, Mayfield virus 2, River Luinaeg virus and Loch Morlich virus was analysed using Stan version 2.16 (Carpenter et al. 2017) using the rstan interface (Stan Development Team 2016a) in R version 3.3.2 (R Core Development Team 2016). A multivariate probit model was fitted, with random host and location effects and maximum temperature, precipitation, solar radiation, vapour pressure and wind speed as fixed effects in each virus. As the small number of locations from which samples were collected was small, we expected that our ability to accurately determine the size and direction of effects caused by ecological covariates would be limited. In order to counter this, we applied regularisation as recommended by Lemoine et al. (2016). The global intercept for each virus was given a Normal (0,10) prior, which does not substantially penalise low probabilities. Each fixed effect coefficient was given a Normal (0,1) prior, which, with this little data, should dominate the likelihood if the effect is small. Host and location random effects were drawn from normal distributions centred at 0 with estimated standard deviations. In both cases, the standard deviations were given Cauchy (0,25) hyperpriors, which are only weakly informative but proper. The correlation in residuals for the multivariate normal was given a flat prior between -1 and 1 using a LKJ prior with shape parameter 1.

3.4 Results

3.4.1 Community Similarity

There is a clear discontinuity between the South and North of Scotland in community structure, with locations in the south having *B. terrestris*, *B. pascuorum* and *B. lucorum* dominated communities, and communities in the North having *Bombus jonellus* and *Bombus hortorum* dominated communities, as shown in Fig. 3.1. Table 3.2 shows the dissimilarity indexes between the sites, and the same effect is also observed there. Two potentially surprising results stand out. The Pentlands appears to represent a third type of community, separate from the North-South divide described above. The presence of *Bombus monticola,* otherwise only found in the highland sites, and

an equivalent frequency of *B. pascuorum* and *B. lucorum* makes the community look like a blending of Northern and Southern community types. Surprisingly, Iona's community differs strongly from Staffa's, despite their close proximity, with Iona having considerably fewer *B. jonellus* and many more *B. hortorum*.

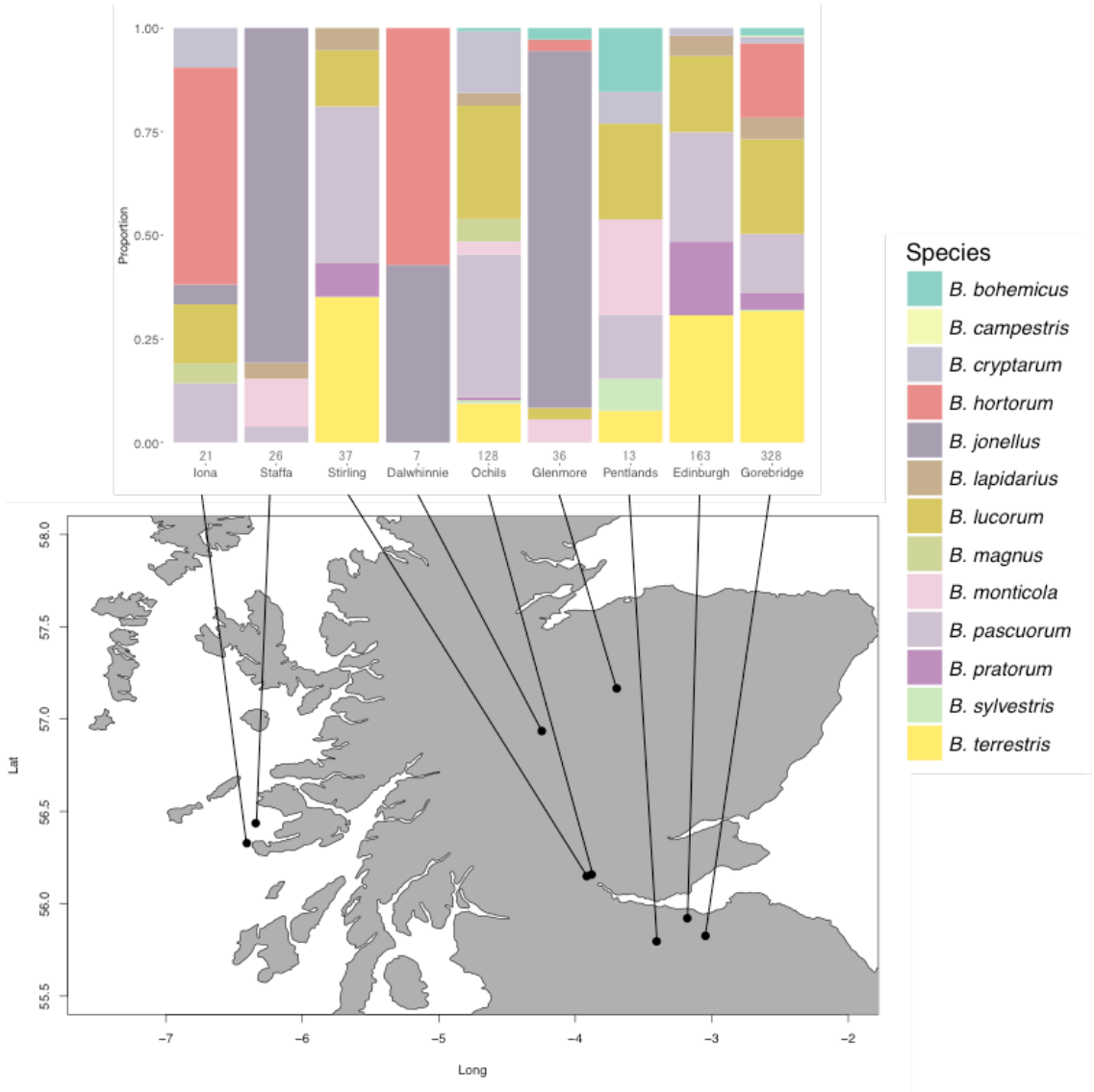


Figure 3.1 The locations of the sampling sites and species distributions of the bumblebees caught at them. Sample sizes for each site are shown above the site names.

Table 3.2 The Morisita-Horn dissimilarities of the different sites. 90% shortest posterior density intervals for the index are in brackets.

	Dalwhinnie	Edinburgh	Glenmore	Gorebridge	Iona	Ochils	Pentlands	Staffa	Stirling
Dalwhinnie	0								
Edinburgh	0.764 (0.548, 0.909)	0							
Glenmore	0.491 (0.228, 0.774)	0.934 (0.855, 0.971)	0						
Gorebridge	0.548 (0.355, 0.733)	0.162 (0.104, 0.217)	0.918 (0.840, 0.964)	0					
Iona	0.220 (0.079, 0.501)	0.653 (0.494, 0.819)	0.817 (0.629, 0.914)	0.415 (0.278, 0.532)	0				
Ochils	0.741 (0.527, 0.894)	0.233 (0.150, 0.325)	0.920 (0.837, 0.962)	0.315 (0.243, 0.415)	0.569 (0.336, 0.737)	0			
Pentlands	0.591 (0.357, 0.795)	0.501 (0.278, 0.674)	0.849 (0.661, 0.934)	0.440 (0.272, 0.663)	0.517 (0.324, 0.748)	0.394 (0.160, 0.572)	0		
Staffa	0.491 (0.190, 0.742)	0.893 (0.780, 0.952)	0.025 (0.004, 0.104)	0.898 (0.802, 0.959)	0.831 (0.602, 0.912)	0.881 (0.754, 0.943)	0.782 (0.572, 0.898)	0	
Stirling	0.780 (0.502, 0.879)	0.070 (0.019, 0.144)	0.907 (0.806, 0.964)	0.174 (0.092, 0.298)	0.637 (0.451, 0.812)	0.213 (0.105, 0.357)	0.450 (0.242, 0.676)	0.853 (0.721, 0.939)	0

3.4.2 Prevalence

Figure 3.2 shows the prevalences of these viruses for each location-species combination measured. When broken down to the specific host-location level, sample sizes for many species become small, so the uncertainty around the modal prevalences is correspondingly large. However, it appears clear that the same species may be infected to a different level with the same virus between

sites. With the exception of SBPV, which is moderately to highly prevalent in all species at all sites, the levels of prevalence depend strongly on both the site and the species, to different degrees depending on the virus.

River Luinaeg virus was present in *B. jonellus* at all sites where the species was sampled, with prevalences of approximately 25% or higher detected at multiple sites. The prevalence was similarly high in *Bombus pratorum,* consistent with the group level estimates presented in Chapter 2. Intermediate prevalences were detected in *Bombus cryptarum*. Low levels of infection with RLV were detected in *B. lucorum* with the prevalences of the virus appearing to be considerably higher in this species in Stirling and the Pentlands. Loch Morlich virus appears to exhibit much higher species specificity with almost all the detections being in *B. jonellus* (13/16) or being coincident with RLV infection (13/16). No species other than *B. jonellus* was infected with LMV without coinfection with RLV. The prevalence of Mayfield virus 1 showed a strong interaction between the host species and site. Edinburgh and Gorebridge, two sites around 15km apart with large sample sizes, have dramatically different prevalences MV1 prevalences in *B. terrestris, B. pratorum* and *B. pascuorum*, with the prevalences being between 30-60% in Edinburgh, and likely being below 15% in all species in Gorebridge. On the species level, MV1 appears to be a true generalist virus, with only *B. jonellus* having relatively certain low prevalence. Mayfield virus 2 shows a similar pattern, but without any obvious differences in infection levels between sites. The prevalences are generally lower across the board in MV2 than MV1, but beyond that, the range of species infected is largely similar. The sample tested for ABPV was considerably smaller than that of the other viruses, but APBV was found at intermediate modal prevalences, of above 10%, in all species apart from *B. terrestris* and *B. lucorum*.

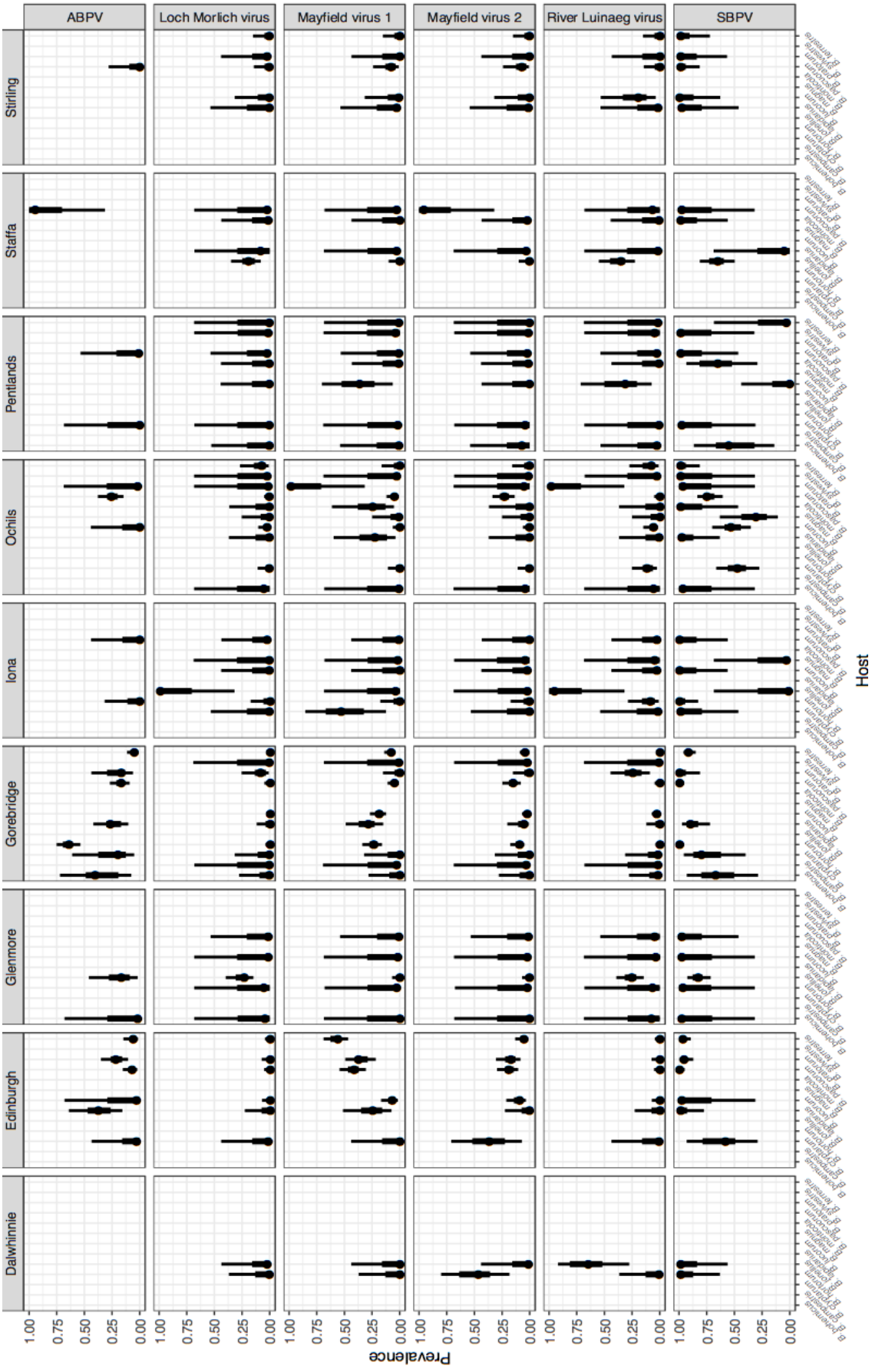


Figure 3.2 The prevalences of ABPV, SBPV, Loch Morlich virus, River Luinaeg virus, Mayfield virus 1 and Mayfield virus 2 in each sampled host species in each site. The point estimate is the

posterior mode, with 50% shortest posterior intervals represented by the thick lines and 90% shortest posterior intervals represented by the thin lines. Untested combinations are left blank.

3.4.3 Factors Influencing Infection

Considering the application of regularisation as described in the methods, the following section should be interpreted with the proviso that this is intended to be an exploratory analysis. Covariates that had estimates shrunk towards zero are unlikely to truly have no effect on prevalence. Given this, under the assumption that no ecological covariate has truly no effect and equal positive and negative prior probabilities for the effect sign, we report the posterior probabilities that the effect is positive or negative for each. These probabilities are shown in Table 3.3, posterior intervals for raw parameter estimates are shown in Figure 3.3.

For most covariates, the relative probabilities that the covariate caused a positive or negative effect after regularisation was roughly equal. Increasing precipitation had a high posterior probability of having a positive effect on the prevalence of River Luinaeg virus (97%), and there was some evidence that the decreasing of precipitation decreased the prevalence of Mayfield virus 1 (90%). There was also weak evidence that higher maximum temperatures and wind speeds similarly increased the prevalence of Mayfield virus 1 (87% and 82% respectively).

Multivariate probit models also allow the calculation of the correlation in the error terms of the multivariate normal latent variable. This measures the degree to which, after accounting for the predictors, there is still shared error, as caused by unobserved factors effecting risk. Posterior correlations are shown in Table 3.4. In this case, it measures the extent to which there is excess coinfection after accounting for the location of sampling, the species of which the bees belong and the various location-level environmental variables. The only error correlation where the 90% shortest posterior interval did not overlap 0 was the error correlation between RLV and LMV, which was strong and positive, consistent with the high levels of coinfection noted above.

Table 3.3 The posterior probabilities of the signs of the effects of the covariates in the table being positive for each virus.

	Precipitation	Radiation	Maximum Temperature	Vapour Pressure	Wind Speed
River Luinaeg virus	0.97	0.47	0.49	0.32	0.43
Loch Morlich virus	0.65	0.45	0.54	0.56	0.52
Mayfield virus 1	0.10	0.75	0.87	0.31	0.82
Mayfield virus 2	0.42	0.51	0.46	0.51	0.29

Table 3.4 The posterior correlations of the errors of each virus from the multivariate probit model. 90% shortest posterior intervals for each correlation are shown in brackets.

	River Luinaeg virus	Loch Morlich virus	Mayfield virus 1	Mayfield virus 2
River Luinaeg virus	1			
Loch Morlich virus	**0.784** (0.594, 0.911)	1		
Mayfield virus 1	0.043 (-0.416, 0.336)	-0.188 (-0.703, 0.217)	1	
Mayfield virus 2	-0.604 (-0.855, 0.236)	-0.366 (-0.919, 0.211)	0.015 (-0.121, 0.240)	1

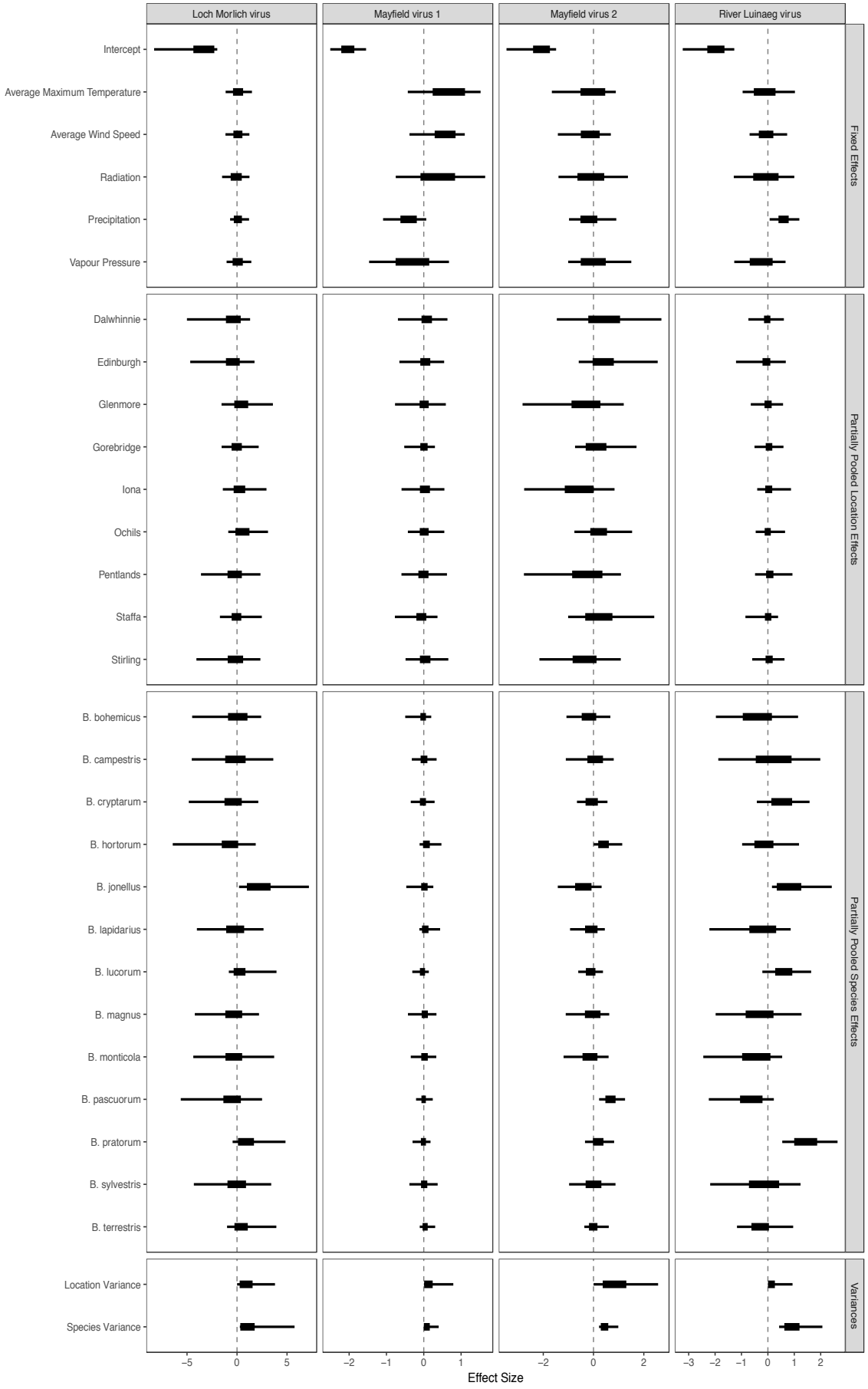


Figure 3.3 The estimates for each parameter in each virus from the multivariate probit model. Thick lines represent 50% shortest posterior density intervals. Thin lines represent 90% shortest posterior density intervals.

3.4.4 Diversity

Haplotype networks were generated for the Sanger sequenced viral samples. As is clearly visible in Figure 3.4, there is considerably more diversity in MV1 and MV2 than is present in SBPV, and more again than is present in ABPV, which showed no sequence variation over the Sanger sequenced region after heterozygous sites were removed. The fact that the same genotypes of MV1 and MV2 are observed in both 2009 when Dalwhinnie, the Ochils and Iona were sampled and 2011 when Edinburgh, Gorebridge and the Pentlands were sampled, implies that the diversity is stable over short periods. This pattern remained when the SNPs called from the raw RNA sequencing data were considered. Over homologous genomic regions within the RdRp, 12.7% of sites in RLV, 11.3% of sites in MV2, 8.4% of sites in MLV, 7.0% of sites in MV1, 1.1% of sites in SBPV and 0.3% of sites in ABPV exhibited polymorphism (defined as the presence of minor alleles at an allele frequency of greater than 5%).

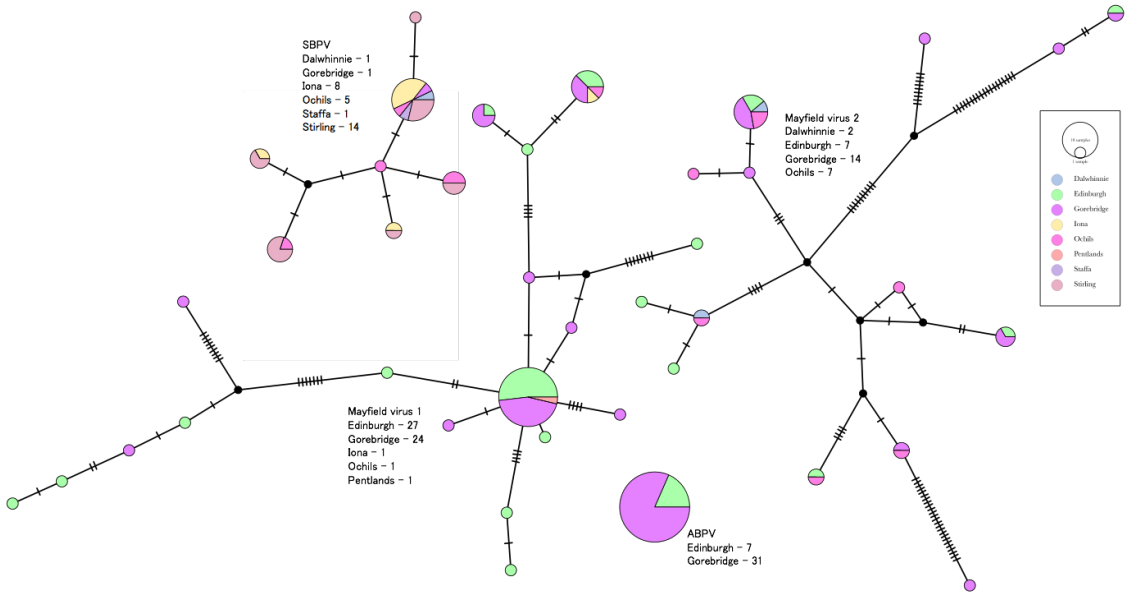


Figure 3.4 Median-joining haplotype networks for Mayfield virus 1, Mayfield virus 2, ABPV and SBPV. Estimated using an epsilon parameter of 0. The dashes represent the number of mutations between sequence, and colours correspond to the sampling location.

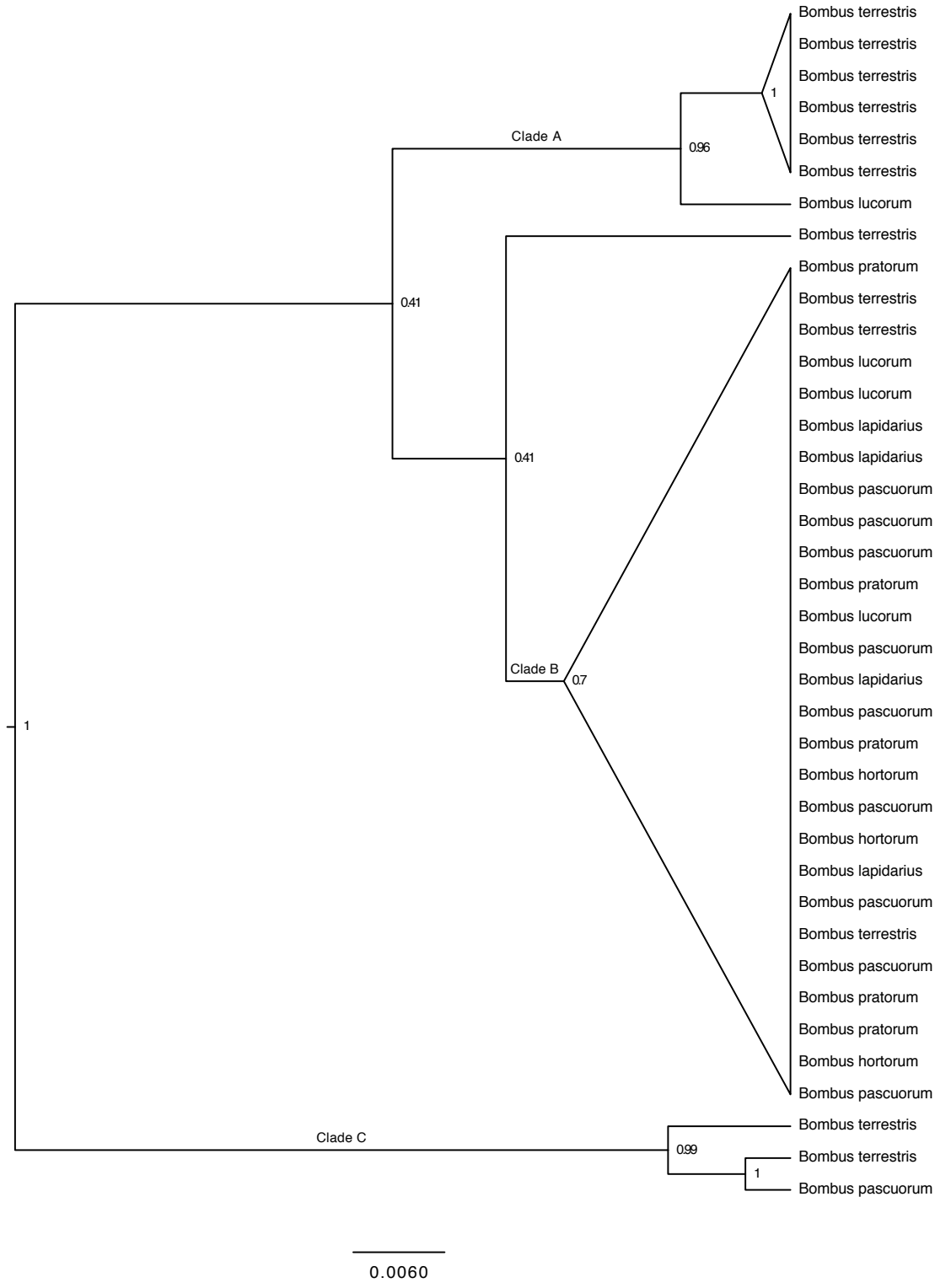


Figure 3.5 A Bayesian phylogenetic tree of the Mayfield virus 1 isolates with their host species. Posterior clade probabilities are shown at the nodes. Non-resolvable polytomic clades were collapsed. Clade assignments represent arbitrary splits for the purposes of discussion.

Post-hoc testing of the Mayfield virus 1 tree (Fig 3.5) appeared to show a clear association between viral type and host species, with the group marked on the tree as Clade A having an excess of *B. terrestris* isolates (AI statistic, p=0.003), and Clade B having fewer than expected *B. terrestris* infections. When the tree

was rerun using only isolates from Gorebridge, to control for a possible effect of spatial structure, the same result was found (AI statistic, p=0.007).

3.5 Discussion

In this study, we have explored the ecological factors influencing the diversity and distribution of the viruses of wild bumblebees. We found evidence for cryptic specialisation within viral species for parts of their host range, and significant differences in prevalence between host species (consistent with Chapter 2). We found that the viruses that have only been detected in bumblebees are considerably more diverse than those previously known from honeybees.

3.5.1 Diversity

Both Acute bee paralysis virus and Slow bee paralysis virus show considerably less diversity than Mayfield virus 1, Mayfield virus 2, River Luinaeg virus and Loch Morlich virus within the study region. ABPV and SBPV are viruses that were initially described in honeybees (Bailey and Gibbs 1964; Bailey et al. 1963), while the other four viruses in the study were found in bumblebees and have not been recorded in honeybees at this point. In multihost systems, different species can differ in their susceptibility and response to an infection (Ruiz-González et al. 2012). Including this as an assumption when modelling multihost pathogens leads to the result that species have very different levels of importance for the maintenance of an infection (Fenton and Pedersen 2005; Gandon 2004). Superspreading dynamics have been observed or inferred in many systems, from small mammals infected with intestinal parasites (Streicker et al. 2013), to large ungulates infected with tuberculosis (Santos et al. 2015), and human sexually transmitted infections (Renton et al. 1995). In these cases, a single heavily infected host species (or population) can act as a source for infection in other sympatric species (or population). It is possible that the ABPV and SBPV isolates detected here might represent spillover from managed honeybee populations, which subsequently lead to epidemic spread, especially as both viruses are very close to the honeybee references on genbank. However, McMahon et al. (2015) found that SBPV is often found at higher prevalences in bumblebees than in sympatric honeybees, bringing in to the

question the direction of the spillover dynamics underlying the system. The patterns observed in SBPV and ABPV are, however, not inconsistent with the genetic diversity that may be observed during an epidemic with a bottlenecking or strong selective event (Agoti et al. 2014; Hapuarachchi et al. 2016), as while the prevalence of both SBPV and ABPV were consistent between sampling years, all sequences isolates were from within years.

The large amounts of diversity in Mayfield virus 1, Mayfield virus 2, River Luinaeg virus and Loch Morlich virus could have multiple causes. MV1 genotypes appear to non-randomly associate with specific host species. This could be driven by different hosts having differential contact with viral strains, or could represent functional differences between viral variants with some variants being able to replicate more efficiently in specific hosts. This potentially implies that at least some of the diversity within MV1 leads to biologically relevant outcomes. It has previously been shown with multiple viruses that the precise viral strain that infected an individual has important phenotypic effects on outcomes of infection (Ilonen et al. 1988; Mayerat, Mantegani, and Frei 1999). It has also been shown that pathogen genotype can correlate with the probability of being found in different host species (Cuevas et al. 2012; Withenshaw et al. 2016). If the variants do have different fitnesses in different hosts, this would increase the viral diversity as more viral types would be maintained by selection. It is also possible that much of the observed variation is neutral, as most variation is at 3rd codon positions and codes for either identical or similarly charged amino acids, which are unlikely to have large fitness effects in either direction, and given the frequent bottlenecking that occurs during transmission (Zwart and Elena 2015) would be inefficiently selected against.

3.5.2 Factors Influencing Infection

We found evidence that the prevalence of River Luinaeg virus was positively associated with increased precipitation, though the size of this effect is uncertain given the bias added by the regularising prior. The direction of this effect is contrary to our hypothesis that higher rainfall would decrease prevalence by reducing the contact rate between the host and virus through mechanical cleaning of contaminated flowerheads and decreased bumblebee flight frequency. However, an alternative explanation consistent with the results

is that high precipitation reduces foraging time and therefore condition, mediated through starvation. Starvation is known to increase the severity of bumblebee viruses (Manley et al. 2017) and has been shown to increase infection risk in humans (reviewed in França et al. 2009; Schaible and Kaufmann 2007) and other mammals (Pedersen et al. 2002). However, we found some evidence of a negative effect of precipitation on Mayfield virus 1, which is consistent with our original hypothesis. We attempted to fit a host community similarity matrix as a random effect in our model, but there were not enough sites to accurately estimate to what extent similar pollinator communities share prevalences. None the less, it seems intuitive to posit that communities that have high proportions of highly susceptible species should have higher viral prevalences in all species, given the ability of highly infected susceptible species to act as source populations for infection (Streicker et al. 2013).

3.5.3 Coinfection

River Luinaeg virus and Loch Morlich virus were rarely found separately in this study and they remained strongly correlated in the errors of the multivariate probit model after taking into account the explanatory variables. Multivariate probit models can be thought of as an extension of the models recommended in Fenton et al. (2010) for estimating interactions between parasite species, though it is only applicable to presence/absence data rather than the faecal egg counts used there. Instead of sequentially estimating the effect of sympatric parasites on a focal parasite, the joint probabilities of presence or absence of all parasites simultaneously are modelled by a multivariate normal distribution, with the values in the correlation being direct estimates of the association strength and direction between the parasites after controlling for the covariates. While the correlations between Mayfield virus 2 and both Loch Morlich virus and River Luinaeg virus did trend negative, given the small sample size relative to the prevalence of Loch Morlich virus and River Luinaeg virus, they could not be estimated precisely.

A potential explanation for the strong association between Loch Morlich and River Luinaeg virus is that one of the viruses is a satellite of the other, as occurs in Chronic bee paralysis virus with Chronic bee paralysis virus satellite virus

(Bailey, Ball, et al. 1980). However, on further inspection this seems unlikely as both virus species are observed in single infections. Another possibility is that both viruses circulate in the population, but infection with one causes damage to the host in such a way that susceptibility to the second is dramatically increased, perhaps a manner similar to HIV's synergism with TB though immune suppression (Kwan and Ernst 2011) or influenza virus' changing of the environment of the nasopharynx so as to allow secondary bacterial invasion (Joseph, Togawa, and Shindo 2013). Viral coinfections are ubiquitously reported in prevalence studies in bees (Anderson and Gibbs 1988; Bacandritsos et al. 2010; Blažytė-Čereškienė et al. 2016; Chen et al. 2004; Choe et al. 2012; Evans 2001; Gajger et al. 2014; McMahon et al. 2015; Mouret et al. 2013; Nielsen, Nicolaisen, and Kryger 2008; Roberts, Anderson, and Durr 2017; Thu et al. 2016), but to our knowledge, only McMahon et al. (2015) tested for a departure from random expectations of infection, and no departure was found. However, non-random associations between parasites appear common in other species, having been reported in, among other species, humans (Griffiths et al. 2011), wood mice (Behnke et al. 2005), buffalo (Jolles et al. 2008), typical white-eyes (Clark et al. 2016), ticks (Václav et al. 2011), moths (Hajek and van Nouhuys 2016) and plants (Biddle, Linde, and Godfree 2012; Seabloom et al. 2009).

3.5.4 Conclusion

The importance of bumblebees both economically and ecologically makes a good understanding of their infections valuable, as they may play an important part in the declines currently observed. In this study, we investigated the ecology of four recently described bumblebee viruses (Mayfield virus 1, Mayfield virus 2, River Luinaeg virus and Loch Morlich virus) and compared their diversity to two previously described viruses: Slow bee paralysis virus and Acute bee paralysis virus. We find evidence that the probability of infection may be modified by the levels of precipitation in the areas in which the host-parasite communities exist implying that in order to get a realistic perspective on the underlying prevalences of bee viruses sampling over a wide range of ecological conditions is required. We detected no strong associations between the prevalence of the four recently described viruses and solar radiation, maximum temperature, vapour pressure or wind speed. The presence of non-random

associations between parasites also implies that simultaneous testing for multiple parasites in samples may be the only way to accurately assess the viral community structure. The recently discovered viruses are more diverse than SBPV and ABPV, however the reasons for this are unknown. There is also evidence that variants of Mayfield virus 1 assort non-randomly with hosts, potentially representing cryptic host specialisation within the virus species. Given the key role that viruses play in the regulation of natural populations and the importance of pollinators, we know very little about the ecological factors that predict infection. This study represents an important first step in isolating predictors of viral prevalence in wild bumblebees.

Chapter 4 - Pulsed pesticide exposure may increase the rate of pathogen uptake in *Bombus terrestris* bumblebees

David J. Pascall containing additional work from Meri Anderson and Thomas Marceau

4.1 Abstract

Bees are economically and ecologically important as pollinators, and many species are currently suffering declines. Global bee declines are caused by a complex interaction between multiple stressors; two key stressors are pathogens and pesticides. It is becoming increasingly clear that neonicotinoid pesticides negatively affect wild bees, but the scale and importance of this damage remains unclear. Evidence is mounting that pathogens and parasites interact synergistically with pesticide exposure increasing the impact of infection. However, it is less clear how pesticide exposure increases the risks of infection. We exposed 20 buff-tailed bumblebee (*Bombus terrestris*) colonies to the neonicotinoid pesticide clothianidin in the laboratory for two weeks and then placed them in the field. This mimicked a pulsed pesticide exposure, as potentially experienced by bees during the bloom of mass-flowering pesticide-treated crops such as oilseed-rape. We then measured their acquisition of a panel of 20 pathogens over an eight-week period. We found marginal evidence for increases in the rate of acquisition under pesticide exposure. This presents new evidence that the use of neonicotinoid pesticies may be lead to bee mortality by increasing disease susceptibility

4.2 Introduction

Pollinators, and particularly bumblebees and honeybees, are key players in global food production, with insect pollination contributing 15.12 billion USD to the US economy in 2009 (Calderone 2012). Many of these pollinators have declined in abundance in recent years (Kosior et al. 2007; Williams and Osborne 2009),while simultaneously, the need for pollination services has increased (Breeze et al. 2014). This decline is driven by a suite of anthropogenic and environmental factors acting in combination (Brown et al. 2016; Vanbergen and the Insect Pollinators Initiative 2013). Besides habitat loss, the increased use of pesticides and fungicides are considered to be an important contributor (Vanbergen and the Insect Pollinators Initiative 2013). Pathogens and parasites have also been implicated, with the movement of infected bees causing the spread of disease in wild populations (Goka, Okabe, and Yoneda 2006; Manley et al. 2015; Wilfert et al. 2016).

The rate of uptake of pathogens from the environment into colonies appears likely to be a strong determinant of the degree of stress that pathogens impose onto bumblebee populations. Bumblebees face multiple pathogens in the wild and there is a stochastic element to how they encounter them. In the tropics, some species produce sexuals all year round (Michener and Amir 1977), but in temperate regions, bumblebees have seasonal colonies, with queens hibernating over winter (Goulson 2010), which puts a hard limit on the time period in which pathogens can affect them. If the transmission of a pathogen species to a seasonal colony is a slow process because of infrequent contact, then that pathogen is unlikely to cause strong effects that species, because on average, colonies will be infected towards the end of the colonies natural lifespan. This is complicated by the fact that sexuals are produced towards the end of a colony's life. Thus, a highly virulent pathogen that is contracted late could still significantly decrease the reproductive output of a colony. But in general, late contracting pathogens seem liable to have less serious impacts. Despite this, little is known about the rate at which naïve colonies uptake pathogens from the environment.

The stress of infection can also be compounded by stress from other sources. Pathogens in honeybees have been shown to act synergistically with pesticides, causing increases in mortality (Alaux et al. 2010; Doublet et al. 2015; Vidau et al. 2011), and infection intensity (Pettis et al. 2012; Di Prisco et al. 2013). Many parasites and pathogens exhibit condition-dependent virulence, where adverse effects are only expressed if the cost of infection is unmasked under stressful conditions. Without resource constraint, as in under *ad libitum* conditions, the cost in resources to both resist disease and repair the damage from infection is affordable. However, if resources are limited, different aspects of pollinator fitness must be traded off against one another, as has been demonstrated both at the individual and colony level in *Bombus terrestris* bumblebees (Moret and Schmid-Hempel 2000, 2001). For example, bumblebee longevity is often not affected by infectious diseases under standard laboratory conditions, but is drastically reduced under starvation (Brown, Loosli, and Schmid-Hempel 2000; Manley et al. 2017). The exact form that interaction of stressors will take depends on the underlying mechanism. A broad review of toxicant-environment

interactions found that they can be both synergistic and antagonistic, but synergistic interactions dominate (Holmstrup et al. 2010). Similarly, both antagonistic and synergistic interactions have been observed in cases of superinfection, with synergistic interactions dominating when coinfection occurs between different classes of parasites, and both occurring about equally when the parasites are of the same type (Griffiths et al. 2011; Kotob et al. 2016). The exact form of these superinfection interactions is also context-dependent and can change based on environmental and host factors (Zheng et al. 2015). Given this uncertainty, it is important to gather data about the nature and direction of these interactions in the wild, rather than in the lab setting.

One particularly important class of stressors in pollinators are neonicotinoid pesticides, the class including the commonly used seed treatments imidacloprid, thiomethoxam and clothianidin, which have been used commercially since the late 1990s (Goulson 2013). Questions over their safety for wild pollinators have led to a moratorium on their use inside the European Union (Barroso 2013), but they are still used extensively in other parts of the world (Tsvetkov et al. 2017). These pesticides have negative effects on both honeybee (Henry et al. 2012; Di Prisco et al. 2013) and bumblebee (Laycock et al. 2012) physiologies leading to adverse outcomes at the colony level (Whitehorn et al. 2012). While negative outcomes in wild and managed bumblebee and honeybee populations from exposure to these pesticides have been repeatedly demonstrated (Doublet et al. 2015; Gill, Ramos-Rodriguez, and Raine 2012; Henry et al. 2012; Laycock et al. 2012; Pettis et al. 2012; Di Prisco et al. 2013; Retschnig, Neumann, and Williams 2014; Vidau et al. 2011; Whitehorn et al. 2012), the effect that exposure to pesticides has on the uptake and risk of infection from natural parasites in the wild has been less quantified. Most studies that conclusively demonstrate the effect of pesticides on the end points of infection take place in laboratory conditions, which may not generalise to the field given the complex buffering interactions exhibited by natural systems (Park et al. 2015).

Here we tested the hypothesis that, on average, pesticide exposure would increase the rate of pathogen uptake and that this effect would be most pronounced immediately after the exposure. We explored both the rate of

uptake of pathogens from the environment in bumblebee colonies and the effect of neonicotinoids on infection rate in a semi-field setting. We exposed colonies to either sugar water or sugar water containing low (1 ppb) concentrations of clothianidin in the laboratory, a design similar to Whitehorn et al (2012) that mimics the pulsed exposure that occurs when a large seed-treated crop plant blooms. We then placed the colonies outside at multiple locations and measured their uptake of a large panel of pathogens from the environment over an eight-week period.

4.3 Methods

4.3.1 Laboratory Exposure

Twenty buff-tailed bumblebee, *Bombus terrestris audax*, colonies were purchased from Biobest ltd. On arrival, each colony received *ad libitum* sugar water, and two teaspoons of radiation sterilised pollen every second day. Ten colonies were randomly assigned into two blocks, A1 and A2. Each block was randomly further split into two treatment groups, pesticide and control. All the following experimental procedures described were carried out on the A1 colonies on one day prior to the A2 colonies to allow adequate time for sample collection once the colonies had been placed out into the field.

The clothianidin sugar solution was generated by dissolving solid clothianidin in acetone to generate a 7.5mg l^{-1} solution, which was then diluted 1:7500 in sugar water generated from a 50:50 dilution of invertebrate sugar syrup and MilliQ H_2O, to generated the 1ppb final solution. A control solution was created by the addition of an equal amount of acetone without dissolved clothianidin to the same stock sugar solution. Immediately prior to exposure, 10 bees were removed from each colony and faecal samples were collected and frozen at -80 degrees centigrade. To account for differences in sugar water consumption, the initial weights of each colony's sugar water containers were taken before treatment. Colonies were allowed to feed *ad libitum* on the sugar solution for two weeks. At the end of the exposure period, each sugar water container was reweighed to measure how much the colony had consumed and a further 10 faecal samples were taken from each colony.

4.3.2 Field Placement

Treatment and pesticide colonies within each group were randomly paired and assigned to one of 10 locations within Falmouth, Cornwall, UK. For reasons of practicality of access, a full randomisation of locations between the two groups was not performed, and sites within a block were closer together than would be expected if sites were assigned fully at random (see Figure 1). The colonies were placed at these locations under bespoke wooden shelters with slanted waterproofed roofs to protect them from the rain. At each site, one shelter had a blue circle painted on it and the other had a green triangle in order to provide a landmark to minimise between colony drifting. Each colony was placed on bricks to avoid surface water entering and weighed down with bricks from above to prevent the colony blowing away in high wind. Every two weeks, five bees were removed from each colony and frozen at -80 degrees centigrade to preserve the RNA. If less than five bees remained in a colony, all were removed. At day 42 in the field, in one of the pesticide exposed colonies only four *B. terrestris* were collected, as when the samples were taken back to the lab and examined, one of the bees collected was observed to be a brood parasitic bee from the subgenus *Psithyrus*. This was the only colony in which brood parasites were detected.

Figure 4.1 The map of locations of the sites at which colonies were placed. A 250m buffer around each site is shown. All sites were within 1km of at least one other site. The five southernmost sites comprise block A1 and the five northernmost sites comprise block A2.

4.3.3 Molecular Work

Pooled RNA extractions were performed using TRIzol (Life Technologies) for each colony at each timepoint. Both samples from before the bees were placed into the field were combined into a single day zero timepoint. Each of these pools were tested for a panel of twenty pathogens known to infect *B. terrestris*; Allermuir Hill virus 1 (AHV1), Boghill Burn virus (BBV), Castleton Burn virus (CBV), Clamshell Cave virus (CCV), Gorebridge virus (GV), Hubei partiti-like virus 34 (HPLV34), Mayfield virus 1 (MV1), Mayfield virus 2 (MV2), Mill Lade virus (MLV), Acute bee paralysis virus (ABPV), Slow bee paralysis virus – Rothamsted strain (SBPV-R), Slow bee paralysis virus – Harpenden strain (SBPV-H), Black queen cell virus (BQCV), Deformed wing virus – type A (DWV-A), Deformed wing virus – type B (DWV-B), Sacbrood virus (SBV), *Nosema ceranae*, *Nosema apis*, *Nosema bombi* and *Crithidia bombi*. The primers and PCR conditions used are available in Chapter 4 Appendix Table A4.1.

4.3.4 Statistical Analysis

The mass of sugar water consumed by the pesticide and treatment colonies was compared using a Gaussian linear model with an identity link in R v3.3.2 (R Core Development Team 2016) using the function 'stan_glm' in package rstanarm (Stan Development Team 2016b) with 5000 warmup draws and 95000 sampling draws across four chains. The data were transformed by subtracting the maximum value + 1 from each observation then square rooting to remove negative skew. Results are given on the back transformed original scale. The priors assigned were Normal (0, 100) for the intercept parameter, Normal (0, 1) for the fixed effect coefficients (representing mild regularisation), and half-Cauchy (0, 2.5) for the standard deviation of the normal distribution. All priors were then scaled by the observed standard deviation of the data, as is the default in rstanarm. The model fitted was:

Sugar water consumed ~ Treatment group + Block

The data for all pathogens that were detected in at least one colony were analysed in MCMCglmm (Hadfield 2010) in R v3.3.2 (R Core Development Team 2016). A series of models were run using the 'nzbinomial' link function to account for the pooled nature of the data. Four models were fitted, treating the data in two different ways. Models 1 and 3 treated the time as a continuous variable, while Models 2 and 4 ignored the continuous nature of time, treating each timepoint as an independent factor. Models 1 and 2 treated the amount of pesticide consumed by each colony as a continuous dose, while Models 3 and 4 treated pesticide as a factor with levels exposed or not exposed. As timepoint 0 was taken as a baseline, when time was treated as a continuous variable, no intercepts were fitted for the pesticide treatment, whether coded as a binary treatment variable or continuous dose variable, nor for the block assignment. These effects were instead estimated as interactions with timepoint representing treatment effects on the rate of update of pathogens over time. When timepoints were treated as factors, the time 0 data was excluded, and intercepts were fitted for both the pesticide treatment and the block.

All models were run with a relatively uninformative parameter expanded prior with V equalling an identity matrix the size of the covariance matrix being

estimated, *nu* equalling the dimension of *V*, *alpha.nu* being a vector of zeros of length *nu* and *alpha.V* being a matrix of the size of the covariance matrix being estimated with 1000 repeated along the diagonal. Models were run for 22000000 iterations including 2000000 iterations of burn-in, and samples were saved every 20000 iterations for a final sample size of 1000 draws from the posterior.

In the models with continuous time, a random timepoint by location interaction was fitted, a random timepoint by colony interaction nested within location and a random virus effect with an estimated intercept and random slopes by location. This accounted for the fact that at day 0 all colonies were in the lab, so the intercept for all locations and colonies was considered to equal the global intercept, as in theory there has been no contact with pathogens to allow uptake at this point. Therefore, all differences driven by different locations and colonies having different uptake rates must be driven by changes in the rate of uptake after day 0, an interaction with time, explaining the random effect formulation above. As some pathogens were detected in the colonies at day 0, a random pathogen specific intercept was fitted for each pathogen, then pathogens were allowed to have different uptake rates through time. The covariance between the pathogen intercepts and slopes was estimated to investigate the possibility that the pathogens detected at day 0 had different uptake patterns than the other pathogens.

In the models where timepoints where analysed separately with no relation between them and the day 0 data was excluded, the location, colony and virus were fitted as random intercepts, and a pathogen-pesticide treatment random interaction was added to test whether different pathogens responded differently to the pesticide treatment.

The models run were as follows:

Model 1 - continuous time, continuous pesticide dose

y_i ~ intercept + time(continuous) + time(continuous):pesticide(dose) + time(continuous):block + (time(continuous)|Location) + (time(continuous)| Location/Colony) + (1 + time(continuous)|Pathogen)

Model 2 - discrete timepoints, continuous pesticide dose

y_i ~ intercept + timepoint + + pesticide(dose) + block + timepoint:pesticide(dose) + (1|Location) + (1 |Location/Colony) + (1 + pesticide(dose)|Pathogen)

Model 3 - continuous time, discrete pesticide treatment

y_i ~ intercept + time(continuous) + time(continuous):pesticide(factor) + time(continuous):block + (time(continuous)|Location) + (time(continuous)| Location/Colony) + (1 + time(continuous)|Pathogen)

Model 4 - discrete timepoints, discrete pesticide treatment

y_i ~ intercept + timepoint + pesticide(factor) + block + timepoint: pesticide(factor) + (1|Location) + (1 |Location/Colony) + (1 + pesticide(factor)|Pathogen)

All reported shortest posterior intervals were calculated in the SPIn package(Liu et al. 2015).

4.4 Results

4.4.1 Pesticide Consumption

The estimate of differential consumption of sugar water between the two treatment groups was highly uncertain with a posterior mode of -72.07g (90% SPI: -148.05g, 7.44g) change in the pesticide group over the 14-day period relative to the control group. This provides little evidence of an effect. However, given the uncertainty in the estimate, if an effect exists, it is potentially large with the 90% shortest posterior interval, being equivalent to a modal average daily change of -5.15g (90% SPI: -10.57g, 0.53g) in consumption, which is more than a 20^{th} of the total sugar water consumed over the entire two-week period for some colonies (Figure 4.2).

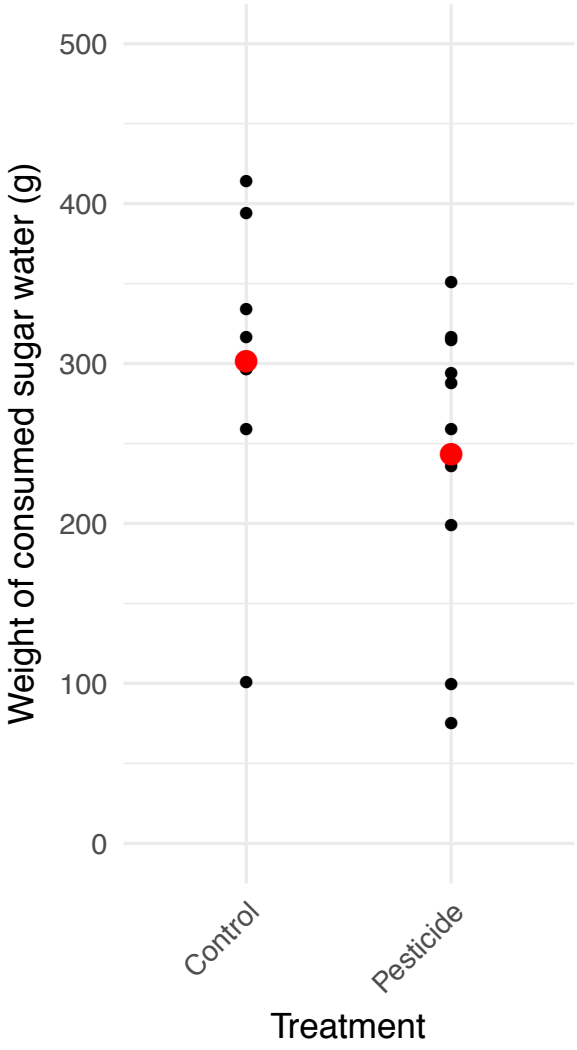


Figure 4.2 The mass of sugar water consumed by each treatment group

4.4.2 Survival and Health of Colonies Post-placement

Colonies failed throughout the experiment, with the rate of failure increasing rapidly after 28 days outside, see Figure 4.3. Losses were roughly equivalent between treatment groups. One colony in the pesticide treatment was infested by a bumblebee brood parasite from the subgenus *Psithyrus* during the experiment (detected at day 42 post-placement), but the colony persisted until the final timepoint, when one *B. terrestris* worker remained. In total, three collections consisted of less than the full five bees due to less than five workers remaining in the colony (2 controls, 1 pesticide).

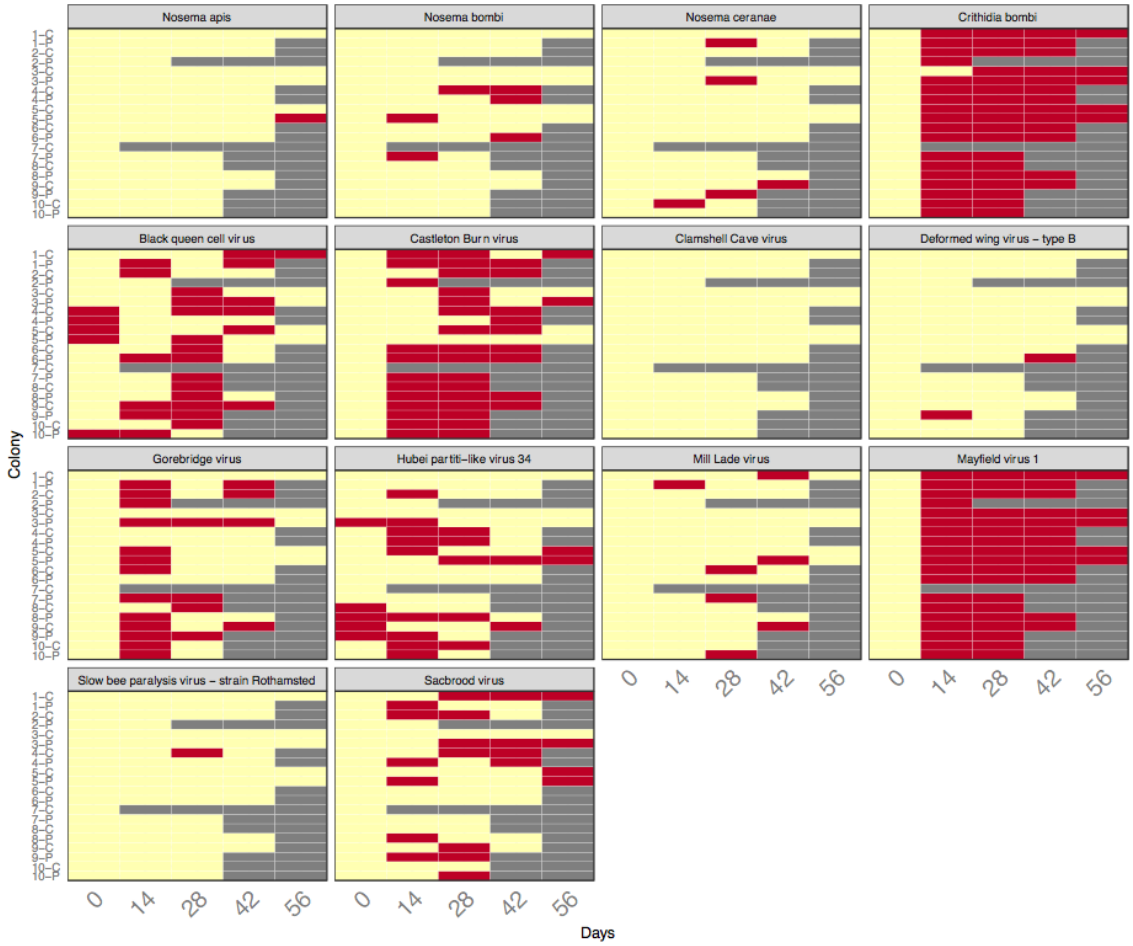


Figure 4.3 The presence or absence of each pathogen in each colony at each time point. Yellow represents absence, red represents presence and grey indicates that the colony is dead or there were too few bees remaining for a full sample. Clamshell Cave virus was detected in the experiment, but was only detected in an incomplete final sample in one colony, so the grid is marked grey rather than red.

4.4.3 Presence of Pathogens

Of the panel of pathogens tested for, six were absent in all tested colonies (Allermuir Hill virus 1, Boghill Burn virus, Mayfield virus 2, Acute bee paralysis virus, Deformed wing virus – type A and Slow bee paralysis virus – strain Harpenden). All other pathogens were detected in at least one colony, see Table 4.1. Table 4.1 shows the number colonies each pathogen was detected in, as well as the number of colonies split by treatment group.

Table 4.1 The number of colonies that each pathogen was detected in

Pathogen	Pesticide	Control	Total
Acute bee paralysis virus	0/10	0/9	0/19
Allermuir Hill virus 1	0/10	0/9	0/19
Black queen cell virus	9/10	9/9	18/19
Boghill Burn virus	0/10	0/9	0/19
Castleton Burn virus	9/10	9/9	18/19
Clamshell Cave virus	1/10	0/9	1/19
Crithidia bombi	10/10	9/9	19/19
Deformed wing virus - type A	0/10	0/9	0/19
Deformed wing virus - type B	2/10	0/9	2/19
Gorebridge virus	8/10	6/9	14/19
Hubei partiti-like virus 34	6/10	6/9	12/19
Mayfield virus 1	10/10	9/9	19/19
Mayfield virus 2	0/10	0/9	0/19
Mill Lade virus	4/10	3/9	7/19
Nosema apis	1/10	0/9	1/19
Nosema bombi	4/10	1/9	5/19
Nosema ceranae	3/10	2/9	5/19
Sacbrood virus	7/10	5/9	12/19
Slow bee paralysis virus - strain Harpenden	0/10	0/9	0/19
Slow bee paralysis virus - strain Rothamsted	0/10	1/9	1/19

Two viruses were detected in the day zero samples before placement into the field. Hubei partiti-like virus 34 (HPLV34) and Black queen cell virus (BQCV) were each found to infect five colonies (three pesticide, two control) colonies. No colony was observed to be infected with both HPLV34 and BQCV in the day 0 samples collected.

The pattern of infection over time varied considerably between the different pathogens, see Figure 4.2. Some pathogens such as Mayfield virus 1 and *Crithidia bombi* were immediately present in nearly every colony by the first sampling period 14 days in (19 and 18 colonies infected respectively). From this point on, they were detected in every colony going forward, implying that they reached high prevalences within the colonies. Gorebridge virus (GV) was found in a large number of the colonies at 14 days (14/19), after which it was only sporadically detected. Infections with Castleton Burn virus (CBV) proceeded at a similar rate initially, with 12/19 colonies being infected at day 14, but did not

exhibit the same drop off in infection observed in Gorebridge virus (GBV). BQCV showed a similar pattern to Gorebridge virus, but with its peak of infection at 28 days rather than 14 days (12/18), before showing a similar reduction in infection rate. Others still maintained a low (*Nosema bombi*, *Nosema apis, Nosema ceranae,* Clamshell Cave virus, DWV-B, SBPV-R and Mill Lade virus) or intermediate (HPLV34 and Sacbrood virus) frequency throughout the experiment.

4.4.4 The Effect of Pesticide on Pathogen Uptake Rate

There is some evidence that pesticides increase the uptake rate of pathogens from the environment, regardless of the parameterisation. Table 4.2 shows the modal estimates and 90% shortest posterior intervals of the size of the pesticide time interaction for the continuous time parameterisations and the pesticide effect for the discrete timepoint parameterisations. The posterior probability that the effect of pesticide on viral uptake is positive is also shown, given the directionality of our hypothesis.

Table 4.2 The estimated effect of pesticide on infection risk. Note that the effect sizes are measured on different scales for the different types of effect estimated, so are not directly comparable between models

Model	Estimate type	Posterior effect size	P(x)>0
Continuous time Continuous pesticide dose	Continuous-Continuous Interaction	0.044 (0.000, 0.196)	0.938
Timepoints as factors Continuous pesticide dose	Continuous	2.853 (0.489, 6.275)	0.971
Continuous time Pesticide treatment as factor	Continuous-Factor Interaction	0.019 (-0.004, 0.047)	0.907
Timepoints as factors Pesticide treatment as factor	Factor	0.811 (0.222, 1.885)	0.985

When time was treated as a continuous variable (Models 1 and 3), the effect was positive, meaning the rate of infection increased over time. We found a modal daily increase of 0.111 of the latent variable (90% SPI: 0.000, 0.213) in the continuous pesticide dose model and 0.124 (90% SPI: 0.011, 0.225) in the factorial pesticide model. The timepoint models (Models 2 and 4) indicated that the level of infection at day 28 exceeded that at day 14, and then fell back to be indistinguishable from the day 14 value at day 42 and 56, as the uncertainty

around the estimates increased. There was very weak evidence for a small positive effect on uptake rate observed in the second block, with the modal increases in all cases being of small to intermediate size (M1: 0.044, M2: 0.711, M3: 0.043, M4: 0.444) and the lower bound of the 90% SPI falling below zero. There was very little evidence for considerable differences in up the rate of uptake of pathogens between locations or colonies, with very small estimated variances for both (location variance posterior modes: M1: 0.000, M2: 0.002, M3: 0.000, M4: 0.006; colony variance posterior modes: M1: 0.001, M2: 0.220, M3: 0.000, M4: 0.005).

There was little evidence that the different pathogens in the study responded differently to the pesticide treatment with small modal variances being estimated for the pathogen-pesticide random interaction and no variation not being excluded (M2: 0.080, 90% SPI: 0.000, 9.083; M4: 0.009, 90% SPI: 0.000, 0.627), though this variance was imprecisely estimated in the factorial timepoint/ continuous pesticide model (Model 2). Pathogens did, however, have very different background probabilities of presence and absence, with large intercept random effect variances (M1: 2.830, 90% SPI: 0.805, 8.256; M2: 8.205, 90% SPI: 4.069, 23.455; M3: 2.909, 90% SPI: 0.622, 7.723; M2: 9.569, 90% SPI: 4.730, 23.277). The considerably higher intercept variances in Models 2 and 4 are likely due to the lack of a random interaction of the pathogen with time to control for the differential rate of uptake, which was estimated to be of intermediate size in models where it was included (M1: 0.028, 90% SPI: 0.014, 0.102; M2: 0.014, 90% SPI: 0.029, 0.109), and the exclusion of the day 0 data. No evidence was found for a correlation between intercepts and slopes in any model.

For Models 2 and 4 that treated the timepoint as a factor, there was a general trend for the pesticide timepoint interactions to be negative after day 14, consistent with the hypothesis that a pulsed pesticide exposure should lead to a reduction in the effect over time, but the imprecision in the estimates caused by colony death meant that this could not be confirmed.

4.5 Discussion

In this study, we have provided evidence for an increased rate of pathogen uptake after exposure to field-realistic levels of pesticide, in concordance with our prior hypothesis.

4.5.1 Pathogen Level Differences in Uptake Rate

Incidence strongly differed between pathogens. This may be due to differences in the rate of exposure between each pathogen and the bees, or differential probabilities of infection on contact when exposure did occur. Initial exposure for a colony is thought to predominantly occur at flowers that have been contaminated by secretions from infected insects in the community, with different flowers providing different efficiencies of transmission from bee to flower (Durrer and Schmid-Hempel 1994; Graystock et al. 2015). Once an individual within a colony becomes infected, both additional infection from external sources and intracolony transmission can drive infection within that colony. As bumblebees defecate within the colony (Goulson 2010) contaminating the colony, as well as becoming physically contaminated with pathogens themselves, intracolony infection does occur, with an individual bee's risk being related to the direct contact rate that that bee has with infected individuals within its colony (Otterstatter and Thomson 2007). For rare pathogens with an appreciable chance of infection on contact, this implies that the time to initial contact is likely to be the limiting factor for colony level prevalence, as once a bee within the colony becomes infected, within-colony contact changes the pathogen from being globally-rare to locally-common. This leads to the expectation that, in many cases, a lag would occur, followed by constant infection of a pathogen in the colony. However, Figure 4.2 shows that this pattern was rarely observed; in almost all cases, pathogens were detected sporadically throughout the experiment, or detected in nearly all colonies at day 14 then were either lost or persisted at a detectable level until the colony failed. This is complicated by the fact that, with a sample of 5 bees being taken at each timepoint, we are more likely to fail to detect a low prevalence infection than detect it, with the switch to an infection being more likely to be detected than missed at approximately 13% prevalence. Additionally, new workers are constantly being produced by the colony, which drives down prevalence,

assuming that infection in the larval stage is rare. As such, sporadic detections are expected for infections maintained at a low level within colonies, simply due the sampling effects.

There was no clear pattern in the pathogens that were rapidly taken up versus those that were gained more slowly. The most quickly gained pathogens were Mayfield virus 1, Castleton Burn virus, Gorebridge virus and *Crithidia bombi*; two picornavirus-like viruses, a tombus-like virus and a trypanosmatid. The three viruses have been previously reported at intermediate frequencies of 15 to 20% in the general *B. terrestris* population in Scotland (see Chapter 2), but the consistency of detection in the colonies observed here implies a considerably higher prevalence within these specific colonies. The rate of uptake implies a considerable background prevalence, given that in order to become infective within the colony, replication first has to occur in the first infected individual. Therefore, the first infection must have occurred considerably early than day 14 in most colonies, unless multiple foragers picked up the pathogen simultaneously due to high levels of environmental contamination. *C. bombi* has been long recognised as a very effective parasite of multiple bumblebee species, which consistently reaches high prevalences in the summer months (Plischuk et al. 2017; Popp, Erler, and Lattorff 2012; Ruiz-González et al. 2012; Shykoff and Schmid-Hempel 1991), making its rapid infection of the colonies unsurprising. Gorebridge virus shows an interesting pattern where at day 14, most colonies test positive for the virus then after that point it only becomes sporadically detected. This could be due to the virus being highly pathogenic, with the workers that get infected quickly dying, and so not passing the virus on or due to quick clearance of the virus and a lack of reinfection. The most slowly acquired pathogens were *Nosema apis*, *Nosema ceranae*, *Nosema bombi*, Slow bee paralysis virus - Rothamsted strain (SBPV-R) and Deformed wing virus - type B (DWV-B). SBPV-R and DWV-B have previously been detected at low levels in *B. terrestris* (Fürst et al. 2014; McMahon et al. 2015) though both studies combined estimates for DWV-A and DWV-B for their prevalence reporting. SBPV-R has been found at considerably higher levels in both *B. terrestris* and other bumblebee species (Fürst et al. 2014, Chapter 3). The fact that *B. terrestris* can be heavily infected with both SBPV-R and DWV-B implies that the lack of infection in this study is due to the local rarity of the viruses.

Two of the more commonly detected viruses in the study, HPLV34 and BQCV, were detected in our colonies before they were placed into the field. We cannot confirm whether the colonies were infected upon arrival, or whether they acquired the infection from pollen in the laboratory feeding procedure. Extensive infection with HPLV34 in breeding facilities would be unsurprising, as it is both a recently discovered virus (Cornman et al. 2012; Shi et al. 2016) and had never previously been described in bumblebees and is therefore unlikely to have been detected or tested for. Common BQCV infection would be more surprising, given that it consistently reported infecting bumblebees (McMahon et al. 2015; Peng et al. 2011; Zhang et al. 2012) and has documented pathological effects in honeybees (Doublet et al. 2015). Irrespective of the origin of the infection before the colonies were placed into the field, it does appear that additional infections of these viruses were gained in colonies that were truly uninfected during the control period, where our sample size was larger giving us more ability to detect rare pathogens.

4.5.2 Pesticide Effects

While we found no conclusive evidence of bees reducing their feeding rate when fed on clothianidin contaminated sugar water relative to bees fed on uncontaminated sugar water, this effect has been previously been reported in the literature, but only at higher concentrations of clothianidin (Kessler et al. 2015). The potential effect sizes of the pesticide-associated increase in the pathogen acquisition rate in the study range from functionally non-existent to highly biologically significant. The modal effect size represents a small positive shift in the rate of pathogen uptake relative to the non-pesticide exposed control group of approximately a 6^{th} of the total increase due to time. Qualitatively similar estimates of the pesticide effect were given irrespective of whether pesticide was treated as a factorial treatment effect or the dose itself was directly used. The small modal size of the effect is consistent with our hypothesis that if there was a visible effect at field realistic pesticide levels, it would be small. Low doses of pesticide, even when there are no other options for feeding would likely be capable of being quickly detoxified (Cresswell et al. 2014), assuming the detoxification rate for clothianidin is similar to that of imidocloprid. The effect size that we estimated may differ slightly from the real

effect size that an intense clothainidin oilseed rape bloom might bring about. We assumed that bees had no other choice than to feed on the contaminated sugar water, but in a field setting, there are likely to be other flowers in the environment during a bloom of a pesticide treated crop, which could lead to a reduction in the amount of pesticide consumed. On the other hand, we did not add any clothianidin to the pollen that the bees consumed, which would bias our estimate downward, given that other neonicotinoids are often found at similar concentrations in pollen as in nectar (Carreck and Ratnieksi 2014).

A secondary prediction is that the effect would be strongest at the first timepoint after the exposure, i.e.14 days. We expected this for four reasons. Firstly, the amount of stored sugar water still contaminated with the experimentally administered pesticide will decrease over time. Secondly, the bees will have longer to recover from the initial effects of direct exposure. Thirdly, there will be more newly emerged bees in the colony that never experienced the original pesticide exposure. Finally, non-pesticide exposed bees will have longer to acquire infections, meaning that if the effect is a change in rate leading to the same final prevalence, the effect will be less visible later in the experiment. When treating timepoints as statistically independent, we find weak evidence of this, with day 14 having a noticeably larger estimated interaction with the pesticide treatment but due to the imprecision with which the later timepoint pesticide interactions were estimated, this effect could not be confirmed. This is consistent with the hypothesis that there should be a pesticide effect of decaying size over time, but is it purely indicative given the low ability to detect effects within the experimental period, due to the lack of replication.

4.5.3 Conclusion

In this study, we found evidence supporting a small to intermediate increase in the rate of pathogen acquisition in bumblebee colonies stressed with a pulsed exposure to a field realistic concentration of the neonicotinoid pesticide clothianidin. Given previous data on pathogen-pesticide interactions, all effects were of the expected sign, but the data does not provide evidence that the size of the increase in risk of infection is large enough to be biologically meaningful at field realistic levels of clothianidin, as the effect size estimate varies from large to insignificantly small. The measured increase in the rate of infection

under pesticide exposure means that it is potentially important, however, a considerably larger study longitudinal study at multiple locations would be required to determine the general relevance neonicotinoid pesticides on the uptake of pathogens into wild populations.

Chapter 5 - Between-host fitness landscape correlation drives multiple outcomes in the evolution of multihost digital parasites

David J. Pascall and Lena Wilfert

5.1 Abstract

The evolution of multihost parasites is complicated by the fact that they partition their time across multiple host species. These host species will represent different environments from the perspective of the parasite, leading to incongruity between the fitness landscapes provided by each host. We used a simple simulation model to evolve digital parasites in two host species under a variety of different degrees of fitness landscape correlation. We also varied epidemiologically important parameters such as the contact rate between hosts and the presence or absence of an adaptive immune system. We found that the rate of adaptation was considerably higher in two-host systems relative to one-host systems, and that increasing correlation between the fitness landscapes led both to increased generalism in the evolved digital parasites and reduced diversity. These results are consistent with previous literature on the evolution of multihost parasites and provide good evidence of the importance of between host fitness landscape correlations in the driving of the dynamics of the evolution of multihost parasites.

5.2 Introduction

Multihost parasites face a complex evolutionary environment as they spend time, and therefore evolve, in multiple host species, either at different stages of their lifecycle or after cross-species transmission. Any two hosts of the same or different species, present distinct challenges for a parasite. These challenges are brought about by the parasite facing a different environment in each host due to, for instance, host genetics (Carpenter et al. 2012; Longdon et al. 2011) or microflora (Brotman et al. 2014; Sekirov et al. 2008). These different environments will likely also cause selective differences in the parasites of individual hosts, an extreme example being the extensive local adaptation that occurs within an infected individual in long term HIV infection (Bordería, Codoñer, and Sanjuán 2007). From the perspective of the parasite, the environment provided by two hosts of the same species will be more similar than that provided by two hosts of different species. As such, it is unsurprising that divergent selection between host species has been observed (Bedhomme, Lafforgue, and Elena 2012; Turner and Elena 2000; Vale et al. 2012). This

divergent selection is likely, in part, driven by incongruences between the fitness landscapes experienced by parasites between host species. The degree of this divergence or convergence in fitness landscapes is expected to strongly influence the outcome of the evolution of a parasite over time, with parasites preferentially host-switching to species that have correlated fitness landscapes (Longdon et al. 2014). Highly divergent landscapes should lead to specialisation within a parasite population, while convergent landscapes would lead to generalism, as there is no associated cost to being capable of infecting multiple host types (Gandon 2004).

This argument implicitly assumes that the host itself evolves at a much slower rate than the parasite, so coevolutionary dynamics are unimportant relative to the ecological dynamics driven by cross-species transmission. If the hosts were coevolving with the parasite, the adaptive landscape experienced by the parasite on each host would themselves change over time (Burmeister, Lenski, and Meyer 2016), which may favour or disfavour generalism, depending on the precise nature of the change. The assumption that the host evolves at a much slower rate than the pathogen is likely to be true for viruses, as they tend to have extremely short generation times, on the order of minutes (Yarwood 1956), with RNA viruses also having similarly high mutation rates (Sanjuan et al. 2010). This combination should meet the requirement that host evolution occur at a different rate to pathogen evolution, in most metazoan hosts.

We expected the amount of genetic variation present in the population to be determined by the degree of correlation between the host fitness landscapes, along with the degree of specialism. Fluctuation in environmental conditions has long been theorised to be a mechanism for maintaining diversity in natural populations (Haldane and Jayakar 1963). This theory has developed a large empirical backing (reviewed in Kassen (2002)). The effect of fluctuation comes about through different variants being selectively favoured in the different environmental conditions, and the extensive examples of antagonistic pleiotropy between hosts are reviewed in Bedhomme, Hillung and Elena (2015). In addition to host-switching, the host immune system can act as a fluctuating environmental condition for a parasite. Immune systems that exhibit immunological memory provide long-lasting specific rather than general

protection, and have been recorded in highly divergent groups of species, from both jawless and jawed vertebrates (Flajnik and Kasahara 2009), to insects (Pham et al. 2007; Sadd and Schmid-Hempel 2006), to bacteria (Barrangou et al. 2007). In this study, we envision a situation in which fine strain differences can be detected by the immune system, and immunity can arise to each one separately. With such an adaptive immune system in both hosts, a pathogen transmitting between host, species faces a fitness landscape fluctuating on two scales. There are large systematic fluctuations between species, driven by the between-host differences in the environment experienced by the parasite, and local fluctuations within-species dependent on the frequency of the variants of the parasite, the strength of the immune system, and the length of the immunological memory.

Parasites evolve over discreet fitness landscapes, due to the underlying discreteness of the genetic code. This discreetness imposes a limit on the areas of the fitness landscape that are accessible from a particular position. A large theoretical literature has built up on evolution across discreet fitness landscapes in the single static case (reviewed in de Visser and Krug (2014)). One of the most important parameters is the ruggedness of the underlying landscape. A rugged landscape has many local optima, where the fitness is less than the maximum on the landscape, but higher than the fitnesses of all variants one mutational step away. The degree of ruggedness of real fitness landscapes is an open empirical question, but the assumption of a smooth landscape leading to a single peak is highly unrealistic and evidence is mounting that high levels of epistatis may be the norm (Cervera, Lalić and Elena 2016; Stern et al. 2017). House-of-Cards landscapes, where the fitnesses within the landscape are uncorrelated, are also generally unrealistic given the high levels of redundancy, both at the protein and genetic code level, but less so than smooth fitness landscapes (Cervera, Lalić and Elena 2016). Theory suggests that adaptive walks on highly rugged landscapes tend to be short (Kauffman and Levin 1987), as a single lineage will quickly evolve by hill climbing to a local optimum and become stuck. However, empirical work with replicate populations of Tobacco etch virus suggests that actually getting stuck on a local optima may be a rare occurrence in reality, even on highly rugged landscapes (Cervera, Lalić and Elena 2016). Fluctuation between fitness landscapes may allow escape from

local optima (Cervera, Lalić and Elena 2016; Cheetham 1993), as incomplete correlations between the fitness landscapes will mean that a local optimum on one landscape is unlikely to be a local optimum on the other. This may potentially allow pathogens evolving across multiple hosts to adapt faster than those in single hosts.

Based on this prior work, we had several assumptions about the effects of evolution across multiple hosts. When fitness landscapes between hosts are positively correlated, generalism should be favoured, as variants with high fitness on one host on average have high fitness on the other host. Directly following from this, increasing correlations between hosts should decrease the variation in the pathogen population as specialist variants are outcompeted by generalist variants. Evolution should occur more quickly when a pathogen is transmitting between two hosts than when it is transmitting in one, as the second host allows escape from local optima on the first. The presence of an adaptive immune system should increase the diversity, and by increasing the diversity, interact synergistically with evolution across two hosts to increase the rate of adaptation. To test these hypothesises, we simulated a highly-abstracted model of pathogen evolution via hillclimbing across an uncorrelated discrete fitness landscapes within an epidemiological framework, with and without the presence of an adaptive immune system.

5.3 Methods

All modelling was done in R version 3.3.2 (R Core Development Team 2016). Functions from the packages MASS (Venables and Ripley 2002), stringi (Gagolewski 2017), gtools (Warnes, Bolker, and Lumley 2015), plyr (Wickham 2011), fdrtool (Klaus and Strimmer 2015) and BiasedUrn (Fog 2015) were used in the modelling procedure.

The general modelling structure is as follows, specific departures from this general model will be noted in their specific sections. One thousand hosts are generated and placed into grid of a defined size. The only host traits are the species, infection status and immunity profile. All possible sequences of ten nucleotides in length are generated, and assigned a within-host reproduction

rate in the host species from a probability distribution. These represent the parasites. Ten hosts are infected with ten random sequences. Each model iteration, hosts are randomly placed onto the grid. Hosts that are placed in the same location are able to interact and can infect one another. The probability of infection is predefined. For simplicity, we do not assume a trade-off between within host reproduction rate and transmission probability, and the parasite is assumed to be avirulent. Each model iteration, every host has the chance to clear any infection with a defined probability. When a host clears an infection, it gains that strain in its immunity profile (if an adaptive immune system is being modelled).

Each iteration, after all hosts have been placed, and clearance and infection have occurred, every pathogen sequence is given the chance to evolve. A random sequence one mutational step away from the current sequence is generated and the fitnesses are compared between the resident and mutant strains. Following nearly neutral theory (Ohta 1992), the probability that the mutant strain replaces the resident strain is $1\text{-}e^{-s}/1\text{-}e^{-sN}$, where in this case N is the presumed within-host viral effective population size, set to $5x10^4$ as a realistic figure, following work from the HIV literature (Maldarelli et al. 2013; Seo et al. 2002). We assume that the replacement is immediate to avoid tracking the relative frequencies of variants within a host. The model was run for 200 iterations, and then summary statists were generated.

Within-host reproduction rates were generated from uniform (0,1), standard halfnormal distribution, and gamma (1,1) distributions. These all provide within landscape correlations of 0. Multiple distributions were used to test for strong dependences of the results on the distribution of fitnesses across the landscape. While zero fitness variants are commonly observed in nature (Cervera, Lalić and Elena 2016), we did not add more variants with fitness 0 into the model than were generated at random.

The general modelling algorithm is provided in psuedocode below. For each parameter combination, the models were run 10 times and all results were stored. To enable simple analysis, two experiments were run. Experiment 1 explored the effect of varying the infection and clearance probabilities, the

number of hosts and the presence and absence of an adaptive immune system. Experiment 2 explored the effects of varying the contact rate between the two host species, and varying the correlation of the fitness landscapes between the two host species with the presence or absence of an adaptive immune system. Models that could be generated both from the combinations of variables in Table 5.1 and in Table 5.2 were run 20 times (10 in each Experiment). For analysis, all models with the same parameter values were included, and as a result, some models had sample sizes of 10 while a minority had sample sizes of 20.

its = number of model iterations = 200

s = (new genotype fitness/resident genotype fitness)-1

$p_{replace} = 1-e^{-s}/1-e^{-sN}$

generate 1000 *hosts*;

generate 2000 *spaces* for hosts to interact in;

generate every possible nucleotide string of 10 base pairs;

assign each nucleotide string a fitness drawn from a probability distribution;

infect 10 random hosts with 10 random sequences;

for (i in 1:*its*) {

    for (q in 1:*hosts*) {

        place a host on space

        determine if any other hosts share that space based on a draw from a binomial distribution with the number of draw representing the number of unplaced hosts and the probability of sharing being the number of spaces with no assigned hosts this iteration;

        allow all selected hosts to clear any current infections with probability p_{clear};

        allow all selected hosts to infect one another with probability p_{inf} if the genotype of the infecting strain has a higher within host fitness and probability 0 if it has a lower or equal within host fitness;

        correct host iterator for the number of placed hosts so hosts are not placed twice;

    }

    for (e in 1:infected hosts) {

        randomly generate a new nucleotide sequence of Hamming distance 1 from the currently infecting genotype;

        allow the new genotype to replace the current genotype with probability $p_{replace}$;

    }

    save the number of infected hosts, number of variants, and fitness distribution of variants;

}

5.3.1 Experiment 1 – Models 1-4

5.3.1.1 Model 1 – One Host Adaptation

This model followed the schema defined above. The models were run for all combinations of parameter values in Table 5.1, excepting those where p_{clear} exceeded p_{inf} to filter out the uninteresting cases where the infection is simply lost.

Table 5.1 Parameter combinations used in Experiment 1

Fitness Landscape Distributions	p_{inf}	p_{clear}
Uniform (0,1)	0.05	0.01
Standard half-normal	0.1	0.05
Gamma (1,1)	0.3	0.1
	0.5	0.3

5.3.1.2 Model 2 - One Host Adaptation with Immunity

This model follows the structure of Model 1, but with the addition of an adaptive immune system. A simple (perfect) adaptive immune system is modelled by the addition of a rule that once a host clears an infection, the probability of infection for that genotype in that host becomes 0. Without host death, this will eventually lead to complete immunity in hosts, so this assumption is only valid for a short timescale. Models were run for the same set of parameter values as Model 1.

5.3.1.3 Model 3 - Two Host Adaptation

The model follows the same structure as Model 1, but with two host species. Instead of 1000 hosts of one species, instead there are two species each of 500 hosts. There is no preferential association with conspecifics. Fitness landscapes for each host are uncorrelated both within and between hosts, using the same probability distributions as the single host case. The model was run with the same combination of parameter values as Model 1.

5.3.1.4 Model 4 - Two Host Adaptation with Immunity

This model is a combination of Model 2 and Model 3. Two hosts were set up in the same way as Model 3, with an adaptive immune system as modelled as in Model 2. The model was run with the same combination of parameter values as Model 1.

5.3.2 Experiment 2 - Model 5

5.3.2.1 Model 5 - Two Host Adaptation with Correlated Between Host Landscapes and Differential Contact Rates

This model has the same basic structure as Model 4. However, correlated fitness landscapes were generated between host species using a multivariate Gaussian copula with the desired correlation, then transforming the corresponding correlated uniform variables to the relevant distribution (Nelsen 2013). These distributions were correlated uniform, gamma and standard halfnormal distributions with the same marginal distributions as those given above. Differential contact rates between the host species were modelled by weighting of probabilities of conspecific versus heterospecific contact. This was done by use of random draws from the univariate Wallenius' noncentral hypergeometric distribution, which describes the probability distribution of *z* draws from a set of two finite pools, with biased selection probabilities, as described by the odds of selecting the first type relative to the second (Chesson 1976). Both increased and decreased within-species contact rates relative to random mixing were included to account for both gregarious species on the one hand and predator-prey or symbiotic interactions on the other. For these models p_{inf} was fixed at 0.3 and p_{clear} was fixed at 0.05. Models were run for all combinations of parameter values in Table 5.2.

Table 5.2 Parameter combinations used in Experiment 2

Fitness Landscape Distributions	Correlation	Odds of conspecific versus heterospecific contact
Uniform (0,1)	-0.9	0.1/0.9
Standard Half-normal	-0.5	0.25/0.75
Gamma (1,1)	0	0.5/0.5
	0.5	0.75/0.25
	0.9	0.9/0.1

5.3.3 Summary Statistics

For all models, the following summary statistics were generated; the number of model iterations until the viral fitness in host 1 reached the 99.9th quantile of the underlying distribution, the number of unique variants circulating in host 1 at

iteration 200, the variance in fitness of viral strains across all hosts of host 1 at iteration 200 and the number of hosts of host type 1 infected at iteration 200. Additionally, in the two-host models, the following statistics were generated; the total number of unique variants at iteration 200, the number of infected hosts of host type 2, and the proportion of unique variants that exceeded the marginal median fitnesses of their distributions in both host types.

5.3.4 Statistical Analyses

Any models with no infected hosts at iteration 200 were discarded. In all cases, 90% highest posterior density intervals were calculated by the SPIn method (Liu et al. 2015). Highest posterior density intervals were used as central intervals will generally not include the mode of the posterior distribution if the most likely parameter value is 0. The 90% interval rather than the typical 95% interval was presented, both because 90% intervals are more stable than 95% intervals, as they rely on the lower and upper 5% of samples to estimate their position rather than the lower and upper 2.5% of samples (Stan Development Team 2016a).

5.3.4.1 Rate of Adaptation

5.3.4.1.1 Experiment 1

The number of model iterations until a fitness equal to the 99.9th percentile of the underlying fitness distribution was achieved in host 1 was analysed using a proportional hazards survival analysis censored at 200 iterations using the ‘cph’ function from the package ‘rms’. A Bayesian analysis using the function ‘survregbayes2’ from the package ‘spBayesSurv’, with a prior consisting of a mixture of 10 beta (1,1) distributions was attempted, but due to convergence issues the frequentist analysis was run instead. The model fitted was:

Time till threshold fitness ~ $p_{infection}$ + $p_{clearance}$ + Number of hosts + Fitness landscape distribution + Reinfection on clearance + $p_{infection}$:$p_{clearance}$ + Reinfection on clearance:$p_{clearance}$ + Reinfection on clearance:Number of hosts

5.3.4.1.2 Experiment 2

A survival analysis model censored at 200 iterations was also run for Experiment 2, again only using the model iterations until maximum observed fitness was achieved in host 1, as host 2 should behave the same way, by the

symmetry of the model. The same function was used as in the previous model. The model fitted was:

Time till threshold fitness ~ Correlation between host landscapes +$p_{conspecific}$ + Fitness landscape distribution + Reinfection on clearance + Correlation between host landscapes:$p_{conspecific}$

5.3.4.2 Variation

5.3.4.2.1 Experiment 1

The factors influencing the number of viral variants present in the population at iteration 200 were analysed by a Poisson GLM square root link with an observation level random effect (OLRM), as a test Poisson GLM indicated overdispersion. A negative binomial GLM was also run as an attempt to correct for the overdispersion, but posterior predictive checks indicated that the OLRM provided a better fit to the observed data and 10-fold cross validation using the 'kfold' function in the 'loo' package indicated no strong preference for the one model over the other, so the OLRM model is presented. The GLM was fitted in the package 'rstanarm', using the 'stan_glmer' function. Normal (0,10) priors were placed on the intercept and regression coefficients with a half-Cauchy (0, 2.5) prior being placed over the standard deviation of the normal distribution of the random effect. The model fitted was:

Number variants in host 1 ~ $p_{infection}$ + $p_{clearance}$ + Number of hosts + Fitness landscape distribution + Reinfection on clearance + $p_{infection}$:$p_{clearance}$ + Reinfection on clearance:$p_{clearance}$ + Reinfection on clearance:Number of hosts

5.3.4.2.2 Experiment 2

A similar model was fitted for the Experiment 2 analysis, the same priors and functions were used. The model fitted was:

Number variants in host 1 ~ Correlation between host landscapes + $p_{conspecific}$ + Fitness landscape distribution + Reinfection on clearance + Correlation between host landscapes:$p_{conspecific}$

5.3.4.3 Generalism vs. Specialism

5.3.4.3.1 Experiment 2

We assayed the tendency for viruses to evolve towards generalism or specialism in the model by looking at the fitnesses of the variants present at iteration 200 over both hosts and comparing that to the median fitness of the fitness landscape. We defined a viral pathogen as a generalist, for this analysis, if it exceeded the median fitness on the fitness landscape in both hosts. For each model run, we assayed the number of generalists and specialists, then modelled the factors that lead to an increased probability of generalism versus specialism using a binomial GLM using a logit link with an observation level random effect to account for overdispersion, as implemented in by the 'stan_glmer' function in rstanarm. A Normal (0,10) prior was placed over the intercept, with Normal (0,5) priors being placed over the regression coefficients and a half-Cauchy (0,2.5) prior being placed over the standard deviation of the normal distribution of the observation level random effect. The model fitted was:

Proportion of variants exceeding the median on both fitness landscapes ~ Correlation between host landscapes + $p_{conspecific}$ + Fitness landscape distribution + Reinfection on clearance + Correlation between host landscapes:$p_{conspecific}$

5.4 Results

5.4.1 Rate of Adaptation

5.4.1.1 Experiment 1 - Host Number and Infection Probabilities

Models that simulated evolution across two host species reached the 99.9th percentile of the fitness landscape much more quickly than those with one host species ($\chi^2_{(2)}$=311.0, p<0.0001). Higher infection probabilities deceased the time until the 99.9th percentile of the landscape was reached (including all interactions: $\chi^2_{(2)}$=192.1, p<0.0001), presumably due to higher infection probabilities leading to more hosts being infected simultaneously, allowing more variants to be generated per iteration. The rate of adaptation also changed depending the underlying distribution of the fitness landscape ($\chi^2_{(2)}$=74.1, p<0.0001), with the uniform distribution showing the slowest rate of adaptation, the half-normal an intermediate rate and the gamma distribution showing the

fastest rate, see Fig 5.2 for the pattern in the correlation dataset. There was evidence that increasing the probability of clearance leads to a reduction in the rate of adaptation (including all interactions: $\chi^2_{(3)}$=45.8, p<0.0001). However, the pattern indicates that the effect may well not be linear, with the highest rates of adaptation apparently occurring at intermediate clearance rates, see Fig 5.1. There was no evidence to suggest an effect caused by whether reinfection was possible after clearance (including all interactions: $\chi^2_{(3)}$=2.6, p=0.46), nor any evidence to suggest interactions between the probability of infection and the probability of clearance ($\chi^2_{(1)}$=2, p=0.16), the probability of clearance and whether reinfection was possible after clearance ($\chi^2_{(1)}$=1.1, p=0.30) and the number of hosts and whether reinfection was possible after clearance ($\chi^2_{(1)}$=1.4, p=0.24).

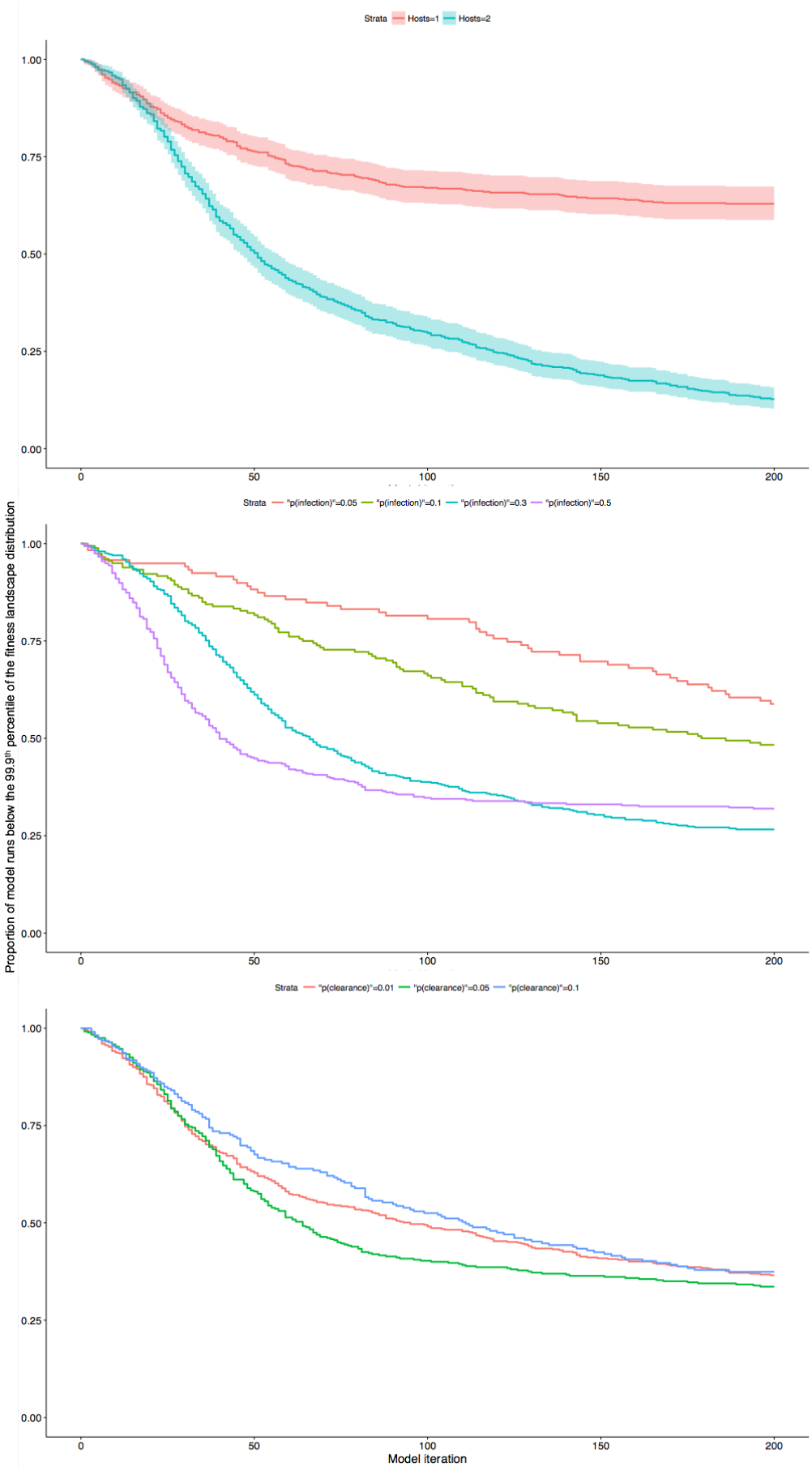


Figure 5.1 Curves showing the number of model iterations required to reach the 99.9th quantile of the fitness landscape for different numbers of host species, different infection probabilities and different clearance probabilities. 95% confidence intervals are provided for factors.

5.4.1.2 Experiment 2 - Correlation and Contact

The different fitness landscape distributions provided effects consistent with Experiment 1 ($\chi^2_{(2)}$=172.9, p<0.0001). Increasing both the correlation between the landscapes, and the levels of preferential host mixing with conspecifics lead to decreased rates of adaptation (Correlation: including all interactions: $\chi^2_{(2)}$=221.4, p<0.0001; $p_{conspecific}$: including all interactions: $\chi^2_{(2)}$= 85.28, p<0.0001), see Fig 5.2. However, when the correlation between landscapes and the degree of preferential host mixing with conspecifics were both high, there was a small ameliorative effect to the rate of adaptation ($\chi^2_{(1)}$=4.9, p=0.027), see Fig 5.3. There was no evidence for an effect of whether hosts could be reinfected after clearance ($\chi^2_{(1)}$= 1.64, p=0.199).

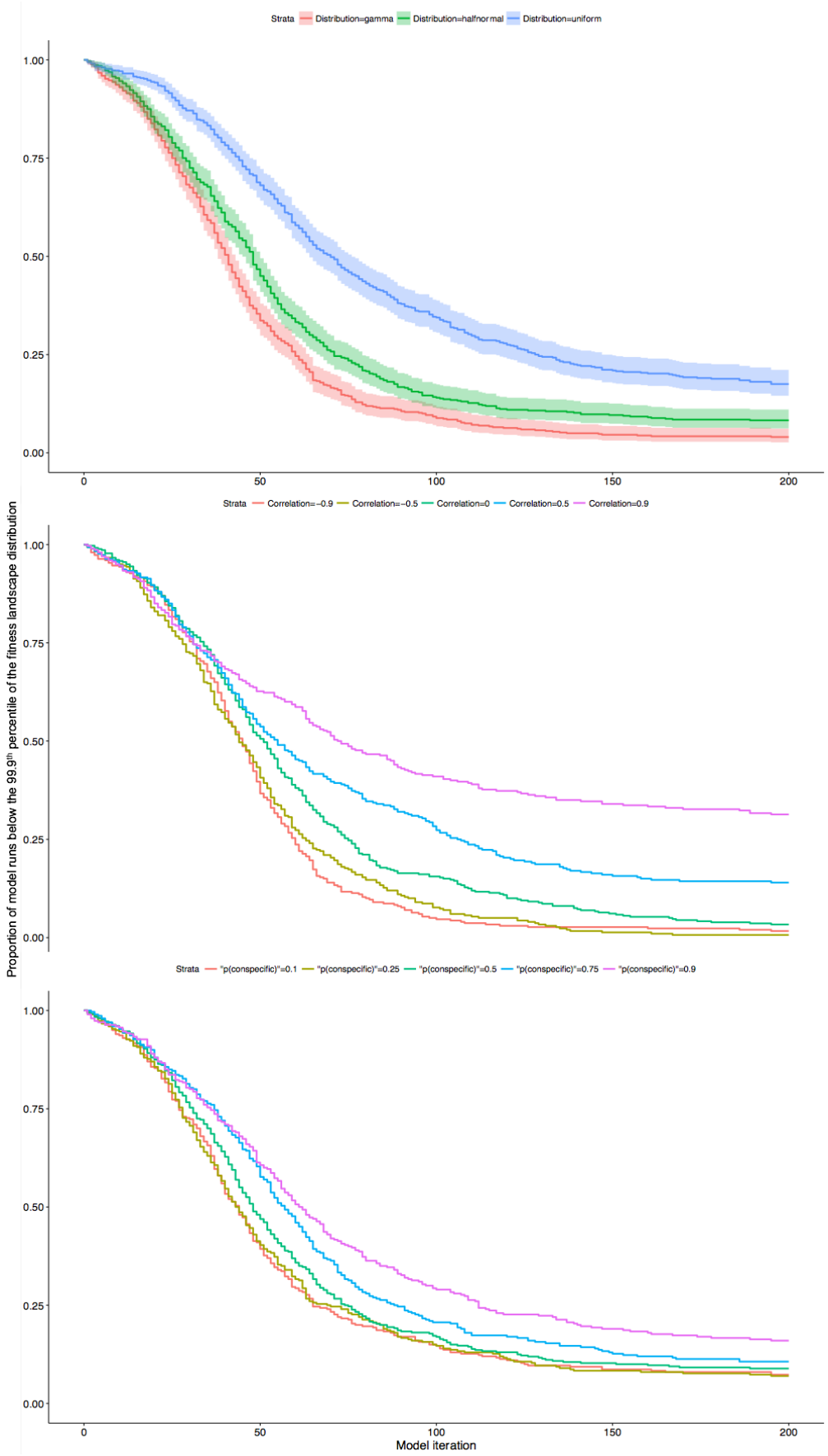


Figure 5.2 Curves showing the number of model iterations required to reach the 99.9th quantile of the fitness landscape for different fitness landscape distributions, different between host fitness landscape correlations and probabilities of contact with conspecifics relative to heterospecifics. 95% confidence intervals are provided for factors.

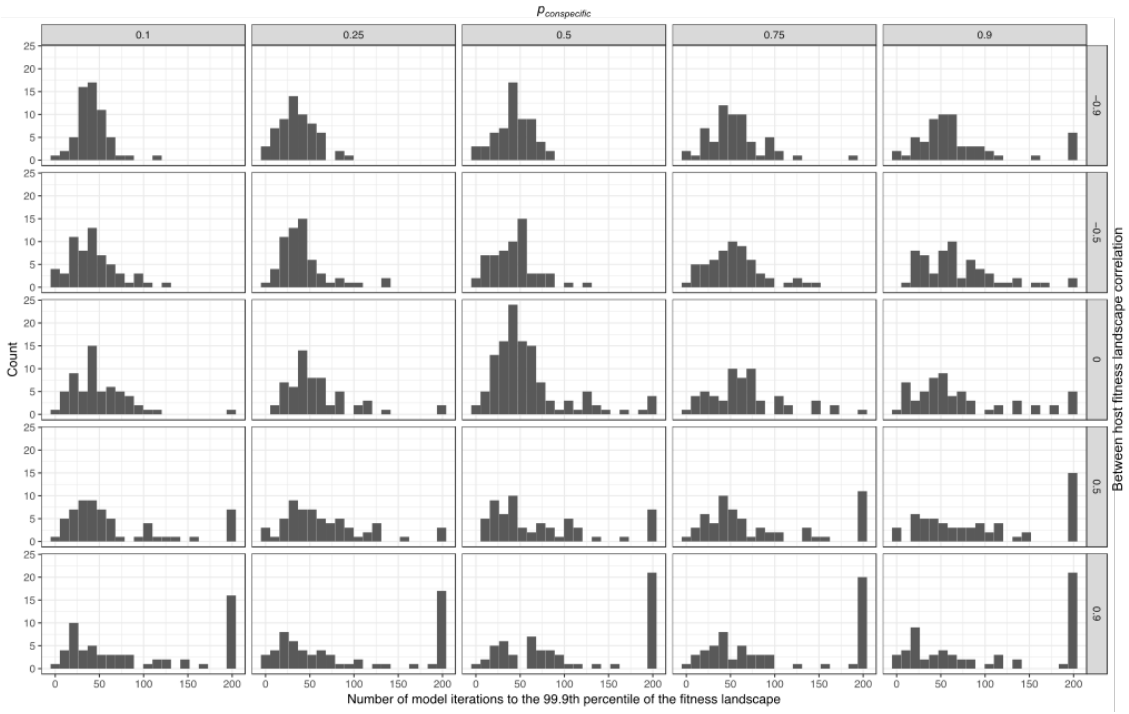


Figure 5.3 The distribution of number of model iterations required to reach the 99.9th percentile of the underlying fitness landscape by the probability of contact with conspecifics and the between host fitness landscape correlation.

5.4.2 Variation

5.4.2.1 Experiment 1 - Host Number and Infection Probabilities

Evolution across two hosts lead to a large increase in the number of circulating variants (modal change: 5.963; 90% SPI: 5.631, 6.272). The size of this change was not dramatically affected by whether reinfection was possible after clearance (modal change: -0.256; 90% SPI: -0.676, 0.218). Evolution on halfnormal fitness landscapes showed roughly the same response as evolution over gamma distributed landscapes (modal change: 0.027; 90% SPI: -0.295, 0.249), but evolution on uniform landscapes lead to reduction in the number of variants (modal change: -0.335; 90% SPI: -0.752, -0.206). There were dramatic changes in the number of circulating variants caused by different probabilities of infection and clearance. Increases in the probability of clearance lead to exceptionally large decreases in the number of variants (modal change: -39.961; 90% SPI: -54.571, -33.476). While increases in the probability of infection lead also to moderate reductions in the number of variants (modal change: -1.904; 90% SPI: -2.807, -0.606), there was a very large positive interaction between the probabilities of clearance and the probabilities of infection (modal change: 88.838; 90% SPI: 57.639, 109.862), which offsets the

effect of very high reductions in variation when the probability of clearance is high, if the probability of infection is also high. These three results must be interpreted carefully as the probability of clearance was never allowed to exceed the probability of infection, due to the inevitable loss of infection in that case. The reduction in variation caused by the probability of clearance increasing was more dramatic when reinfection with a lost variant was impossible (modal change: -12.095; 90% SPI: -19.546, -6.0394), presumably because mutation was not generating variation at a high enough rate to account for the build-up of resistance to variants present in the population. Preventing reinfection after clearance did not have a strong effect outside of its interaction with the probability of clearance (modal change: -0.160; 90% SPI: -0.594, 0.269).

5.4.2.2 Experiment 2 - Correlation and Contact

The modal estimate of the intercept of the number of variants on the latent scale was 12.638, representing a posterior predictive mode of 102 variants for a datum with average values for all continuous covariates and a gamma distributed fitness landscape. Uniform (modal change: -0.595; 90% SPI: -0.695, -0.433) and half-normal (modal change: -0.127; 90% SPI: -0.258, -0.003) distributed fitness landscapes lead to fewer variants than evolution on gamma distributed landscapes. The number of variants was strongly negatively associated with the degree of correlation between the fitness landscapes of the two hosts, with a modal change of -7.674 (90% SPI: -7.797, -7.485) on the latent scale. Increasing propensity for contact with conspecifics had a moderately smaller negative effect on the number of variants present (modal change: -5.070; 90% SPI: -5.234, -4.886). A positive interaction on the same scale as the original effects was found between the correlation and propensity for contact with conspecifics with a modal size of 5.564 (90% SPI: 5.261, 5.779), see Fig 5.5. Contrary to expectations, the number of variants decreased appreciably when hosts could not be reinfected after clearance, relative to the case in which reinfection with the same strain was possible (modal change: -0.840; 90% SPI: -0.949, -0.736), see Fig 5.4.

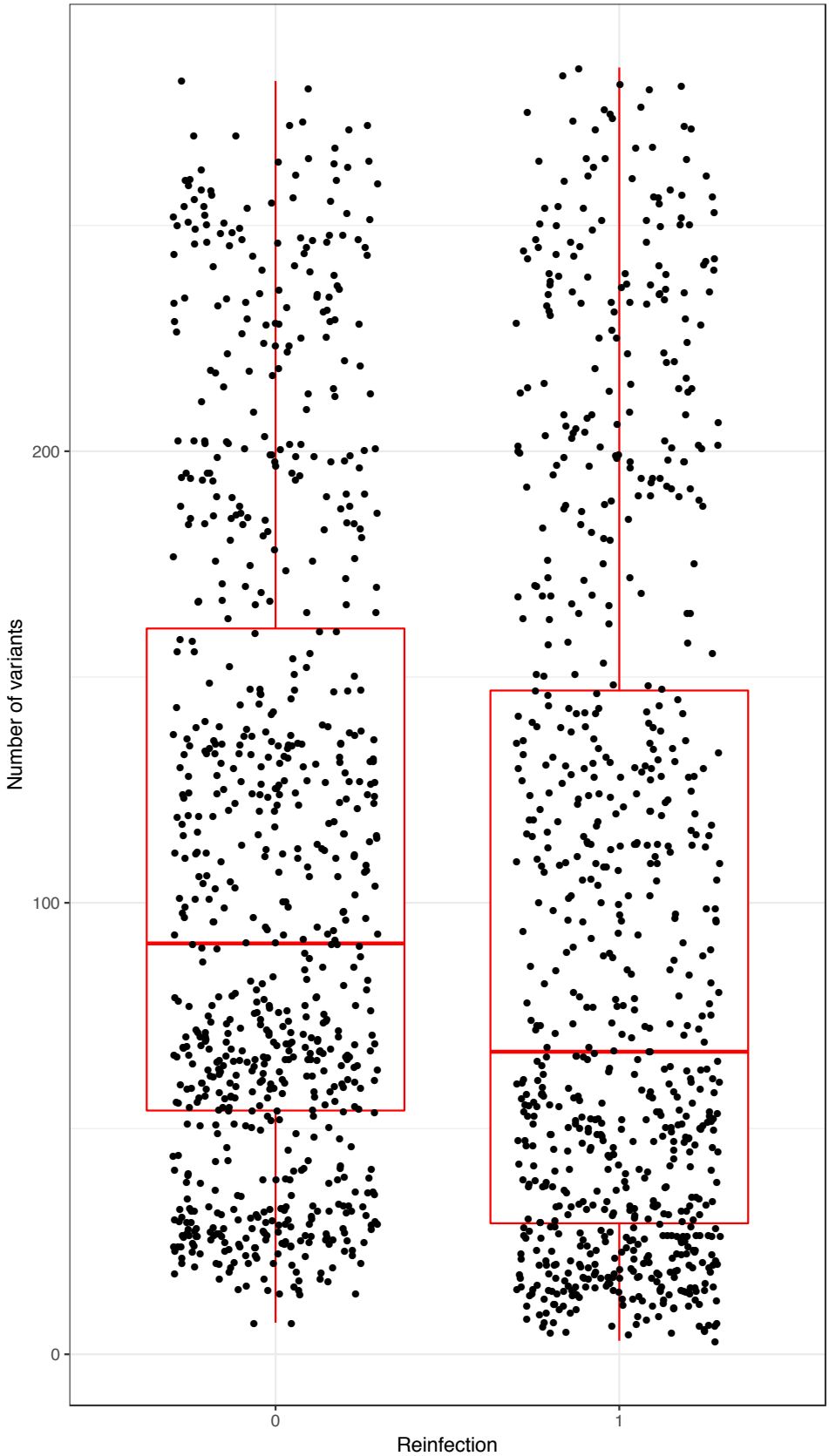


Figure 5.4 The distribution of the number of variants at model iteration 200 with and without reinfection being possible

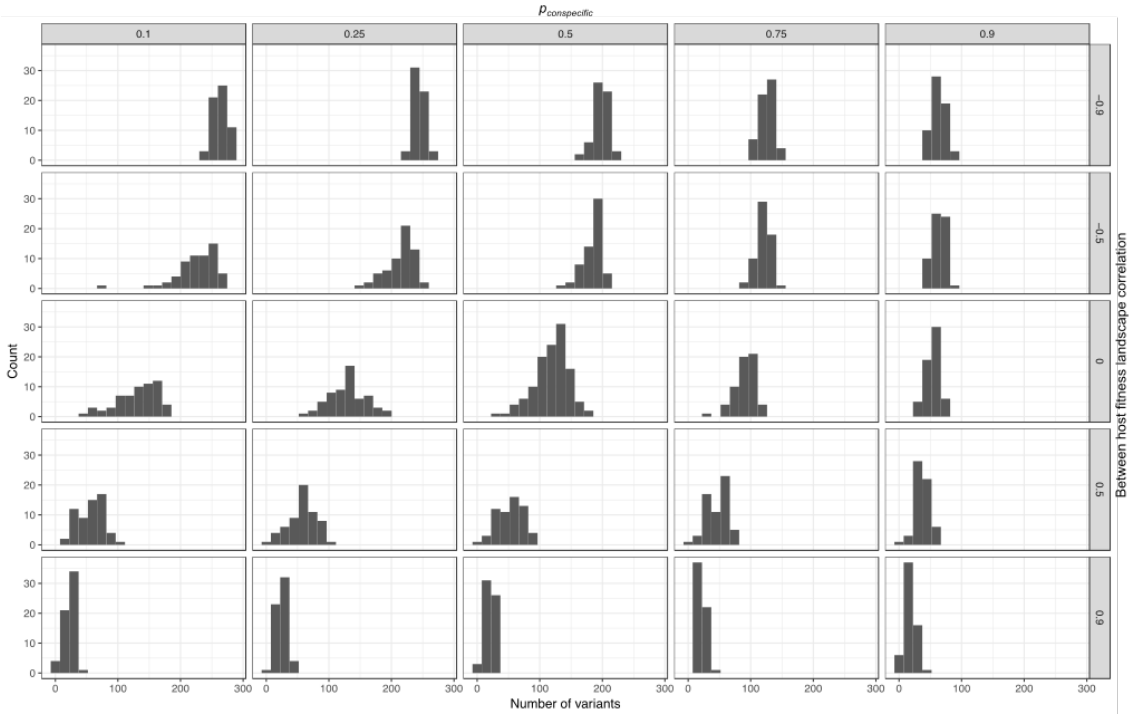


Figure 5.5 The distribution of number of variants at model iteration 200 by the probability of contact with conspecifics and the between host fitness landscape correlation.

5.4.3 Specialism vs. Generalism

By far the largest effect on the degree of generalism exhibited by the variants present at the end of the simulation was the degree of correlation between the fitness landscapes, with a per unit change in correlation providing a modal 3.891 (90% SPI: 3.758, 3.895) change in the latent variable. There is an effect of the underlying fitness landscape, with half-normal landscapes providing a small modal increase of 0.061 (90% SPI: 0.017, 0.095) in the value of the latent variable relative to gamma landscapes, and uniform landscapes providing a larger modal shift of 0.226 (90% SPI: 0.186, 0.264) relative to gamma landscapes. Preventing reinfection upon clearance of a strain provided a very small positive increase in the propensity for generalism in variants (modal change: 0.034; 90% SPI: 0.006, 0.069), potentially as generalist variants when rare were not removed from the population due to a build-up of resistance. As predicted by theory, increasing the degree of preferential contact with conspecifics decreased the probability of generalist variants existing the population (modal change: -0.429; 90% SPI: -0.493, -0.360). This change is relatively small, representing a maximum shift on the probability scale of around than 0.1, assuming that the initial probably is near 0.5. An imprecisely estimated interaction between the degree of correlation in the fitness landscape and the

probability of contact between conspecifics was found, with the reduction in generalism due to highly preferential interaction with conspecifics being reduced when the correlation between the fitness landscapes across hosts is high (modal change: 0.151; 90% SPI: 0.021, 0.270).

5.5 Discussion

Using a simple model, we have shown that evolution on a fluctuating House of Cards-like fitness landscape increases the rate of adaptation relative to a stable landscape. Fluctuating fitness landscapes are experienced by all parasites to some degree, as the precise make up of hosts differs even within species. However, hosts that are of the same species must, on average, have a more highly correlated fitness landscape than hosts that are of different species, given the existence of host barriers.

5.5.1 Specialism, Generalism and Fitness Landscapes

As expected in this simple case, the degree of correlation between the fitness landscapes in the different hosts has significant impact on the outcome of the evolutionary process, at least over the short timescales considered here. High correlation of the fitness landscapes between hosts corresponds to the case where there is no real cost to adapting to both host species, as regions of the sequence space that provide high fitness in one host, on average, provide high fitnesses in the other, i.e. cross-adaptation occurs. This means that two of the main genetic mechanisms thought to be important in driving specialism in pathogens would not apply: antagonistic pleiotropy (Fry 1996) and mutation accumulation (Kawecki 1994). The non-applicability of mutation accumulation is more a function of the modelling procedure than any biological factors, as we did not directly model a set of mutations without selective value. Some sites will have fitnesses close to other sites at random, but, as in this model evolution is occurring over House-of-Cards-like landscapes, mutations are uncorrelated and the effect of a mutation in one background is independent of that mutation in another background, i.e. epistasis is maximal. Therefore, there is no population of sites in which mutation does not change the underlying fitness, or change it in such a small way to be inappreciable to selection, in all backgrounds. As such, variants that are neutral on one host but advantageous on the other, a

requirement of mutation accumulation (Kawecki 1994), are highly unlikely to occur in this model. Antagonistic pleiotropy fundamentally assumes a negative correlation structure over regions of the fitness landscapes between hosts, with a proportion of mutations having effects of directly opposing signs between species. By manipulating the correlation between the fitness landscapes directly, we changed the average area of the fitness landscape that exhibits antagonistic pleiotropy, and this had the expected effect of decreasing the number of realised variants that had high fitnesses on both hosts when the correlation was low, and increasing the number of variants with high fitnesses on both hosts when it was high.

Given an adaptive walk on a paired set of landscapes with negative correlation, maintaining high fitness on both hosts provides a double challenge. Firstly, by definition, sequences with high fitness on both hosts are rare, but secondly, even if one is found, maintaining position at that site is difficult, as the site is unlikely to be optimal for either host species. Therefore, long adaptive walks in either species are likely to push the sequences away from the section of the sequence space that provides cross-adaptation. This has been repeated observed in viral experimental evolution studies (Crill, Wichman, and Bull 2000; Remold et al. 2008). The empirical results therefore suggest that the rate of cross-species transmission should be an important parameter determining the degree of specialism versus generalism observed in the viral population. Our model, consistent with other modelling literature on the generalism-specialism divide (Gandon 2004), agrees with this empirical finding. We manipulated the rate of cross-species transmission by adjusting the contact rate between host species. When the probability of conspecific interactions increased relative to heterospecific interactions, fewer variants of high fitness on both host species were observed, consistent with the theory that adaptive walks in one species were driving sequences away from regions of high fitness in both.

The statistical interaction between the degree of preferential contact between conspecifics and the correlation in fitness landscapes between hosts on the degree of generalism exhibited by variants observed in the model follows directly from the facts noted above. Long adaptive walks on a single host species lead to reductions in fitness on the linked host species. High

correlations between fitness landscapes reduce the region of the sequence space that exhibits high fitness on one host and low fitness on the other. Therefore, an adaptive walk on a highly positively correlated fitness landscape is likely to lead to a lower average change in fitness on the other host than a walk of an equivalent length on a anticorrelated landscape. This drives the observed statistical interaction, the expected change in generalism driven by preferential interaction between conspecifics depends on how correlated the underlying fitness landscapes are.

5.5.2 Variation and Rate of Adaptation

Variation is increased when there are two host types, presumably because divergent fitnesses of the same genotype between hosts exert selection after between-host transmission. When a parasite switches from one host to the other, the fitnesses are unlikely be to aligned for the reasons discussed above, when there is no correlation between the landscapes adaptive walks in one host push sequences away from regions of the fitness landscape providing coadaptation, so the variant, when transmitted into the new host species, moves away from the (potentially adapted) sequence in the original host. Fitness change on environment switch has been noted repeatedly in both pathogens, such vesicular stomatitis virus (Remold et al. 2008), and non-pathogens, such as *Pseudomonas fluorescens* evolving in low quality media (Buckling et al. 2007). This also explained the dramatic increase in the rate of adaptation seen in two-host systems versus one-host systems in this model. Simulations of adaptive walks in House-of-Cards landscapes indicate that the average walk length to a local optimum is short (Kauffman and Levin 1987). In elitist implementations, where a lower fitness sequence cannot replace a higher fitness sequence, this represents a dead end. In our implementation and others like it (Heredia et al. 2017), there exists a non-zero probability of fixation of a non-optimal sequence through drift effects, so given a long enough time horizon populations can escape from fitness optima. However, this is unlikely to be important in short run-time simulations such as ours, so a sequence becoming trapped at a local optimum is still likely to be a common occurrence.

A second host provides a pathway to escape a local optimum in the first, as noted by Cheetham (1993), with regards to fluctuating environments generally,

and Cervera, Lalić and Elena (2016), specifically in the pathogen evolution case. Because the rank correlation in fitness between the two hosts is not complete, a local optimum in one host is unlikely to represent a local optimum in another, so cross species transmission allows escape. When a cross-species transmission event occurs, the sequence can evolve down a new path from the position it starts on in the second host, then when cross-species transmission back to the original host occurs, the sequence will have escaped the local optimum. This results in a dramatic increase the in the rate of adaptation, and explains the considerably higher proportion of two host runs relative to one host runs reaching the 99.9^{th} percentile of the fitness landscape.

The underlying distribution of the fitnesses across a landscape also strongly influenced the rate of adaptation. This is likely due to the adaptation strategy chosen for the model, where the fixation probability is a function of the fitness difference between the resident and mutant sequences. The three distributions used to generate the fitnesses have very different tails. The uniform distribution used here was bounded at 1, meaning as the adaptation process proceeds, the difference between the current fitness and next highest fitness value must go towards 0. This means that probability of replacement converges to the neutral fixation probability as higher fitness values are achieved. This effect does not occur with the gamma or half-normal landscape distributions, as these distributions are unbounded to the right. The forcing of the probability of replacement towards the neutral fixation probability likely leads to the observed reduction in the rate of adaptation in the uniform distribution relative to the others. This theory about the form of the replacement probability being the cause of the differences between fitness landscapes is consistent with the gamma distributed fitness landscape having a higher rate of adaptation than the half-normal distributed landscape, as the variance of the gamma distribution is higher. So, on average the distance between variants will be higher, leading to higher acceptance probabilities.

Contrary to our stated hypothesis, the presence of a highly efficient adaptive immune system capable of completely blocking reinfection with a strain lead to a clear reduction of diversity in the two-host case, despite the fact that the lack of cross-immunity between strains would be expected to lead to stable

coexistence (Gupta 1998). While the selection between strains may be such as to overpower this effect (Cobey 2014), it would not be expected to reverse it.

5.5.3 Conclusion

With a simple model explicitly evolving a population of digital organisms across a defined sequence space, we find strong evidence that positive correlations between fitness landscapes between hosts represent a strong driver of generalism, and that the rate of contact between host species is also important. We find that evolution in multiple hosts increases diversity, with degree of increase dependent on the degree of correlation between the environments. The rate of adaptation is found to be dramatically increased when evolution occurs across multiple environments. This is likely due to the ability of sequences caught on local optima of the fitness landscape on any one host to escape these optima through cross-species transmission. The results generated here are consistent with previous literature on the topic, and this simple stereotyped model adds weight to the already strong empirical and theoretical literature on the evolution of generalism, and generation of diversity in heterogeneous environments.

Chapter 6 - General discussion

6.1 A synopsis of the results

The aim of this thesis was to explore the factors that explain the distribution and diversity of multihost pathogens in bumblebees with special focus on their viruses. Bumblebees are an important part of the pollinator community, providing pollination services, such as buzz pollination, which other pollinators are inefficient at (Heinrich 2004). Despite this, remarkably little is known about their diseases in the wild. This thesis explored both the factors that lead to infection of bumblebees in the wild and the diversity of pollinators that bumblebees are infected with. Chapter 2 described a series of new multihost pollinator viruses and used them to investigate the effects the evolutionary history of hosts and pathogens on the current distribution of disease in a natural multi-host mutli-parasite system. Chapter 3 studied four of the viruses found in Chapter 2 in more detail looking at their diversity and factors that predict their prevalence in the wild. Chapter 4 presented evidence that the use of neonicotinoid pesticides may be leading to greater infection in exposed individuals by increasing their susceptibility to disease. Chapter 5 explored the impact of evolution across correlated fitness landscapes of the type experienced by multihost parasites using digital organisms.

6.1.1 Chapter 2 - Evolution of host range: The prevalence of novel bumblebee viruses is explained by both host and pathogen phylogenies

Previously, only one virus has been initially discovered in bumblebees, which was unnamed (Clark 1982). However, viruses discovered in other species are frequently found in bumblebees (Levitt et al. 2013; McMahon et al. 2015; Singh et al. 2010). In Chapter 2, I described and confirmed by PCR 18 new bumblebee viruses, as well as describing bioinformatic evidence of the potential existence of 19 more. As some of these 19 may constitute multiple contigs from the same virus, we do not report them as separate viruses. In assessing group level prevalence by PCR, I find considerable differences in the levels of host usage across the viruses. Some viruses, such as River Luinaeg virus, and Allermuir Hill viruses 1, 2, and 3 are almost completely limited to one, two or three hosts, while others, such as Castleton Burn virus and Gorebridge virus are found at intermediate frequencies in all hosts. Large variations in the degree of generalism between pathogens are common, for instance compare the

extreme generalism of rabies virus (Lembo et al. 2007) to the near completely host limitation of hepatitis virus A to humans and non-human primates (Purcell and Emerson 2001). Interestingly, at least within the bees, strong host barriers are rarely observed, and these highly-limited cases represent the exception rather than the norm. However, the normal caveats about attempting to determine host range from point prevalence data apply, in that a non-detection from a single sample is not strong evidence of absence from a species.

I used the novel viruses to assess the extent to which the bumblebee-virus infection network is explained by the relatedness of the hosts and parasites. Most interaction studies are plagued with the issue of false negatives (Poulin et al. 2016), biasing estimates of parameters. In a manner similar in concept to Walker et al. (2017), I attempted to control for this by directly including the sampling uncertainty in a non-standard link function, hopefully rendering the results robust to this bias. I showed that related hosts are infected by similar sets of viruses, related viruses infect similar sets of hosts and that related hosts are infected by related viruses. A large proportion of the variance in infection was due to phylogenetically-correlated effects, which is consistent with the literature on host shifts, where related hosts are often found to share pathogen assemblages (e.g. Hadfield et al. 2014; Waxman et al. 2014). This is often thought to be due to biased-host switching. However, the fact that related viral sequences were found in different host species in Chapter 3 indicates that these host shifts are occurring on ecological rather than evolutionary timescales. I therefore believe that prevalences are correlated between related species by either the shared environment from the perspective of the parasite causing cross-species transmission to occur more efficiently, or by preferential mixing between related bee species at flowers driven by phylogenetically-correlated flower preferences (Goulson, Lye, and Darvill 2008; Harmon-Threatt and Ackerly 2013). Effects of the relatedness of viral species on their host assemblages are less commonly observed, potentially because they are less commonly tested for. In this case, a small effect of viral relatedness on the degree of sharing of host assemblage was observed, implying that some aspect of conserved viral physiology leads to more or less efficient reproduction in certain host species. A similarly sized coevolutionary interaction was also

observed indicating that related hosts tend to have similar prevalences of related pathogens.

6.1.2 Chapter 3 - The ecological predictors of viral infection in wild bumblebees

I performed an exploratory study to investigate the abiotic predictors of bumblebee infection using a subset of novel virus from Chapter 2 as well as two previously known viruses that infect bumblebees in the wild: Acute bee paralysis virus and Slow bee paralysis virus. Previously, the only factor that had be shown to be predictive in wild bumblebee infection was the presence of sympatric managed bees (Fürst et al. 2014). I acquired prevalence data at the individual level of four of the novel viruses and tested for their associations with daily average maximum temperature, solar radiation, vapour pressure, wind speed and precipitation during the months of sampling. As infection is likely to come about though physical contamination of flowerheads (Graystock, Goulson, and Hughes 2015; McArt et al. 2014), these abiotic factors were chosen as either they had been reported in being important in the inactivation of viruses in the environment in other studies (temperature, vapour pressure, solar radiation), or in bumblebee foraging behaviour (wind speed, precipitation). This analysis showed that River Luinaeg virus was positively associated with increased rainfall, even after strong regularisation of the posterior distribution to account for the fact that the number of site-level covariates was greater than the number of sites assayed. The model also indicated there were strong associations between the probability of infection with River Luinaeg virus and Loch Morlich virus even after accounting for environmental covariates, location and host species. I expected rainfall to decrease prevalence but the results showed the opposite, although, the results are consistent with high precipitation being a signal of a harsh environment. Condition-infection correlations have been reported in other species (e.g. bank voles (Beldomenico et al. 2008)), and as a reduction in foraging time caused by frequent rainfall could feasibly put the colony into a state of starvation, increasing its susceptibility to disease. The cause of the River Luinaeg virus-Loch Morlich virus association is uncertain, though I personally favour an explanation of facilitation between the viruses through one or the other’s actions leading to coinfection (Eswarappa et al. 2012; Pedersen and Fenton 2007). I favour this explanation predominately due to the

flaws in the other possible explanations in this case. The association does not appear to seem to represent a general immunosuppression, as there was no correlated increase in infection with Mayfield virus 1 and Mayfield virus 2. And the viruses are not likely to be replication deficient (López-Ferber et al. 2003), or satellite viruses of one another, as both occur in single infections.

I also show that the diversity of the novel viruses found is greater than that of the previously known bee viruses: Acute bee paralysis virus (ABPV) or Slow bee paralysis virus (SBPV). ABPV was lacking in diversity that of the 38 sequences sampled, there were no detectable genetic differences. Both ABPV and SBPV were also at high prevalence in the sampled populations. I believe the most parsimonious explanation in this case, is that there was a strong bottlenecking event that removed diversity followed by an epidemic in these species, similar to the outbreak of Dengue in Singapore described by Hapuarachchi et al. (2016). However, the results could also be consistent with a selective sweep associated with a highly fit new viral variant causing an epidemic. A possible argument against this is that sampling in 2015 found roughly equivalent prevalences of SBPV in Scottish *Bombus terrestris* populations (Manley 2017), implying that the prevalences were not unusually high during sampling. As between sampling year sequences were not generated for SPBV and ABPV, it is impossible to tell whether a different strain dominates in different years, or the same strains become prevalent every year. A randomisation test indicated that, after accounting for phylogenetic uncertainty, Mayfield virus 1 variants assorted non-randomly with host species, with the set of sequences designated Clade A preferentially assorting with *B. terrestris* individuals, and Clade B showing a lack of *B. terrestris* infection. Sampling of the Clade A was not intense enough to determine whether sequences infecting other *Bombus* species are also present in it at intermediate frequencies, but falling within the clade was one *Bombus lucorum* infecting variant, so there is some degree of cross species transmission, even within that set of variants. Interestingly, this could imply that Mayfield virus 1 actually consists of two phenotypically different, but genetically similar strains, one which is a psuedospecialist on *B. terrestris* and one that is a generalist all infected species engaging in frequent cross-species transmission.

6.1.3 Chapter 4 - Pulsed pesticide exposure may increase the rate of pathogen uptake in *Bombus terrestris* bumblebees

Chapter 4 investigated the effect of field realistic doses of the neonicotinoid pesticide clothianidin on the update of pathogens from the environment by *B. terrestris* workers in order to determine whether pesticides are likely to be having a major effect on the distribution of pesticides in the wild. Pesticide studies in the past have illustrated negative effects on survivorship (Doublet et al. 2015), and other infection endpoints (Di Prisco et al. 2013; Vidau et al. 2011) when the pesticide is coincident with infection. We found that exposure to a pulse of neonicotinoid at field realistic levels provides small increases to the rate of infection in exposed colonies.

6.1.4 Chapter 5 - Between-host fitness landscape correlation drives multiple outcomes in the evolution of multihost digital parasites

In Chapter 5, I developed and applied a highly stereotyped model of the evolution of avirulent digital organisms across a multihost fitness landscape. I explored how evolution over a fluctuating fitness landscape with and without an immune system capable of immunological memory would affect the rate of adaptation, the propensity for generalism and the amount of variation generated. I find that parasite diversity increases in cases when the probability of infection is high, and therefore there is large stable infected host population. Additionally, parasite diversity is higher with low interspecific contact, in multihost relative to single host infections, and with low correlations in the fitness landscapes between host species. The results from the model align with expectations from other theoretical studies either directly, or by analogy. Fluctuation in environmental conditions is known to maintain variation (reviewed in Kassen 2002). Lower rates of interaction between host species result in lower rates of cross species transmission. The displacement from a fitness optimum by cross-species transmission appears to be the main driver of the generation of variation in the system, so correspondingly reduced rates of cross species transmission lead to less genetic variants being maintained. This displacement on cross-species transmission also appears to increase the rate of adaptation by displacing the variants from local optima and allowing the adaptive walk to continue consistent with work by Cheetham (1993) and Cervera et al. (2016).

6.2 Implications and open questions

The discovery of a new set of multihost viruses of bumblebees increases our understanding of the viral biodiversity present in this important group, which has value in and of itself. However, there are also practical benefits. The transfer of managed bumblebee colonies around the world is leading to increased disease transmission between distant areas (Goka, Okabe, and Yoneda 2006; Graystock et al. 2013). The discovery of these viruses will allow breeding facilities to add them to their routine testing in order to minimise the spread of disease to non-exposed wild populations. As these viruses are only known from bumblebees, this also allows studies to be performed where bumblebee viruses are experimentally infected into honeybees to help determine what the risk might be if a bumblebee virus emerged into the managed honeybee population.

The implication that exposure to field realistic doses of pesticide increases susceptibility to a broad suite pathogens has potentially important implications for agriculture and conservation. While a moratorium on the use of neonicotiniods such as clothianidin has been issued in the EU, they are still commonly used in other countries, and this common practice might be contributing to the ongoing declines in bumblebee populations (Williams and Osborne 2009). The size of the effect could not be accurately estimated in Chapter 4, and clearly this is a very important parameter both for the epidemiology of bumblebee disease, and for conservation purposes. If the effect is small, then it is likely insignificant in the face of factors like land-use changes (Vanbergen and the Insect Pollinators Initiative 2013) and the direct impact of pesticides themselves on colony growth and output (e.g. Whitehorn et al. 2012). However, given that the possibility exists that the effect is large, further studies should be performed in order to quantify the risk.

The model I presented in Chapter 5 came to conclusions broadly similar with the literature on evolution in heterogeneous environments. Given the importance of epistasis between landscapes in the evolution of RNA viruses (Elena, Solé, and Sardanyés 2010), an extension of the model to handle virulent pathogens may give interesting results. Especially as the literature on virulent pathogens makes very different predictions about the maintenance of diversity (e.g. Sasaki 2000).

6.3 Conclusion

This thesis has investigated the factors influencing the diversity and distribution of multihost viruses in bumblebees. Chapter 2 described a new collection of bumblebee viruses, and use these viruses to show that evolutionary histories of the hosts and parasites have impacts on prevalence in the wild. Chapter 3 showed that abiotic factors may affect the distribution of infection across space, and that there is considerably less diversity in ABPV and SBPV than is present in the novel viruses tested. Chapter 4 found that field realistic doses of the neonicotinoid pesticide clothianidin likely increase susceptibility to infection with a wide panel of infectious agents. Chapter 5 used a simple model of evolution across fitness landscapes to show that, consistent with theory, cross-species transmission is likely to increase the rate of adaptation, increase variation and at high rates of switching promote generalism. In conclusion, there are a complex array of factors that influence the diversity and distribution of bumblebee viruses in natural communities, and we found evidence for viral genetics, host genetics, the weather and pesticides all playing a role.

Appendices

Chapter 2 Appendix

Table A2.1 Locations of sites where bees were collected.

Location	Lat	Long
Dalwhinnie	56.935	-4.245
Edinburgh	55.921	-3.179
Glenmore	57.166	-3.696
Gorebridge	55.826	-3.046
Iona	56.328	-6.406
Ochils	56.158	-3.88
Pentlands	55.796	-3.404
Staffa	56.436	-6.341
Stirling	56.149	-3.916

Table A2.2 The number of bees in each pool and their species.

Pool ID	Species	Sample Size	Location
BO	*Bombus bohemicus*	10	Gorebridge, the Ochils, the Pentlands and Glenmore
C1	*Bombus cryptarum*	10	Ochills
C2	*Bombus cryptarum*	10	Ochills
C3	*Bombus cryptarum*	5	Iona, the Ochils and the Pentlands
C4	*Bombus cryptarum*	8	Gorebridge and Edinburgh
DIV	Mixed	7	Gorebridge, Staffa, the Ochils and the Pentlands
H1	*Bombus hortorum*	10	Gorebridge
H2	*Bombus hortorum*	10	Gorebridge
H3	*Bombus hortorum*	10	Gorebridge
H4	*Bombus hortorum*	10	Gorebridge
H5	*Bombus hortorum*	10	Gorebridge
H6	*Bombus hortorum*	6	Gorebridge
H7	*Bombus hortorum*	7	Gorebridge
H8	*Bombus hortorum*	7	Stirling, Glenmore and Dalwhinnie
H9	*Bombus hortorum*	11	Iona
J1	*Bombus jonellus*	10	Stafa
J2	*Bombus jonellus*	10	Stafa
J3	*Bombus jonellus*	5	Iona and Staffa
J4	*Bombus jonellus*	10	Glenmore
J5	*Bombus jonellus*	10	Glenmore
J6	*Bombus jonellus*	10	Glenmore

J7	*Bombus jonellus*	5	Dalwhinnie and Glenmore
L1	*Bombus lucorum*	10	Stirling and the Ochils
L2	*Bombus lucorum*	10	Stirling and the Ochils
L3	*Bombus lucorum*	10	Gorebridge
L4	*Bombus lucorum*	10	Gorebridge
L5	*Bombus lucorum*	10	Gorebridge
L6	*Bombus lucorum*	10	Gorebridge
L7	*Bombus lucorum*	10	Gorebridge
L8	*Bombus lucorum*	10	Gorebridge
L9	*Bombus lucorum*	10	Gorebridge
L10	*Bombus lucorum*	10	Gorebridge
L11	*Bombus lucorum*	8	Gorebridge
L12	*Bombus lucorum*	10	Edinburgh
L13	*Bombus lucorum*	10	Edinburgh
L14	*Bombus lucorum*	10	Edinburgh
L15	*Bombus lucorum*	10	Stirling, the Ochils and the Pentlands
L16	*Bombus lucorum*	10	Iona, the Ochils and the Pentlands
L17	*Bombus lucorum*	2	The Ochils
L18	*Bombus lucorum*	10	Edinburgh
L19	*Bombus lucorum*	10	Edinburgh
L20	*Bombus lucorum*	2	Edinburgh
LA1	*Bombus lapidarius*	11	Gorebridge
LA2	*Bombus lapidarius*	8	Gorebridge
LA3	*Bombus lapidarius*	8	Edinburgh
LA4	*Bombus lapidarius*	7	Stirling and the Ochils
M	*Bombus monticola*	10	Glenmore, the Ochils and the Pentlands
MA	*Bombus magnus*	11	Iona, the Ochils and Glenmore
P1	*Bombus pascuorum*	10	Stirling
P2	*Bombus pascuorum*	10	Ochills
P3	*Bombus pascuorum*	9	Iona, Staffa and the Pentlands
P4	*Bombus pascuorum*	10	Ochills
P5	*Bombus pascuorum*	10	Ochills
P6	*Bombus pascuorum*	10	Ochills
P7	*Bombus pascuorum*	8	Stirling
P8	*Bombus pascuorum*	10	Ochills
P9	*Bombus pascuorum*	5	Ochills
P10	*Bombus pascuorum*	10	Edinburgh
P11	*Bombus pascuorum*	10	Edinburgh
P12	*Bombus pascuorum*	10	Edinburgh
P13	*Bombus pascuorum*	10	Edinburgh
P14	*Bombus pascuorum*	10	Edinburgh

P15	*Bombus pascuorum*	5	Edinburgh
P16	*Bombus pascuorum*	10	Gorebridge
P17	*Bombus pascuorum*	10	Gorebridge
P18	*Bombus pascuorum*	10	Gorebridge
P19	*Bombus pascuorum*	10	Gorebridge
P20	*Bombus pascuorum*	10	Gorebridge
P21	*Bombus pascuorum*	10	Gorebridge
P22	*Bombus pascuorum*	10	Gorebridge
P23	*Bombus pascuorum*	5	Gorebridge
PR1	*Bombus pratorum*	10	Edinburgh
PR2	*Bombus pratorum*	10	Edinburgh
PR3	*Bombus pratorum*	9	Edinburgh
PR4	*Bombus pratorum*	10	Gorebridge
PR5	*Bombus pratorum*	8	Gorebridge, Stirling and the Ochils
T1	*Bombus terrestris*	10	Stirling and the Ochils
T2	*Bombus terrestris*	10	Gorebridge
T3	*Bombus terrestris*	10	Gorebridge
T4	*Bombus terrestris*	10	Gorebridge
T5	*Bombus terrestris*	10	Gorebridge
T6	*Bombus terrestris*	10	Gorebridge
T7	*Bombus terrestris*	10	Gorebridge
T8	*Bombus terrestris*	10	Gorebridge and Edinburgh
T9	*Bombus terrestris*	10	Edinburgh
T10	*Bombus terrestris*	10	Edinburgh
T11	*Bombus terrestris*	10	Edinburgh
T12	*Bombus terrestris*	10	Edinburgh
T13	*Bombus terrestris*	10	Edinburgh
T14	*Bombus terrestris*	10	Stirling and the Ochils
T15	*Bombus terrestris*	10	Stirling and the Ochils
T16	*Bombus terrestris*	5	Stirling and the Ochils
T17	*Bombus terrestris*	10	Gorebridge
T18	*Bombus terrestris*	10	Gorebridge
T19	*Bombus terrestris*	8	Gorebridge
T20	*Bombus terrestris*	10	Edinburgh
T21	*Bombus terrestris*	10	Edinburgh
T22	*Bombus terrestris*	6	Edinburgh
T23	*Bombus terrestris*	10	Gorebridge
T24	*Bombus terrestris*	10	Gorebridge
T25	*Bombus terrestris*	10	Gorebridge
T26	*Bombus terrestris*	10	Gorebridge

Table A2.3 The PCR conditions for all reactions. Conditions code are given below the table.

Primer	Virus	Sequence	Conditions	Reference
MV1Forward	Mayfield virus 1	TATCCGCCGGCGT AATCTTC	PCR 60	This study
MV1Reverse	Mayfield virus 1	GGATCTGATCCGTA GCGTGG		
MV2Forward	Mayfield virus 2	CGGCTGCGTTGCG TAGTATA	T62	This study
MV2Reverse	Mayfield virus 2	ACCTGCCGTGCTA ACAAATA		
AHV1Foward	Allermuir Hill virus 1	TGGGGAAGGAATA TTTGCAGGT	T62	This study
AHV1Reverse	Allermuir Hill virus 1	GGCATCTTTGAAGA TAACCTACGC		
AHV2Foward	Allermuir Hill virus 2	TGCTGGTGCTGAT GTTACATCT	T62	This study
AHV2Reverse	Allermuir Hill virus 2	TTCGAAACACAACT GCAATACA		
AHV3Foward	Allermuir Hill virus 3	GGGGCTTCGGCTG AATCTAG	PCR 60	This study
AHV3Reverse	Allermuir Hill virus 3	TGCAAACAATTAGA GTTGGCCA		
MLVFoward	Mill Lade virus	TCTCGCAATCCATA CGTACTTCA	T62	This study
MLVReverse	Mill Lade virus	CGTCAACAAGGTC GTTTCTTCC		
CBVForward	Castleton Burn virus	TTCTCTATCGAGCG GCCTTG	T62	This study
CBVReverse	Castleton Burn virus	TGTGCTTCCATGTA GGCGAA		
SVForward	Sheriffmuir virus	GTTTTGACCAGCA CCAGAGC	PCR 60	This study
SVReverse	Sheriffmuir virus	CCAGTTCGGGGTG GCTAAAT		
DVForward	Dumyat virus	CGGAGATAACGGA GGTGTGG	T62	This study
DVReverse	Dumyat virus	GCAAGGAGACAAG GCTCCTT		
CMVForward	Cnoc Mor virus	TCAGCCGAATTAG AATGTGTACA	T62	This study
CMVReverse	Cnoc Mor virus	GCGCTTTCGAATA GATGCCT		
ARVForward	Agassiz Rock virus	TGAATGGTAGGAG CATGCGT	PCR 60	This study
ARVReverse	Agassiz Rock virus	TGTAGTAGATGCCT GGGTTTGA		
ELVForward	Elf Loch virus	GAACAAGCGCGAG TGGAAAC	PCR 62	This study
ELVReverse	Elf Loch virus	TCGAGATTATCTGC GTGGCC		
CCVForward	Clamshell Cave virus	GGCCTCAAGGTAT GTTGAATAACA	T62	This study

CCVReverse	Clamshell Cave virus	TGCACTTATTATCT GCTGTCTTAGA	T62	This study
BBVForward	Boghill Burn virus	TGGATCACCTGAT GGATTCCT	T62	This study
BBVReverse	Boghill Burn virus	TCGATCTCTCTGTG AGTCTCTGT		
BHVForward	Black Hill virus	ACATCTATGTGCTG CAGCGA	T62	This study
BHVReverse	Black Hill virus	CCATGACCGTGTG CTAGCAT		
RLVForward	River Liunaeg virus	ACCAGGTGGAACT CGTGTTT	PCR 60	This study
RLVReverse	River Liunaeg virus	GTACTCTGGACCT TTGCCGT		
LMVForward	Loch Morlich virus	AGTGGTGGAGATG GAGACGA	PCR 60	This study
LMVReverse	Loch Morlich virus	CCACAGATACCAG TGGCGTA		
GVForward	Gorebridge virus	GGATAGATACACTA AAGGATGCTAAAA	T62	This study
GVReverse	Gorebridge virus	ATTCGGTGCATCAA GGAGCA		
HPLV34Forward	Hubei partiti-like virus 34	TGGCTTAGATTTAA TGCTACGAT	T62	This study
HPLV34Reverse	Hubei partiti-like virus 34	CCTCATAGCTCCAC CAGTAACC		
SBPV 774F	Slow bee paralysis virus	GAGATGGATMGRC CTGAAGG	T62	Lena Wilfert (*pers comm*)
SBPV 1698R	Slow bee paralysis virus	CATGAGCCAKGAR TGTGAA		
ABPV 5088F	Acute bee paralysis virus	CYATGGACACACC CTATGTG	T62	Lena Wilfert (*pers comm*)
ABPV 6122R	Acute bee paralysis virus	CGCCATTTTGCTAC TTCTCC		

PCR 62: 95°C 5:00, (95°C 0:15, 62°C 0:30, 72°C 0:45)x40, 72°C 7:00, 4°C
PCR 60: 95°C 5:00, (95°C 0:15, 60°C 0:30, 72°C 0:45)x40, 72°C 7:00, 4°C
T62: 95°C 5:00, (95°C 0:15, 62°C [minus 1°C per step] 0:30, 72°C 0:45)x10, (95°C 0:15, 52°C 0:30, 72°C 0:45)x30, 72°C 7:00, 4° C

Table A2.4 The genbank numbers IDs of the sequences used to build the host tree

	16S ribosomal RNA gene	**Cytochrome oxidase subunit I (COI) gene**	**Phosphoenol pyruvate carboxykinase**	**Long-wavelength rhodopsin**	**Elongation factor-1 alpha gene**	**Arginine kinase gene**
Bombus terrestris	DQ788118.1	AY181170.1	EF050865.1	AF493022.1	DQ788288.1	AF492888.1
Bombus cryptarum	DQ787995.1	AY181100.1	EF050855.1	AY739461.1	DQ788175.1	DQ788416.1
Bombus lucorum	DQ788051.1	AY181120.1	EF050857.1	AF493021.1	DQ788225.1	AF492887.1
Bombus patagiatus	DQ788078.1	AF279499.1	EF050862.1	AF493020.1	DQ788252.1	AF492886.1
Bombus lapidarius	DQ788045.1	AY181115.1	EF050902.1	AF493005.1	DQ788219.1	AF492871.1

Bombus monticola	DQ788064.1	AY181132.1	EF050848.1	AY739483.1	DQ788238.1	DQ788466.1
Bombus pascuorum	DQ788077.1	AY181137.1	EF050932.1	AF493001.1	DQ788251.1	AF492867.1
Bombus hortorum	DQ788024.1	AY181105.1	EF050999.1	AF492987.1	DQ788200.1	AF492853.1
Bombus pratorum	DQ788087.1	AY181149.1	EF050819.1	AF493033.1	AF492966.1	AF492899.1
Bombus jonellus	DQ788039.1	AY181113.1	EF050814.1	AY739473.1	DQ788214.1	DQ788446.1
Bombus bohemicus	DQ787980.1	AY181181.1	EF050979.1	AF492992.1	AF492925.1	AF492858.1
Apis mellifera	L06178.1	L06178.1				
Episyrphus balteatus	AY573115.1	KR262632.1				
Drosophila melanogaster	KY559386.1	KY559386.1				
Anopheles dirus	JX219731.1	JX219731.1				
Bombyx mori	KM875545.1	KM875545.1				
Heliconius melpomene	KP100653.1	KP100653.1				
Allantus luctifer	KJ713152.1	KJ713152.1				
Philotrypesis pilosa	JF808723.1	JF808723.1				
Bombus magnus		AY181123.1				
Eristalis pertinax		KX055520.1				

Table A2.5 The genbank IDs of the sequences used to build the viruses tree.

Genbank ID	Virus	Negative Sense
NP_066241.1	ABPV	
This study	Allermuir Hill virus 1	
This study	Allermuir Hill virus 2	
This study	Allermuir Hill virus 3	
JX045858.1	Aphid lethal paralysis virus	
NC_032494.1	Beihai paphia shell virus 2	
NC_032972.1	Beihai sobemo-like virus 8	
YP_009336943.1	Beihai weivirus-like virus 17	
YP_009337445.1	Beihai zhaovirus-like virus 4	
JF423196.1	Big Sioux River virus	
NC_003784.1	Black queen cell virus	
This study	Boghill Burn virus	
NC_024297.1	Bovine astrovirus strain BAstV-GX7/CHN/2014	

This study	Castleton Burn virus	
APG79015.1	Changjiang picorna-like virus 17	
KY937971.1	Chronic bee paralysis virus segment	
This study	Clamshell Cave virus	x
NC_003924.1	Cricket paralysis virus	
AF188515.1	Cryphonectria hypovirus 3	
JQ413340.1	Deformed wing virus A	
AY251269.2	Deformed wing virus B	
Unpublished, Lena Wilfert (*pers comm*)	Deformed wing virus C	
NC_004169.1	Drosophila x virus	
This study	Dumyat virus	
This study	Elf Loch virus	
APT68154.1	Ganda bee virus	x
This study	Gorebridge virus	
NC_003607.2	Helminthosporium victoriae virus	
NC_032203.1	Hubei astro-like virus	
NC_033015	Hubei chuvirus-like virus 3	
APG78254.1	Hubei odonate virus 13	
APG78322.1	Hubei partiti-like virus 34	
APG77985.1	Hubei picorna-like virus 15	
APG78030.1	Hubei picorna-like virus 57	
YP_009336934.1	Hubei tombus-like virus 29	
KX578271.1	Iflavirus sp. isolate VDV-2	
NC_001916.1	Infectious pancreatic necrosis virus	
KX421583.1	Israeli acute paralysis virus	
NC_004807.1	Kashmir bee virus	
HQ871931.2	Lake Sinai virus 1	
HQ888865.2	Lake Sinai virus 2	
KR021357.1	Lake Sinai virus 6	
This study	Loch Morlich virus	
This study	Mayfield virus 1	
This study	Mayfield virus 2	
This study	Mill Lade virus	
AMO03223.1	Muthill virus	
AF174533.1	Nodamura virus	
M87661.2	Norwalk virus	
This study	River Liunaeg virus	

NC_001545.2	Rubella virus	
NP_049374.1	Sacbrood virus	
YP_003622540.1	SBPV	
KY053857.1	Scaldis River bee virus	x
AOF41423.1	Seattle Prectang virus	x
This study	Sheriffmuir virus	
NC_001547.1	Sindbis virus	
KJ556849.1	Tobacco ringspot virus isolate SK	
KX578272.1	Varroa destructor virus 3	
NC_033164.1	Wenzhou shrimp virus 10	
NC_033300.1	Wenzhou shrimp virus 9	
YP_009304995.1	Wuchang Cockroach Virus 1	x
NC_033764.1	Wuhan cricket virus 2	
NC_033473.1	Wuhan insect virus 14	
KX884291.1	Wuhan spirurian nematodes virus 1	
YP_009345041.1	Xinzhou nematode virus 1	

A2.6 Aggregate prevalence analysis

Introduction and Methods

A reanalysis of the data using an aggregated presence-absence matrix for each host-virus combination was performed during corrections. The methods used were the same as presented for the main text, excepting the changes outlined below:

As a reduction to a presence-absence matrix removes the replication within host-virus combinations, both the effect of pool ID and spatial composition are no longer identified, so these were removed from the model. All other terms were maintained. A full phylogenetic uncertainty analysis was not performed, only the maximum clade credibility trees for the hosts and viruses were used. These trees were calculated using the median node heights. Only the model using both the positive and negative sense viruses was fitted, as this maximises the amount of data available to fit the model.

The model fitted is shown below, where y' is the value of the latent variable, i is the index of the data point, μ is the global mean of the latent variable, ε is a fixed unidentified error term and all other effects are estimated by partial pooling:

$$y'_i = \mu + host_i + virus_i + host\ phylogenetic\ effect_i + virus\ phylogenetic\ effect_i + host\ evolutionary\ assemblage\ effect_i + virus\ evolutionary\ assemblage\ effect_i + coevolutionary\ interaction_i + \varepsilon$$

Results and Discussion

The reduction in the size of the dataset due to the aggregation meant that (even with the maximal data inclusion gained by the addition of the negative sense virus) there was not

enough information left in the data to fit the model, and the model was not capable of converging with the weak priors provided. Given that aggregation constituted a loss of approximately 90% of the data, this is not a surprising outcome. While the model would be capable of converging to a posterior distribution if more informative priors were used, this would not be a justified approach considering the lack of certainty about what the correct parameter values should be *a priori*.

As such, no conclusions can be drawn from this aggregate analysis.

Chapter 3 Appendix

Table A3.1 The PCR conditions for all reactions. Conditions code are given below the table.

Primer	Virus	Sequence	Conditions	Reference
RLVForward	River Liunaeg virus	ACCAGGTGGAACT CGTGTTT	T62	This study
RLVReverse	River Liunaeg virus	GTACTCTGGACCT TTGCCGT		
LMVForward2	Loch Morlich virus	GCATGTGGCTCATT TTTGTTCC	T62	This study
LMVReverse2	Loch Morlich virus	AGTCCAGAGGAGA AGCACAA		
MayGenericForward	Mayfield generic	TGAGACAACTCGG CCATGAC	T62	This study
MayGenericReverse	Mayfield generic	CTATGCCGATGGTT TGCGTG		
SBPV 774F	Slow bee paralysis virus	GAGATGGATMGRC CTGAAGG	T62	Lena Wilfert (*pers comm*)
SBPV 1698R	Slow bee paralysis virus	CATGAGCCAKGAR TGTGAA		
ABPV 5088F	Acute bee paralysis virus	CYATGGACACACC CTATGTG	T62	Lena Wilfert (*pers comm*)
ABPV 6122R	Acute bee paralysis virus	CGCCATTTTGCTAC TTCTCC		

T62: 95°C 5:00, (95°C 0:15, 62°C [minus 1°C per step] 0:30, 72°C 0:45)x10, (95°C 0:15, 52°C 0:30, 72°C 0:45)x30, 72°C 7:00, 4° C

Chapter 4 Appendix

Table A4.1 The PCR conditions for all reactions. Conditions code are given below the table.

Primer	Virus	Sequence	Conditions	Reference
MV1Forward	Mayfield virus 1	TATCCGCCGGCGT AATCTTC	T62	This study
MV1Reverse	Mayfield virus 1	GGATCTGATCCGTA GCGTGG		
MV2Forward	Mayfield virus 2	CGGCTGCGTTGCG TAGTATA	T62	This study
MV2Reverse	Mayfield virus 2	ACCTGCCGTGCTA ACAAATA		
AHV1Foward	Allermuir Hill virus 1	TGGGGAAGGAATA TTTGCAGGT	T62	This study

			T62	This study
AHV1Reverse	Allermuir Hill virus 1	GGCATCTTTGAAGA TAACCTACGC		
MLVFoward	Mill Lade virus	TCTCGCAATCCATA CGTACTTCA	T62	This study
MLVReverse	Mill Lade virus	CGTCAACAAGGTC GTTTCTTCC		
CBVForward	Castleton Burn virus	TTCTCTATCGAGCG GCCTTG	T62	This study
CBVReverse	Castleton Burn virus	TGTGCTTCCATGTA GGCGAA		
SVForward	Sheriffmuir virus	GTTTTGACCAGCA CCAGAGC	PCR 60	This study
SVReverse	Sheriffmuir virus	CCAGTTCGGGGTG GCTAAAT		
DVForward	Dumyat virus	CGGAGATAACGGA GGTGTGG	T62	This study
DVReverse	Dumyat virus	GCAAGGAGACAAG GCTCCTT		
CMVForward	Cnoc Mor virus	TCAGCCGAATTAG AATGTGTACA	T62	This study
CMVReverse	Cnoc Mor virus	GCGCTTTCGAATA GATGCCT		
ARVForward	Agassiz Rock virus	TGAATGGTAGGAG CATGCGT	PCR 60	This study
ARVReverse	Agassiz Rock virus	TGTAGTAGATGCCT GGGTTTGA		
ELVForward	Elf Loch virus	GAACAAGCGCGAG TGGAAAC	PCR 62	This study
ELVReverse	Elf Loch virus	TCGAGATTATCTGC GTGGCC		
CCVForward	Clamshell Cave virus	GGCCTCAAGGTAT GTTGAATAACA	T62	This study
CCVReverse	Clamshell Cave virus	TGCACTTATTATCT GCTGTCTTAGA		
BBVForward	Boghill Burn virus	TGGATCACCTGAT GGATTCCT	T62	This study
BBVReverse	Boghill Burn virus	TCGATCTCTCTGTG AGTCTCTGT		
GVForward	Gorebridge virus	GGATAGATACACTA AAGGATGCTAAAA	T62	This study
GVReverse	Gorebridge virus	ATTCGGTGCATCAA GGAGCA		
HPLV34Forward	Hubei partiti-like virus 34	TGGCTTAGATTTAA TGCTACGAT	T62	This study
HPLV34Reverse	Hubei partiti-like virus 34	CCTCATAGCTCCAC CAGTAACC		
ABPV_LF	Acute bee paralysis virus	TGAAACGGAACAA ATCACCA	PCR 57	Lena Wilfert (*pers comm*)
ABPV_LR	Acute bee paralysis virus	TTCGCCAC CTTGTTAACTCC		
SBPV611Roth_F	Slow bee paralysis virus-Rothamsted	AGGTGAGGCTGCT AATTCAAT	PCR 60	Lena Wilfert (*pers comm*)
SBPV789Roth_R	Slow bee paralysis virus-Rothamsted	TCGAGACAAGCTC CATAGACA		

SBPV307Harp F	Slow bee paralysis virus-Harpenden	TGGCATTGAAGGT GAGCAAC	PCR 60	Lena Wilfert (*pers comm*)
SBPV459Harp R	Slow bee paralysis virus-Harpenden	GCTAAATCTGGAGT TGCAGC		
BQCV-B8150	Black queen cell virus	GGAGGTGAAGTGG CTATATC	PCR 57	Lena Wilfert (*pers comm*)
BQCV-F7893	Black queen cell virus	AGTGGCGGAGATG TATGC		
SBV1F	Sacbrood virus	ACCACCCGATTCC TCAGTAG	PCR Neil	Nielsen et al. [1]
SBV1R	Sacbrood virus	CCTTGGAACTCTG CTGTGTA		
DWV-F1a	Deformed wing virus-A	GGAAACATCTGGA ATTAGCGACAAA	PCR 57	Lena Wilfert (*pers comm*)
DWVDV-7A-R	Deformed wing virus-A	AATCCGTGAATATA GTGTGAGG		
VDV-F1A	Deformed wing virus-B	GAAAACATTTGGAA TTAGCAACGAC	PCR 57	Lena Wilfert (*pers comm*)
DWVDV-7A-R	Deformed wing virus-B	AATCCGTGAATATA GTGTGAGG		
Mnceranae-F	*Nosema ceranae*	CGTTAAAGTGTAGA TAAGATGTT	PCR N	Fries et al. [2]
Mnuniv-R	*Nosema ceranae*	GACTTAGTAGTAGC CGTCTCTCT		
Mnapis-F	*Nosema apis*	GCATGTCTTTGAC GTACTATG	PCR N	Fries et al. [2]
Mnuniv-R	*Nosema apis*	GACTTAGTAGTAGC CGTCTCTCT		
Mnbombi-F	*Nosema bombi*	TTTATTTTATGTRYA CMGCAG	PCR N	Fries et al. [2]
Mnuniv-R	*Nosema bombi*	GACTTAGTAGTAGC CGTCTCTCT		
CB_ITS1-F	*Crithidia bombi*	GGAAACCACGGAA TCACATAGACC	PCR C	Schmid-Hempel and Tognazzo [3]
CB_ITS1-R	*Crithidia bombi*	AGGAAGCCAAGTC ATCCATCG		

PCR 62: 95°C 5:00, (95°C 0:15, 62°C 0:30, 72°C 0:45)x40, 72°C 7:00, 4°C
PCR 60: 95°C 5:00, (95°C 0:15, 60°C 0:30, 72°C 0:45)x40, 72°C 7:00, 4°C
T62: 95°C 5:00, (95°C 0:15, 62°C [minus 1°C per step] 0:30, 72°C 0:45)x10, (95°C 0:15, 52°C 0:30, 72°C 0:45)x30, 72°C 7:00, 4°C
PCR 57: 95°C 3:00, (95°C 0:15, 57°C 0:15, 72°C 0:30)x35, 72°C 7:00, 4°C
PCR C: 95°C 1:00, (95°C 0:15, 56°C 0:15, 72°C 1:00)x40, 72°C 7:00, 4°C
PCR N: 95°C 2:00, (94°C 0:30, 60.7°C 0:30, 72°C 1:00)x40, 72°C 3:00, 4°C
PCR Niel: 94°C 5:00, (94°C 0:30, 56°C 0:45, 72°C 1:00)x40, 72°C 7:00, 4°C

Bibliography

Abascal, Zardoya, and Telford. 2010. “TranslatorX: Multiple Alignment of Nucleotide Sequences Guided by Amino Acid Translations.” *Nucleic Acids Research* 38(Supplement 2):W7–13.

Agoti et al. 2014. “Examining Strain Diversity and Phylogeography in Relation to an Unusual Epidemic Pattern of Respiratory Syncytial Virus (RSV) in a Long-Term Refugee Camp in Kenya.” *BMC Infectious Diseases* 14(1):178.

Alaux et al. 2010. “Interactions between Nosema Microspores and a Neonicotinoid Weaken Honeybees (Apis Mellifera).” *Environmental Microbiology* 12(3):774–82.

Altschul, Gish, Miller, Myers, and Lipman. 1990. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology* 215(3):403–10.

Amiri, Meixner, and Kryger. 2016. “Deformed Wing Virus Can Be Transmitted during Natural Mating in Honey Bees and Infect the Queens.” *Scientific Reports* 6(1):33065.

Anderson and Gibbs. 1988. “Inapparent Virus Infections and Their Interactions in Pupae of the Honey Bee (Apis Mellifera Linnaeus) in Australia.” *Journal of General Virology* 69(7):1617–25.

Anderson and May. 1992. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press.

Antonovics et al. 2017. “The Evolution of Transmission Mode.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 372(1719): 20160083.

Araujo et al. 2015. “Undestanding Host-Switching by Ecological Fitting.” *PLoS ONE* 10(10):e0139225.

Arbulo, Salvarrey and Invernizzi. 2011. “Proboscis Length and Resource Utilization in Two Ruguayan Bumblebees: Bombus Atratus Franklin and Bombus Bellicosus Smith (Hymenoptera: Apidae).” *Neotropical Entomology* 40(4):483–88.

Arneberg, Skorping, Grenfell and Read. 1998. “Host Densities as Determinants of Abundance in Parasite Communities.” *Proceedings of the Royal Society B: Biological Sciences* 265(1403):1283–89.

Bacandritsos et al. 2010. “Sudden Deaths and Colony Population Decline in Greek Honey Bee Colonies.” *Journal of Invertebrate Pathology* 105(3):335–40.

Bailey, Ball, Carpenter, and Woods. 1980. “Small Virus-like Particles in Honey Bees Associated with Chronic Paralysis Virus and with a Previously Undescribed Disease.” *Journal of General Virology* 46(1):149–55.

Bailey, Ball, and Woods. 1976. “An Iridovirus from Bees.” *Journal of General Virology* 31(3):459–61.

Bailey, Carpenter, Govier, and Woods. 1980. “Bee Virus Y.” *Journal of General Virology* 51(2):405–7.

Bailey, Carpenter, and Woods. 1979. “Egypt Bee Virus and Australian Isolates of Kashmir Bee Virus.” *Journal of General Virology* 43(3):641–47.

Bailey, Carpenter, and Woods. 1981. “Properties of a Filamentous Virus of the Honeybee, Apis Mellifera.” *Virology* 114(1):1–7.

Bailey, Carpenter, and Woods. 1982. “A Strain of Sacbrood Virus from Apis Cerana.” *Journal of Invertebrate Pathology* 39(2):264–65.

Bailey and Gibbs. 1964. “Acute Infection of Bees with Paralysis Virus.” *Journal of Insect Pathology* 6(4):395–407.

Bailey, Gibbs, and Woods. 1963. “Two Viruses from Adult Honey Bees (Apis Mellifera Linnaeus).” *Virology* 21(3):390–95.

Bailey, Gibbs, and Woods. 1964. “Sacbrood Virus of the Larval Honey Bee (Apis Mellifera Linnaeus).” *Virology* 23(3):425–29.

Bailey and Woods. 1974. “Three Previously Undescribed Viruses from the Honey Bee.” *Journal of General Virology* 25(2):175–86.

Bailey and Woods. 1977. “Two More Small RNA Viruses from Honey Bees and Further Observations on Sacbrood and Acute Bee-Paralysis Viruses.”

Journal of General Virology 37(1):175–82.
Bailey and Ball. 1991. *Honey Bee Pathology*. 2nd Edition. London: Academic Press Inc.
Ball, Overton, and Buck. 1985. “Relationships between the Multiplication of Chronic Bee-Paralysis Virus and Its Associate Particle.” *Journal of General Virology* 66(7):1423–29.
Baqir et al. 2012. “Infectious Diseases in the Aftermath of Monsoon Flooding in Pakistan.” *Asian Pacific Journal of Tropical Biomedicine* 2(1):76–79.
Barrangou et al. 2007. “CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes.” *Science* 315(5819):1709–12.
Barroso. 2013. *COMMISSION IMPLEMENTING REGULATION (EU) No 485/2013*.
Bedhomme, Hillung, and Elena. 2015. “Emerging Viruses: Why They Are Not Jacks of All Trades?” *Current Opinion in Virology* 10:1–6.
Bedhomme, Lafforgue, and Elena. 2012. “Multihost Experimental Evolution of a Plant RNA Virus Reveals Local Adaptation and Host-Specific Mutations.” *Molecular Biology and Evolution* 29(5):1481–92.
Beekman and Stratum. 2000. “Does the Diapause Experience of Bumblebee Queens Bombus Terrestris Affect Colony Characteristics?” *Ecological Entomology* 25(1):1–6.
Behnke, Gilbert, Abu-Madi and Lewis. 2005. “Do the Helminth Parasites of Wood Mice Interact?” *Journal of Animal Ecology* 74(5):982–93.
Benaets et al. 2017. “Covert Deformed Wing Virus Infections Have Long-Term Deleterious Effects on Honeybee Foraging and Survival.” *Proceedings of the Royal Society B: Biological Sciences* 284(1848):20162149.
Bertsch, Schweer, Titze and Tanaka. 2005. “Male Labial Gland Secretions and Mitochondrial DNA Markers Support Species Status of Bombus Cryptarum and B. Magnus (Hymenoptera, Apidae).” *Insectes Sociaux* 52(1):45–54.
Biddle, Linde and Godfree. 2012. “Co-Infection Patterns and Geographic Distribution of a Complex Pathosystem Targeted by Pathogen-Resistant Plants.” *Ecological Applications* 22(1):35–52.
Blažytė-Čereškienė et al. 2016. “A Three-Year Survey of Honey Bee Viruses in Lithuania.” *Journal of Apicultural Research* 55(2):176–84.
Boccardo, Milne, Luisoni, Lisa and Accotto. 1985. “Three Seedborne Cryptic Viruses Containing Double-Stranded RNA Isolated from White Clover.” *Virology* 147(1):29–40.
Bonning and Miller. 2010. “Dicistroviruses.” *Annual Review of Entomology* 55(1):129–50.
Bordería, Codoñer, and Sanjuán. 2007. “Selection Promotes Organ Compartmentalization in HIV-1: Evidence from Gag and Pol Genes.” *Evolution* 61(2):272–79.
Bouckaert et al. 2014. “BEAST 2: A Software Platform for Bayesian Evolutionary Analysis”. *PLoS Computational Biology* 10(4):e1003537.
Bouckaert and Drummond. 2015. “bModelTest: Bayesian Phylogenetic Site Model Averaging and Model Comparison.” *bioRxiv* 20792.
Bouza and Rodriguez-Créixems. 1999. “Bacteremic Infections in the HIV-Infected Patient and Recurrent Bacteremia.” *Clinical Microbiology and Infection* 5:2S33-2S39.
Bratbak, Heldal, and Egge. 1991. “Termination of Algal Blooms: Viral Mortality of the Marine Coccolithophorid Emiliania Huxleyi.” *Nature* 93:39–48.
Brcak, Svoboda, and Kralik. 1963. “Electron Microscopic Investigation of

Sacbrood of the Honey Bee." *Journal of Insect Pathology* 5(3):385.
Breeze et al. 2014. "Agricultural Policies Exacerbate Honeybee Pollination Service Supply-Demand Mismatches across Europe." *PLoS ONE* 9(1):e82996.
Brennecke et al. 2007. "Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila." *Cell* 128(6):1089–1103.
Bronkhorst et al. 2012. "The DNA Virus Invertebrate Iridescent Virus 6 Is a Target of the Drosophila RNAi Machinery." *Proceedings of the National Academy of Sciences* 109(51):E3604–13.
Brotman et al. 2014. "Interplay between the Temporal Dynamics of the Vaginal Microbiota and Human Papillomavirus Detection." *Journal of Infectious Diseases* 210(11):1723–33.
Brown, Loosli, and Schmid-Hempel. 2000. "Condition-Dependent Expression of Virulence in a Trypanosome Infecting Bumblebees." *Oikos* 91(3):421–27.
Brown, Mark J. F. et al. 2016. "A Horizon Scan of Future Threats and Opportunities for Pollinators and Pollination" edited by G. Benelli. *PeerJ* 4(August):e2249.
Brussaard, Kempers, Kop, Riegman, and Heldal. 1996. "Virus-like Particles in a Summer Bloom of *Emiliania Huxleyi* in the North Sea." *Aquatic Microbial Ecology* 10(1992):105–13.
Bryden, Gill, Mitton, Raine, and Jansen. 2013. "Chronic Sublethal Stress Causes Bee Colony Failure." *Ecology Letters* 16(12):1463–69.
Buchfink, Xie, and Huson. 2015. "Fast and Sensitive Protein Alignment Using DIAMOND." *Nature Methods* 12(1):59–60.
Buckling, Brockhurst, Travisano, and Rainey. 2007. "Experimental Adaptation to High and Low Quality Environments under Different Scales of Temporal Variation." *Journal of Evolutionary Biology* 20(1):296–300.
Burmeister, Lenski, and Meyer. 2016. "Host Coevolution Alters the Adaptive Landscape of a Virus." *Proceedings of the Royal Society B: Biological Sciences* 283(1839):20161528.
Calderone. 2012. "Insect Pollinated Crops, Insect Pollinators and US Agriculture: Trend Analysis of Aggregate Data for the Period 1992-2009." *PLoS ONE* 7(5):e37235.
Cameron, Hines, and Williams. 2007. "A Comprehensive Phylogeny of the Bumble Bees (Bombus)." *Biological Journal of the Linnean Society* 91(1): 161–88.
Carolan et al. 2012. "Colour Patterns Do Not Diagnose Species: Quantitative Evaluation of a DNA Barcoded Cryptic Bumblebee Complex." *PLoS ONE* 7(1).
Carpenter et al. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76(1):1–32.
Carpenter, Hadfield, Bangham, and Jiggins. 2012. "Specific Interactions between Host and Parasite Genotypes Do Not Act as a Constraint on the Evolution of Antiviral Resistance in Drosophila." *Evolution* 66(4):1114–25.
Carreck and Ratnieksi. 2014. "The Dose Makes the Poison: Have 'field Realistic' rates of Exposure of Bees to Neonicotinoid Insecticides Been Overestimated in Laboratory Studies?" *Journal of Apicultural Research* 53(5):607–14.
Cervera, Lalić, and Elena. 2016. "Effect of Host Species on Topography of the Fitness Landscape for a Plant RNA Irus." *Journal of Virology* 90(22): 10160–69.

Cervera, Lalić, and Elena. 2016. “Efficient Escape from Local Optima in a Highly Rugged Fitness Landscape by Evolving RNA Virus Populations.” *Proceedings of the Royal Society B-Biological Sciences* 283(1836).
Chandra. 1983. “Nutrition, Immunity, and Infection: Present Knowledge and Future Directions.” *The Lancet* 321(8326):688–91.
Cheetham. 1993. “Adaptation on Fluctuating Fitness Landscapes: Speciation and the Persistence of Lineages.” *Journal of Theoretical Biology* 161(3): 287–97.
Chejanovsky et al. 2014. “Characterization of Viral siRNA Populations in Honey Bee Colony Collapse Disorder.” *Virology* 454–455(1):176–83.
Chen et al. 2004. “Multiple Virus Infections in the Honey Bee and Genome Divergence of Honey Bee Viruses.” *Journal of Invertebrate Pathology* 87(2–3):84–93.
Chesson. 1976. “A Non-Central Multivariate Hypergeometric Distribution Arising from Biased Sampling with Application to Selective Predation.” *Journal of Applied Probability* 13(4):795–97.
Choe et al. 2012. “Prevalence and Distribution of Six Bee Viruses in Korean Apis Cerana Populations.” *Journal of Invertebrate Pathology* 109(3):330–33.
Civitello et al. 2015. “Biodiversity Inhibits Parasites: Broad Evidence for the Dilution Effect.” *Proceedings of the National Academy of Sciences* 112(28): 8667–71.
Clark, Wells, Dimitrov, and Clegg. 2016. “Co-Infections and Environmental Conditions Drive the Distributions of Blood Parasites in Wild Birds”. *Journal of Animal Ecology* 85(6):1461–70.
Clark. 1978. “A Filamentous Virus of the Honey Bee.” *Journal of Invertebrate Pathology* 32(3):332–40.
Clark. 1982. “Entomopoxvirus-like Particles in Three Species of Bumblebees.” *Journal of Invertebrate Pathology* 39(1):119–22.
Cobey. 2014. “Pathogen Evolution and the Immunological Niche.” *Annals of the New York Academy of Sciences* 1320(1):1–15.
Cordes et al. 2012. “Interspecific Geographic Distribution and Variation of the Pathogens Nosema Bombi and Crithidia Species in United States Bumble Bee Populations.” *Journal of Invertebrate Pathology* 109(2):209–16.
Cornman et al. 2012. “Pathogen Webs in Collapsing Honey Bee Colonies.” *PLoS ONE* 7(8):1–21.
Posada and Crandall. 2001. “Selecting the Best-Fit Model of Nucleotide Substitution.” *Systematic Biology* 50(4):580–601.
Cresswell, Robert, Florance, and Smirnoff. 2014. “Clearance of Ingested Neonicotinoid Pesticide (Imidacloprid) in Honey Bees (Apis Mellifera) and Bumblebees (Bombus Terrestris).” *Pest Management Science* 70(2):332–37.
Crill, Wichman, and Bull. 2000. “Evolutionary Reversals during Viral Adaptation to Alternating Hosts.” *Genetics* 154(1):27–37.
Cuevas et al. 2012. “Phylogeography and Molecular Evolution of Potato Virus Y.” *PLoS ONE* 7(5):e37853.
Cui and Holmes. 2012. “Endogenous RNA Viruses of Plants in Insect Genomes.” *Virology* 427(2):77–79.
Curtis. 2014. “Infection-Avoidance Behaviour in Humans and Other Animals.” *Trends in Immunology* 35(10):457–64.
Dainat, Evans, Chen, Gauthier and Neumanna. 2012. “Dead or Alive: Deformed

Wing Virus and Varroa Destructor Reduce the Life Span of Winter Honeybees." *Applied and Environmental Microbiology* 78(4):981–87.
Darvill, Knight and Goulson. 2004. "Use of Genetic Markers to Quantify Bumblebee Foraging Range and Nest Density." *Oikos* 107(3):471–78.
DePristo et al. 2011. "A Framework for Variation Discovery and Genotyping Using next-Generation DNA Sequencing Data." *Nature Genetics* 43(5): 491–98.
Devleesschauwer et al. 2014. "Prevalence: Tools for Prevalence Assessment Studies."
Dimmic, Rest, Mindell, and Goldstein. 2002. "rtREV: An Amino Acid Substitution Matrix for Inference of Retrovirus and Reverse Transcriptase Phylogeny." *Journal of Molecular Evolution* 55(1):65–73.
Ding and Voinnet. 2007. "Antiviral Immunity Directed by Small RNAs." *Cell* 130(3):413–26.
Doublet, Labarussias, de Miranda, Moritz, and Paxton. 2015. "Bees under Stress: Sublethal Doses of a Neonicotinoid Pesticide and Pathogens Interact to Elevate Honey Bee Mortality across the Life Cycle." *Environmental Microbiology* 17(4):969–83.
Drummond, Ho, Phillips, and Rambaut. 2006. "Relaxed Phylogenetics and Dating with Confidence" edited by D. Penny. *PLoS Biology* 4(5):699–710.
Drummond, Suchard, Xie, and Rambaut. 2012. "Bayesian Phylogenetics with BEAUti and the BEAST 1.7." *Molecular Biology and Evolution* 29(8):1969–73.
Duffy, Turner, and Burch. 2006. "Pleiotropic Costs of Niche Expansion in the RNA Bacteriophage Φ6." *Genetics* 172(2):751–57.
Duggal and Emerman. 2012. "Evolutionary Conflicts between Viruses and Restriction Factors Shape Immunity." *Nature Reviews Immunology* 12(10): 687–95.
Durrer and Schmid-Hempel. 1994. "Shared Use of Flowers Leads to Horizontal Pathogen Transmission." *Proceedings of the Royal Society B: Biological Sciences* 258(1353):299–302.
Ebert, Brlansky, and Rogers. 2010. "Reexamining the Pooled Sampling Approach for Estimating Prevalence of Infected Insect Vectors." *Annals of the Entomological Society of America* 103(6):827–37.
Ehlenfeldt and Martin. 2009. "Seed Set, Fruit Weight, and Yield in Highbush (Vaccinium Corymbosum L.) Blueberry Cultivars 'Duke' and 'Bluecrop'" edited by D. Steinke. *Acta Horticulturae* 810(8):93–96.
Erler and Lattorff. 2010. "The Degree of Parasitism of the Bumblebee (Bombus Terrestris) by Cuckoo Bumblebees (Bombus (Psithyrus) Vestalis)." *Insectes Sociaux* 57(4):371–77.
Evans. 2001. "Genetic Evidence for Coinfection of Honey Bees by Acute Bee Paralysis and Kashmir Bee Viruses." *Journal of Invertebrate Pathology* 78(4):189–93.
Fauquet and Stanley. 2005. "Revising the Way We Conceive and Name Viruses below the Species Level: A Review of Geminivirus Taxonomy Calls for New Standardized Isolate Descriptors." *Arch Virol* 150:2151–79.
Félix et al. 2011. "Natural and Experimental Infection of Caenorhabditis Nematodes by Novel Viruses Related to Nodaviruses" edited by J. Hodgkin. *PLoS Biology* 9(1):e1000586.
Fenton, Knowles, Petchey, and Pedersen. 2014. "The Reliability of Observational Approaches for Detecting Interspecific Parasite Interactions:

Comparison with Experimental Results." *International Journal for Parasitology* 44(7):437–45.
Fenton and Pedersen. 2005. "Community Epidemiology Framework for Classifying Disease Threats." *Emerging Infectious Diseases* 11(12):1815–21.
Fenton, Streicker, Petchey, and Pedersen. 2015. "Are All Hosts Created Equal? Partitioning Host Species Contributions to Parasite Persistence in Multihost Communities." *The American Naturalist* 186(5):610–22.
Fenton, Viney, and Lello. 2010. "Detecting Interspecific Macroparasite Interactions from Ecological Data: Patterns and Process." *Ecology Letters* 13(5):606–15.
Fick and Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology*.
Finn et al. 2014. "Pfam: The Protein Families Database." *Nucleic Acids Research* 42(D1):D222–30.
Flajnik and Kasahara. 2009. "Origin and Evolution of the Adaptive Immune System: Genetic Events and Selective Pressures." *Nature Reviews Genetics* 11.
Fleming. 1871. *Animal Plagues: Their History, Nature and Prevention*. London: Chapman and Hall.
Fog. 2015. "BiasedUrn: Biased Urn Model Distributions."
França et al. 2009. "Impact of Malnutrition on Immunity and Infection." *Journal of Venomous Animals and Toxins Including Tropical Diseases* 15(3):374–90.
Franks. 1987. "The Parasitic Strategies of a Cuckoo Bee." *Trends in Ecology and Evolution* 2(11):324–26.
Free. 1955. "The Division of Labour within Bumblebee Colonies." *Insectes Sociaux* 2(3):195–212.
Fry. 1996. "The Evolution of Host Specialisation: Are Trade-Offs Overrated?" *The American Naturalist* 148(148):S84–107.
Fujiyuki et al. 2004. "Novel Insect Picorna-Like Virus Identified in the Brains of Aggressive Worker Honeybees." *Journal of Virology* 78(3):1093–1100.
Fujiyuki et al. 2006. "Prevalence and Phylogeny of Kakugo Virus, a Novel Insect Picorna-like Virus That Infects the Honeybee (Apis Mellifera L.), under Various Colony Conditions." *Journal of Virology* 80(23):11528–38.
Fürst, McMahon, Osborne, Paxton, and Brown. 2014. "Disease Associations between Honeybees and Bumblebees as a Threat to Wild Pollinators." *Nature* 506(7488):364–66.
le Gac, Hood, Fournier, and Giraud. 2007. "Phylogenetic Evidence of Host-Specific Cryptic Species in the Anther Smut Fungus." *Evolution* 61(1):15–26.
Gagolewski. 2017. "R Package Stringi: Character String Processing Facilities."
Gajger, Jolanta, Bakonyi and Nowotny. 2014. "Prevalence and Distribution Patterns of Seven Different Honeybee Viruses in Diseased Colonies: A Case Study from Croatia." *Apidologie* 45(6):701–6.
Gandon. 2004. "Evolution of Multihost Parasites." *Evolution* 58(3):455–69.
Genersch et al. 2010. "The German Bee Monitoring Project: A Long Term Study to Understand Periodically High Winter Losses of Honey Bee Colonies." *Apidologie* 41(3):332–52.
Genersch, Yue, Fries, and de Miranda. 2006. "Detection of Deformed Wing Virus, a Honey Bee Viral Pathogen, in Bumble Bees (Bombus Terrestris

and Bombus Pascuorum) with Wing Deformities." *Journal of Invertebrate Pathology* 91(1):61–63.
Gibbs and Gower. 1960. "The Use of a Multiple-Transfer Method in Plant Virus Transmission Studies - Some Statistical Points Arising in the Analysis of Results." *Annals of Applied Biology* 48(1):75–83.
Gill, Ramos-Rodriguez and Raine. 2012. "Combined Pesticide Exposure Severely Affects Individual- and Colony-Level Traits in Bees." *Nature* 491(7422):105–8.
Goka, Okabe, and Yoneda. 2006. "Worldwide Migration of Parasitic Mites as a Result of Bumblebee Commercialization." *Population Ecology* 48(4):285–91.
Goodwin. 1995. "Seasonal Phenology and Abundance of Early-, Mid-and Long-Season Bumble Bees in Southern England." *Journal of Apicultural Research* 34(2):79–87.
Goulson. 2010. *Bumblebees: Behaviour, Ecology and Conservation*. 2nd Edition. Oxford: Oxford University Press.
Goulson and Williams. 2001. "Bombus Hypnorum (L.)(Hymenoptera: Apidae), a New British Bumblebee?" *British Journal of Entomology and …* (1973):1–3.
Goulson. 2013. "An Overview of the Environmental Risks Posed by Neonicotinoid Insecticides" edited by D. Kleijn. *Journal of Applied Ecology* 50(4):977–87.
Goulson and Darvill. 2004. "Niche Overlap and Diet Breadth in Bumblebees; Are Rare Species More Specialized in Their Choice of Flowers?" *Apidologie* 35(1):55–63.
Goulson, Lye, and Darvill. 2008. "Diet Breadth, Coexistence and Rarity in Bumblebees." *Biodiversity and Conservation* 17(13):3269–88.
Grabherr et al. 2011. "Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome." *Nature Biotechnology* 29(7):644–52.
Graystock, Goulson, and Hughes. 2015. "Parasites in Bloom: Flowers Aid Dispersal and Transmission of Pollinator Parasites within and between Bee Species." *Proceedings of the Royal Society B-Biological Sciences* 282(1813):20151371.
Graystock, Meeus, Smagghe, Goulson, and Hughes. 2016. "The Effects of Single and Mixed Infections of Apicystis Bombi and Deformed Wing Virus in Bombus Terrestris." *Parasitology* 143:358–65.
Griffiths, Pedersen, Fenton, and Petchey. 2011. "The Nature and Consequences of Coinfection in Humans." *Journal of Infection* 63(3):200–206.
Gupta. 1998. "Chaos, Persistence, and Evolution of Strain Structure in Antigenically Diverse Infectious Agents." *Science* 280(5365):912–15.
Hadfield. 2010. "MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package." *Journal of Statistics Software* 33(2): 1–22.
Hadfield, Krasnov, Poulin, and Nakagawa. 2014. "A Tale of Two Phylogenies: Comparative Analyses of Ecological Interactions." *The American Naturalist* 183(2):174–87.
Hafner and Nadler. 1990. "Cospeciation in Host-Parasite Assemblages: Comparative Analysis of Rates of Evolution and Timing of Cospeciation Events." *Systematic Zoology* 39(3):192–204.
Hajek and van Nouhuys. 2016. "Fatal Diseases and Parasitoids: From Competition to Facilitation in a Shared Host." *Proceedings of the Royal*

Society of London B: Biological Sciences 283(1828).
Haldane and Jayakar. 1963. “Polymorphism due to Selection of Varying Direction.” *Journal of Genetics* 58(2):237–42.
Hamilton. 1972. “Altruism and Related Phenomena, Mainly in Social Insects.” *Annual Review of Ecology and Systematics* 3(1):193–232.
Hapuarachchi et al. 2016. “Intra-Epidemic Evolutionary Dynamics of a Dengue Virus Type 1 Population Reveal Mutant Spectra That Correlate with Disease Transmission.” *Scientific Reports* 6(1):22592.
Harder. 1985. “Morphology as a Predictor of Flower Choice by Bumble Bees.” *Ecology* 66(1):198–210.
Harmon-Threatt and Ackerly. 2013. “Filtering across Spatial Scales: Phylogeny, Biogeography and Community Structure in Bumble Bees.” *PLoS ONE* 8(3):e60446.
Hedtke, Patiny and Danforth. 2013. “The Bee Tree of Life: A Supermatrix Approach to Apoid Phylogeny and Biogeography.” *BMC Evolutionary Biology* 13:138.
Henikoff and Henikoff. 1992. “Amino Acid Substitution Matrices from Protein Blocks.” *Proceedings of the National Academy of Sciences of the United States of America* 89(22):10915–19.
Henry et al. 2012. “A Common Pesticide Decreases Foraging Success and Survival in Honey Bees.” *Science* 336(6079):348–50.
Heredia,, Trubenová, Sudholt, and Paixão. 2017. “Selection Limits to Adaptive Walks on Correlated Landscapes.” *Genetics* 205(2):803–25.
Hines. 2008. “Historical Biogeography, Divergence Times, and Diversification Patterns of Bumble Bees (Hymenoptera: Apidae: Bombus).” *Systematic Biology* 57(1):58–75.
Holmstrup et al. 2010. “Interactions between Effects of Environmental Chemicals and Natural Stressors : A Review.” *Science of the Total Environment* 408(18):3746–62.
Horn. 1966. “Measurement of ‘Overlap’ in Comparative Ecological Studies.” *The American Naturalist* 100(914):419–24.
Hueffer et al. 2003. “The Natural Host Range Shift and Subsequent Evolution of Canine Parvovirus Resulted from Virus-Specific Binding to the Canine Transferrin Receptor The Natural Host Range Shift and Subsequent Evolution of Canine Parvovirus Resulted from Virus-Specific Bind.” *Journal of Virology* 77(3):1718–26.
Ilonen, Salonen, Marusyk, and Salmi. 1988. “Measles Virus Strain-Dependent Variation in Outcome of Infection of Human Blood Mononuclear Cells.” *Journal of General Virology* 69(1):247–52.
Inouye. 1978. “Resource Partitioning in Bumblebees: Experimental Studies of Foraging Behavior.” *Ecology* 59(4):672–78.
Johnson et al. 2013. “Phylogenomics Resolves Evolutionary Relationships among Ants, Bees, and Wasps.” *Current Biology* 23(20):2058–62.
Johnson, Lafferty, van Oosterhout, and Cable. 2011. “Parasite Transmission in Social Interacting Hosts: Monogenean Epidemics in Guppies.” *PLoS ONE* 6(8).
Jolles, Ezenwa, Etienne, Turner, and Olff. 2008. “Interactions between Macroparasites and Microparasites Drive Infection Patterns in Free-Ranging African Buffalo.” *Ecology* 89(8):2239–50.
Joseph, Togawa, and Shindo. 2013. “Bacterial and Viral Infections Associated with Influenza.” *Influenza and Other Respiratory Viruses* 7(SUPPL.2):105–

13.
Joshi and Fass. 2011. “Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files (Version 1.33).” *Available at Https:// github.com/najoshi/sickle*.
Kame and Argos. 1984. “Primary Structural Comparison of RNA-Dependent Polymerases from Plant, Animal and Bacterial Viruses.” *Nucleic Acids Research* 12(18):7269–82.
Kanehisa, Goto, Kawashima, and Nakaya. 2002. “The KEGG Databases at GenomeNet.” *Nucleic Acids Research* 30(1):42–46.
Kassen. 2002. “The Experimental Evolution of Specialists, Generalists, and the Maintenance of Diversity.” *Journal of Evolutionary Biology* 15(2):173–90.
Katoh,, Misawa, Kuma, and Miyata. 2002. “MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform.” *Nucleic Acids Research* 30(14):3059–66.
Katoh and Standley. 2013. “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.” *Molecular Biology and Evolution* 30(4):772–80.
Kauffman and Levin. 1987. “Towards a General Theory of Adaptive Walks on Rugged Landscapes.” *Journal of Theoretical Biology* 128(1):11–45.
Kawecki. 1994. “Accumulation of Deleterious Mutations and the Evolutionary Cost Being a Generalist.” *The American Naturalist* 144(5):833–38.
Keane, Creevey, Pentony, Naughton, and Mclnerney. 2006. “Assessment of Methods for Amino Acid Matrix Selection and Their Use on Empirical Data Shows That Ad Hoc Assumptions for Choice of Matrix Are Not Justified.” *BMC Evolutionary Biology* 6(1):29.
Kearse, Moir, and Cheung. 2012. “Geneious Pro.” 6.0.6:1647–49.
Keeling, Bjørnstad, and Grenfell. 2004. “Metapopulation Dynamics of Infectious Diseases.” Pp. 415–45 in *Ecology, Genetics and Evolution of Metapopulations*.
Kessler et al. 2015. “Bees Prefer Foods Containing Neonicotinoid Pesticides.” *Nature* 521(7550):74–76.
Klaus and Strimmer. 2015. “Fdrtool: Estimation of (Local) False Discovery Rates and Higher Criticism.”
Klein et al. 2007. “Importance of Pollinators in Changing Landscapes for World Crops.” *Proceedings of the Royal Society B: Biological Sciences* 274(1608):303–13.
Koonin, Dolja, and Krupovic. 2015. “Origins and Evolution of Viruses of Eukaryotes: The Ultimate Modularity.” *Virology* 479–480:2–25.
Koonin, Wolf, Nagasaki, and Dolja. 2008. “The Big Bang of Picorna-like Virus Evolution Antedates the Radiation of Eukaryotic Supergroups.” *Nature Reviews Microbiology* 6(12):925–39.
Kosior et al. 2007. “The Decline of the Bumble Bees and Cuckoo Bees (Hymenoptera: Apidae: Bombini) of Western and Central Europe.” *Oryx* 41(1):79.
Kotob, Menanteau-Ledouble, Kumar, Abdelzaher and El-Matbouli. 2016. “The Impact of Co-Infections on Fish: A Review.” *Veterinary Research* 47(1):1–12.
Kwan and Ernst. 2011. “HIV and Tuberculosis: A Deadly Human Syndemic.” *Clinical Microbiology Reviews* 24(2):351–76.
Langmead and Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9(4):357–59.

Laycock, Lenthall, Barratt, and Cresswell. 2012. “Effects of Imidacloprid, a Neonicotinoid Pesticide, on Reproduction in Worker Bumble Bees (Bombus Terrestris).” *Ecotoxicology* 21(7):1937–45.
Lee et al. 2014. “MOSAIK: A Hash-Based Algorithm for Accurate next-Generation Sequencing Short-Read Mapping.” *PLoS ONE* 9(3).
Leigh and Bryant. 2015. “POPART: Full-Feature Software for Haplotype Network Construction” edited by S. Nakagawa. *Methods in Ecology and Evolution* 6(9):1110–16.
Lemoine et al. 2016. “Underappreciated Problems of Low Replication in Ecological Feld Studies.” *Ecology* 97(10):2554–61.
Levitt et al. 2013. “Cross-Species Transmission of Honey Bee Viruses in Associated Arthropods.” *Virus Research* 176(1–2):232–40.
Li et al. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25(16):2078–79.
Li et al. 2014. “Systemic Spread and Propagation of a Plant-Pathogenic Virus in European Honeybees, Apis Mellifera.” *mBio* 5(1).
Liu, Liang, Etienne, Gilbert, and Yu. 2016. “Phylogenetic Congruence between Subtropical Trees and Their Associated Fungi.” *Ecology and Evolution* 6(23):8412–22.
Liu, Gelman, and Zheng. 2015. “Simulation-Efficient Shortest Probability Intervals.” *Statistics and Computing* 25(4):809–19.
Longdon et al. 2015. “The Evolution, Diversity, and Host Associations of Rhabdoviruses.” *Virus Evolution* 1(1):vev014-vev014.
Longdon, Brockhurst, Russell, Welch, and Jiggins. 2014. “The Evolution and Genetics of Virus Host Shifts.” *PLoS Pathogens* 10(11):e1004395.
Longdon, Hadfield, Webster, Obbard, and Jiggins. 2011. “Host Phylogeny Determines Viral Persistence and Replication in Novel Hosts.” *PLoS Pathogens* 7(9):e1002260.
Lunau and Maier. 1995. “Innate Colour Preferences of Flower Visitors.” *Journal of Comparative Physiology A* 177(1):1–19.
Lye, Kaden, Park, and Goulson. 2010. “Forage Use and Niche Partitioning by Non-Native Bumblebees in New Zealand: Implications for the Conservation of Their Populations of Origin.” *Journal of Insect Conservation* 14(6):607–15.
Mackenzie and Jeggo. 2013. “Reservoirs and Vectors of Emerging Viruses.” *Current Opinion in Virology* 3(2):170–79.
Magoč and Salzberg. 2011. “FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies.” *Bioinformatics* 27(21):2957–63.
Maldarelli et al. 2013. “HIV Populations Are Large and Accumulate High Genetic Diversity in a Nonlinear Fashion.” *Journal of Virology* 87(18):10313–23.
Manley, Boots and Wilfert. 2015. “Emerging Viral Disease Risk to Pollinating Insects: Ecological, Evolutionary and Anthropogenic Factors.” *Journal of Applied Ecology* 52(2):331–40.
Manley, Boots, and Wilfert. 2017. “Condition-Dependent Virulence of Slow Bee Paralysis Virus in Bombus Terrestris: Are the Impacts of Honeybee Viruses in Wild Pollinators Underestimated?” *Oecologia* 184(2):305–15.
Maori et al. 2007. “Isolation and Characterization of Israeli Acute Paralysis Virus, a Dicistrovirus Affecting Honeybees in Israel: Evidence for Diversity due to Intra- and Inter-Species Recombination.” *Journal of General Virology* 88(12):3428–38.
Martínez-Aquino et al. 2009. “Detecting a Complex of Cryptic Species within

Neoechinorhynchus Golvani (Acanthocephala: Neoechinorhynchidae) Inferred from ITSs and LSU rDNA Gene Sequences." *Journal of Parasitology* 95(5):1040–47.
Mayera, Mantegani and Frei. 1999. "Does Hepatitis B Virus (HBV) Genotype Influence the Clinical Outcome of HBV Infection?" *Journal of Viral Hepatitis* 6(4):299–304.
Mazzei et al. 2014. "Infectivity of DWV Associated to Flower Pollen: Experimental Evidence of a Horizontal Transmission Route." *PLoS ONE* 9(11):e113448.
Mbithim, Springthorpe and Sattar. 1991. "Effect of Relative Humidity and Air Temperature on Survival of Hepatitis A Virus on Environmental Surfaces." *Applied and Environmental Microbiology* 57(5):1394–99.
McArt, Koch, Irwin and Adler. 2014. "Arranging the Bouquet of Disease: Floral Traits and the Transmission of Plant and Animal Pathogens" edited by J. Gurevitch. *Ecology Letters* 17(5):624–36.
McMahon et al. 2015. "A Sting in the Spit: Widespread Cross-Infection of Multiple RNA Viruses across Wild and Managed Bees." *Journal of Animal Ecology* 84(3):615–24.
Meyer, Skinhøj and Prag. 1994. "Bacteremia in HIV-Positive and AIDS Patients: Incidence, Species Distribution, Risk-Factors, Outcome, and Influence of Long-Term Prophylactic Antibiotic Treatment." *Scandinavian Journal of Infectious Diseases* 26(6):635–42.
Michener and Amir. 1977. "The Seasonal Cycle and Habitat of a Tropical Bumble Bee." *Pacific Insects* 17(2/3):237–40.
de Miranda and Fries. 2008. "Venereal and Vertical Transmission of Deformed Wing Virus in Honeybees (Apis Mellifera L.)." *Journal of Invertebrate Pathology* 98(2):184–89.
de Miranda et al. 2015. "Genome Characterization, Prevalence and Distribution of a Macula-like Virus from Apis Mellifera and Varroa Destructor." *Viruses* 7(7):3586–3602.
de Miranda and Genersch. 2010. "Deformed Wing Virus." *Journal of Invertebrate Pathology* 103(SUPPL. 1):S48-61.
Möckel, Gisder and Genersch. 2011. "Horizontal Transmission of Deformed Wing Virus: Pathological Consequences in Adult Bees (Apis Mellifera) Depend on the Transmission Route." *Journal of General Virology* 92(2): 370–77.
Møller. 1995. "Bumblebee Preference for Symmetrical Flowers." *Proceedings of the National Academy of Sciences of the United States of America* 92(6): 2288–92.
Moore et al. 2011. "Recombinants between Deformed Wing Virus and Varroa Destructor Virus-1 May Prevail in Varroa Destructor-Infested Honeybee Colonies." *Journal of General Virology* 92(1):156–61.
Mordecai et al. 2016. "Moku Virus; a New Iflavirus Found in Wasps, Honey Bees and Varroa." *Scientific Reports* 6:34983.
Mordeca, Wilfert, Martin, Jones and Schroeder. 2015. "Diversity in a Honey Bee Pathogen: First Report of a Third Master Variant of the Deformed Wing Virus Quasispecies." *The ISME Journal* 10(5):1–10.
Moret and Schmid-Hempel. 2001. "Immune Defence in Bumble-Bee Offspring." *Nature* 414(6863):506.
Moret and Schmid-Hempel. 2000. "Survival for Immunity : The Price of Immune System Activation for Bumblebee Workers." *Science* 290(November):1166–

69.
Mouret et al. 2013. “Prevalence of 12 Infectious Agents in Field Colonies of 18 Apiaries in Western France.” *Revue de Médecine Vétérinaire* 164(12):577–82.
van Munster et al. 2002. “Sequence Analysis and Genomic Organization of Aphid Lethal Paralysis Virus: A New Member of the Family Dicistroviridae.” *Journal of General Virology* 83(Pt 12):3131–8.
Murray, Fitzpatrick, Brown and Paxton. 2008. “Cryptic Species Diversity in a Widespread Bumble Bee Complex Revealed Using Mitochondrial DNA RFLPs.” *Conservation Genetics* 9(3):653–66.
Nagata et al. 2006. “Infectivity Variation and Genetic Diversity among Strains of Western Equine Encephalitis Virus.” *The Journal of General Virology* 87(2006):2353–61.
Natsopoulou et al. 2017. “The Virulent, Emerging Genotype B of Deformed Wing Virus Is Closely Linked to Overwinter Honeybee Worker Loss.” *Scientific Reports* 7(1):5242.
Nelsen. 2013. *An Introduction to Copulas*. 2nd ed. Springer.
Nielsen, Nicolaisen and Kryger. 2008. “Incidence of Acute Bee Paralysis Virus, Black Queen Cell Virus, Chronic Bee Paralysis Virus, Deformed Wing Virus, Kashmir Bee Virus and Sacbrood Virus in Honey Bees (Apis Mellifera) in Denmark.” *Apidologie* 39:310–14.
Oers. 2010. “Genomics and Biology of Lflaviruses.” Pp. 231–50 in *Insect Virology*, edited by S. Asgari and K. N. Johnson. Norfolk, UK: Caister Academic Press.
Ohta. 1992. “The Nearly Neutral Theory Of Molecular Evolution.” *Annual Review of Ecology and Systematics* 23(1992):263–86.
Oksanen et al. 2017. “Vegan: Community Ecology Package.” *R Package Version 2.4-4*.
Olivier et al. 2008. “Molecular Characterisation and Phylogenetic Analysis of Chronic Bee Paralysis Virus, a Honey Bee Virus.” *Virus Research* 132(1–2):59–68.
Ongus et al. 2004. “Complete Sequence of a Picorna-like Virus of the Genus Iflavirus Replicating in the Mite Varroa Destructor.” *Journal of General Virology* 85(12):3747–55.
Osborne et al. 1999. “A Landscape-Scale Study of Bumble Bee Foraging Range and Constancy, Using Harmonic Radar.” *Journal of Applied Ecology* 36(4): 519–33.
Osborne et al. 2008. “Bumblebee Flight Distances in Relation to the Forage Landscape.” *Journal of Animal Ecology* 77(2):406–15.
Otterstatter and Thomson. 2007. “Contact Networks and Transmission of an Intestinal Pathogen in Bumble Bee (Bombus Impatiens) Colonies.” *Oecologia* 154(2):411–21.
Park, Blitzer, Gibbs, Losey and Danforth. 2015. “Negative Effects of Pesticides on Wild Bee Communities Can Be Buffered by Landscape Context.” *Proceedings of the Royal Society B: Biological Sciences* 282(1809): 20150299–20150299.
Parke, Rambaut and Pybus. 2008. “Correlating Viral Phenotypes with Phylogeny: Accounting for Phylogenetic Uncertainty.” *Infection, Genetics and Evolution* 8(3):239–46.
Parmentier, Smagghe, de Graaf and Meeus. 2016. “Varroa Destructor Macula-like Virus, Lake Sinai Virus and Other New RNA Viruses in Wild Bumblebee

Hosts (Bombus Pascuorum, Bombus Lapidarius and Bombus Pratorum).” *Journal of Invertebrate Pathology* 134:6–11.
Peat and Goulson. 2005. “Effects of Experience and Weather on Foraging Rate and Pollen versus Nectar Collection in the Bumblebee, Bombus Terrestris.” *Behavioral Ecology and Sociobiology* 58(2):152–56.
Pedersen, Saeed, Michaelsen, Friis and Murrell. 2002. “Impact of Protein Energy Malnutrition on Trichuris Suis Infection in Pigs Concomitantly Infected with Ascaris Suum.” *Parasitology* 124(Pt 5):561–68.
Pei, Kim and Grishin. 2008. “PROMALS3D: A Tool for Multiple Protein Sequence and Structure Alignments.” *Nucleic Acids Research* 36(7):2295–2300.
Peng et al. 2011. “Host Range Expansion of Honey Bee Black Queen Cell Virus in the Bumble Bee, Bombus Huntii.” *Apidologie* 42(5):650–58.
Peng, Fang, Xu and Ge. 1987. “The Resistance Mechanism of the Asian Honey Bee, Apis Cerana Fabr., to an Ectoparasitic Mite, Varroa Jacobsoni Oudemans.” *Journal of Invertebrate Pathology* 49(1):54–60.
Perlman and Jaenike. 2003. “Infection Success in Novel Hosts: An Experimental and Phylogenetic Study of Drosophila-Parasitic Nematodes.” *Evolution* 57(3):544–57.
Pettis, vanEngelsdorp, Johnson and Dively. 2012. “Pesticide Exposure in Honey Bees Results in Increased Levels of the Gut Pathogen Nosema.” *Naturwissenschaften* 99(2):153–58.
Pham, Dionne,Shirasu-Hiza and Schneider. 2007. “A Specific Primed Immune Response in Drosophila Is Dependent on Phagocytes.” *PLoS Pathogens* 3(3):e26.
Plath. 1922. “Notes on Psithyrus, With Records of Two New American Hosts.” *The Biological Bulletin* 43(1):23–44.
Plischuk, Antúnez, Haramboure, Minardi and Lange. 2017. “Long-Term Prevalence of the Protists Crithidia Bombi and Apicystis Bombi and Detection of the Microsporidium Nosema Bombi in Invasive Bumble Bees.” *Environmental Microbiology Reports* 9(2):169–73.
Popp, Erler and Lattorff. 2012. “Seasonal Variability of Prevalence and Occurrence of Multiple Infections Shape the Population Structure of Crithidia Bombi , an Intestinal Parasite of Bumblebees (Bombus Spp.).” *MicrobiologyOpen* 1(4):362–72.
di Prisco et al. 2013. “Neonicotinoid Clothianidin Adversely Affects Insect Immunity and Promotes Replication of a Viral Pathogen in Honey Bees.” *Proceedings of the National Academy of Sciences* 110(46):18466–71.
R Core Development Team. 2016. “R: A Language and Environment for Statistical Computing.”
Rambaut, Suchard, Xie and Drummond. 2014. “Tracer v1.6.”
Rao, Poinar and Henley. 2017. “A Scientific Note on Rare Parasitism of the Bumble Bee Pollinator, Bombus Impatiens, by a Mermithid Nematode, Pheromermis Sp. (Nematoda: Mermithidae).” *Apidologie* 48(1):75–77.
Ravoet et al. 2013. “Comprehensive Bee Pathogen Screening in Belgium Reveals Crithidia Mellificae as a New Contributory Factor to Winter Mortality.” *PLoS ONE* 8(8).
Remnant et al. 2017. “A Diverse Range of Novel RNA Viruses in Geographically Distinct Honey Bee Populations.” *Journal of Virology* (May):JVI.00158-17.
Remold, Rambaut and Turner. 2008. “Evolutionary Genomics of Host Adaptation in Vesicular Stomatitis Virus.” *Molecular Biology and Evolution*

25(6):1138–47.
Renton, Whitaker, Ison, Wadsworth and Harris. 1995. “Estimating the Sexual Mixing Patterns in the General Population from Those in People Acquiring Gonorrhoea Infection: Theoretical Foundation and Empirical Findings.” *Journal of Epidemiology and Community Health* 49(2):205–13.
Retschnig, Neumann and Williams. 2014. “Thiacloprid-Nosema Ceranae Interactions in Honey Bees: Host Survivorship but Not Parasite Reproduction Is Dependent on Pesticide Dose.” *Journal of Invertebrate Pathology* 118(0):18–19.
van Rij et al. 2006. “The RNA Silencing Endonuclease Argonaute 2 Mediates Specific Antiviral Immunity in Drosophila Melanogaster.” *Genes and Development* 20(21):2985–95.
van Riper, van Riper, Goff and Laird. 1986. “The Epizootiology and Ecological Significance of Malaria in Hawaiian Land Birds.” *Ecological Monographs* 56(4):327–44.
Roberts. 1988. “The Structure of Particles of Tobacco Ringspot Nepovirus: Evidence from Electron Microscopy.” *Journal of General Virology* 69(8): 1831–40.
Roberts, Anderson and Durr. 2017. “Absence of Deformed Wing Virus and Varroa Destructor in Australia Provides Unique Perspectives on Honeybee Viral Landscapes and Colony Losses.” *Scientific Reports* 7(1):6925.
Ruiz-González et al. 2012. “Dynamic Transmission, Host Quality, and Population Structure in a Multihost Parasite of Bumblebees.” *Evolution* 66(10):3053–66.
Runckel et al. 2011. “Temporal Analysis of the Honey Bee Microbiome Reveals Four Novel Viruses and Seasonal Prevalence of Known Viruses, Nosema, and Crithidia.” *PLoS ONE* 6(6):e20656.
Ryabov et al. 2014. “A Virulent Strain of Deformed Wing Virus (DWV) of Honeybees (Apis Mellifera) Prevails after Varroa Destructor-Mediated, or In Vitro, Transmission.” *PLoS Pathogens* 10(6):e1004230.
Sadd and Schmid-Hempel. 2006. “Insect Immunity Shows Specificity in Protection upon Secondary Pathogen Exposure.” *Current Biology* 16(12): 1206–10.
Sanjuan, Nebot, Chirico, Mansky and Belshaw. 2010. “Viral Mutation Rates.” *Journal of Virology* 84(19):9733–48.
Santos, Almeida, Gortázar and Correia-Neves. 2015. “Patterns of Mycobacterium Tuberculosis-Complex Excretion and Characterization of Super-Shedders in Naturally-Infected Wild Boar and Red Deer.” *Veterinary Research* 46(1):129.
Schaible and Kaufmann. 2007. “Malnutrition and Infection: Complex Mechanisms and Global Impacts.” *PLoS Medicine* 4(5):0806–12.
Schmid-Hempel. 1998. *Parasites in Social Insects*. Princeton University Press.
Schoonvaere et al. 2016. “Unbiased RNA Shotgun Metagenomics in Social and Solitary Wild Bees Detects Associations with Eukaryote Parasites and New Viruses”. *PLOS ONE* 11(12):e0168456.
Seabloom, Hosseini, Power and Borer. 2009. “Diversity and Composition of Viral Communities: Coinfection of Barley and Cereal Yellow Dwarf Viruses in California Grasslands.” *The American Naturalist* 173(3):E79–98.
Sehgal et al. 2006. “Evidence for cryptic speciation of leucocytozoon spp. (Haemosporidia, Leucocytozoidae) in diurnal raptors.” *Journal of Parasitology* 92(2):375–79.

Sekirov et al. 2008. “Antibiotic-Induced Perturbations of the Intestinal Microbiota Alter Host Susceptibility to Enteric Infection.” *Infection and Immunity* 76(10):4726–36.
Sela, Ashkenazy, Katoh and Pupko. 2015. “GUIDANCE2: Accurate Detection of Unreliable Alignment Regions Accounting for the Uncertainty of Multiple Parameters.” *Nucleic Acids Research* 43(W1):W7–14.
Seo, Thorne, Hasegawa and Kishino. 2002. “Estimation of Effective Population Size of HIV-1 within a Host: A Pseudomaximum-Likelihood Approach.” *Genetics* 160(4):1283–93.
Sharma, Duveiller and Ortiz-Ferrara. 2007. “Progress and Challenge towards Reducing Wheat Spot Blotch Threat in the Eastern Gangetic Plains of South Asia: Is Climate Change Already Taking Its Toll?” *Field Crops Research* 103(2):109–18.
Shi et al. 2016. “Redefining the Invertebrate RNA Virosphere.” *Nature* 540:1–12.
Shinya et al. 2006. “Influenza Virus Receptors in the Human Airway.” *Nature* 440(7083):435–36.
Shykoff and Schmid-Hempel. 1991. “Incidence and Effects of Four Parasites in Natural Populations of Bumble Bees in Switzerland.” *Apidologie* 22(2):117–25.
Singh et al. 2010. “RNA Viruses in Hymenopteran Pollinators: Evidence of Inter-Taxa Virus Transmission via Pollen and Potential Impact on Non-Apis Hymenopteran Species.” *PLoS ONE* 5(12):e14357.
Sladen. 1912. *The Humble-Bee, Its Life History and How to Domesticate It*. London: Macmillan and Co. Lmt.
Smith and Markham. 1944. “Two New Viruses Affecting Tobacco and Other Plants.” *Phytopathology* 34:324–29.
Stan Development Team. 2016a. “RStan: The R Interface to Stan.”
Stan Development Team. 2016b. “Rstanarm: Bayesian Applied Regression Modeling via Stan.”
Stern et al. 2017. “The Evolutionary Pathway to Virulence of an RNA Virus.” *Cell* 169(1):35–46.e19.
Streicker, Fenton and Pedersen. 2013. “Differential Sources of Host Species Heterogeneity Influence the Transmission and Control of Multihost Parasites” edited by R. Ostfeld. *Ecology Letters* 16(8):1134–37.
Suttle and Chen. 1992. “Mechanisms and Rates of Decay of Marine Viruses in Seawater.” *Applied and Environmental Microbiology* 58(11):3721–29.
Suttle. 2007. “Marine Viruses--Major Players in the Global Ecosystem.” *Nature Reviews Microbiology* 5(10):801–12.
Suzuki et al. 2017. “Uncovering the Repertoire of Endogenous Flaviviral Elements in Aedes.” *Journal of Virology*.
Svensson. 1979. “Pyrobombus Lapponicus Auct., in Europe Recognized as Two Species: P. Lapponicus (Fabricius, 1793) and P. Monticola (Smith, 1849) (Hymenoptera, Apoidea, Bombinae).” *Insect Systematics & Evolution* 10(4TY–JOUR):275–96.
Telfer et al. 2010. “Species Interactions in a Parasite Community Drive Infection Risk in a Wildlife Population.” *Science* 330(6001):243–46.
Thomson. 1981. “Field Measures of Flower Constancy in Bumblebees.” *American Midland Naturalist* 105(2):377–80.
Thu et al. 2016. “Prevalence of Bee Viruses among Apis Cerana Populations in Vietnam.” *Journal of Apicultural Research* 55(5):379–85.
Tsvetkov et al. 2017. “Chronic Exposure to Neonicotinoids Reduces Honey Bee

Health near Corn Crops." *Science* 356(6345).
Turner and Elena. 2000. "Cost of Host Radiation in an RNA Virus." *Genetics* 156(4):1465–70.
Untergasser et al. 2012. "Primer3-New Capabilities and Interfaces." *Nucleic Acids Research* 40(15):e115.
Václav, Ficová, Prokop, and Betáková. 2011. "Associations Between Coinfection Prevalence of Borrelia Lusitaniae, Anaplasma Sp., and Rickettsia Sp. in Hard Ticks Feeding on Reptile Hosts." *Microbial Ecology* 61(2):245–53.
Vale, Choisy, Froissart, Sanjuán and Gandon. 2012. "The Distribution Of Mutational Fitness Effects Of Phage ΦX174 On Different Hosts." *Evolution* 66(11):3495–3507.
Vanbergen and the Insect Pollinators Initiative. 2013. "Threats to an Ecosystem Service: Pressures on Pollinators." *Frontiers in Ecology and the Environment* 11(5):251–59.
Venables and Ripley. 2002. *Modern Applied Statistics with S*. New York: Springer.
Vidau et al. 2011. "Exposure to Sublethal Doses of Fipronil and Thiacloprid Highly Increases Mortality of Honeybees Previously Infected by Nosema Ceranae" edited by E. Didier. *PLoS ONE* 6(6):e21550.
de Vienne, Hood and Giraud. 2009. "Phylogenetic Determinants of Potential Host Shifts in Fungal Pathogens." *Journal of Evolutionary Biology* 22(12): 2532–41.
de Visse and Krug. 2014. "Empirical Fitness Landscapes and the Predictability of Evolution." *Nature Reviews Genetics* 15(7):480–90.
Walther-Hellwig and Frankl. 2000. "Foraging Distances of Bombus Muscorum, Bombus Lapidarius, and Bombus Terrestris (Hymenoptera, Apidae)." *Journal of Insect Behavior* 13(2):239–46.
Wang et al. 2013. "Sequence Recombination and Conservation of Varroa Destructor Virus-1 and Deformed Wing Virus in Field Collected Honey Bees (Apis Mellifera)." *PLoS ONE* 8(9).
Wang and Jehle. 2009. "Nudiviruses and Other Large, Double-Stranded Circular DNA Viruses If Ubvertebrates: New Insights on an Old Topic." *Journal of Invertebrate Pathology* 101(3):187–93.
Warnes, Bolker and Lumley. 2015. "Gtools: Various R Programming Tools."
Waxman, Weinert and Welch. 2014. "Inferring Host Range Dynamics from Comparative Data: The Protozoan Parasites of New World Monkeys." *The American Naturalist* 184(1):65–74.
Webster et al. 2015. "The Discovery, Distribution, and Evolution of Viruses Associated with Drosophila Melanogaster." *PLoS Biology* 13(7):e1002210.
Westphal, Steffan-Dewenter and Tscharntke. 2006. "Bumblebees Experience Landscapes at Different Spatial Scales: Possible Implications for Coexistence." *Oecologia* 149(2):289–300.
Whitehorn, O'Connor, Wackers and Goulson. 2012. "Neonicotinoid Pesticide Reduces Bumble Bee Colony Growth and Queen Production." *Science* 336(6079):351–52.
Whitehorn, Tinsley, Brown, Darvill and Goulson. 2011. "Genetic Diversity, Parasite Prevalence and Immunity in Wild Bumblebees." *Proceedings. Biological Sciences / The Royal Society* 278(1709):1195–1202.
Wickham. 2011. "The Split-Apply-Combine Strategy for Data Analysis." *Journal of Statistical Software* 40(1).

Wiegmann et al. 2009. “Single-Copy Nuclear Genes Resolve the Phylogeny of the Holometabolous Insects.” *BMC Biology* 7(1):34.
Wilfert et al. 2016. “Deformed Wing Virus Is a Recent Global Epidemic in Honeybees Driven by Varroa Mites.” *Science* 351(6273):594–97.
Williams et al. 2012. “Unveiling Cryptic Species of the Bumblebee Subgenus Bombus S. Str. Worldwide with COI Barcodes (Hymenoptera: Apidae).” *Systematics and Biodiversity* 10(1):21–56.
Williams,. Araújo and Rasmont. 2007. “Can Vulnerability among British Bumblebee (Bombus) Species Be Explained by Niche Position and Breadth?” *Biological Conservation* 138(3–4):493–505.
Williams, Cameron, Hines, Cederberg and Rasmont. 2008. “A Simplified Subgeneric Classification of the Bumblebees (Genus Bombus).” *Apidologie* 39(1):1–29.
Williams and Osborne. 2009. “Bumblebee Vulnerability and Conservation World-Wide.” *Apidologie* 40(3):367–87.
Wilm et al. 2012. “LoFreq: A Sequence-Quality Aware, Ultra-Sensitive Variant Caller for Uncovering Cell-Population Heterogeneity from High-Throughput Sequencing Datasets.” *Nucleic Acids Research* 40(22):11189–201.
Withenshaw, Devevey, Pedersen and Fenton. 2016. “Multihost Bartonella Parasites Display Covert Host Specificity Even When Transmitted by Generalist Vectors.” *Journal of Animal Ecology* 85(6):1442–52.
Wolf and Moritz. 2008. “Foraging Distance in Bombus Terrestris L. (Hymenoptera: Apidae).” *Apidologie* 39(4):419–27.
Wolf, Ellington and Begley. 1999. “Foraging Costs in Bumblebees: Field Conditions Cause Large Individual Differences.” *Insectes Sociaux* 46(3): 291–95.
Wommack, Nasko, Chopyk and Sakowski. 2015. “Counts and Sequences, Observations That Continue to Change Our Understanding of Viruses in Nature.” *Journal of Microbiology* 53(3):181–92.
Woolhouse and Gowtage-Sequeria. 2005. “Host Range and Emerging and Reemerging Pathogens.” *Emerging Infectious Diseases* 11(12):1842–47.
Worden and Papaj. 2005. “Flower Choice Copying in Bumblebees.” *Biology Letters* 1(4):504–7.
Yañez et al. 2012. “Deformed Wing Virus and Drone Mating Flights in the Honey Bee (Apis Mellifera): Implications for Sexual Transmission of a Major Honey Bee Virus.” *Apidologie* 43(1):17–30.
Yang, Peng, Li and Kadowaki. 2013. “Molecular and Phylogenetic Characterization of Honey Bee Viruses, Nosema Microsporidia, Protozoan Parasites, and Parasitic Mites in China.” *Ecology and Evolution* 3(2):298–311.
Yang and Marr. 2012. “Mechanisms by Which Ambient Humidity May Affect Viruses in Aerosols.” *Applied and Environmental Microbiology* 78(19): 6781–88.
Yarwood. 1956. “Generation Time and the Biological Nature of Viruses.” *The American Naturalist* 90(851):97–102.
Yue, Schroder, Gisder and Genersch. 2007. “Vertical-Transmission Routes for Deformed Wing Virus of Honeybees (Apis Mellifera).” *Journal of General Virology* 88(8):2329–36.
Zhang et al. 2012. “New Evidence That Deformed Wing Virus and Black Queen Cell Virus Are Multi-Host Pathogens.” *Journal of Invertebrate Pathology* 109(1):156–59.

Zheng et al. 2015. “Evidence of the Synergistic Interaction of Honey Bee Pathogens Nosema Ceranae and Deformed Wing Virus.” *Veterinary Microbiology* 177(1–2):1–6.

Zioni, Soroker and Chejanovsky. 2011. “Replication of Varroa Destructor Virus 1 (VDV-1) and a Varroa Destructor Virus 1 – Deformed Wing Virus Recombinant (VDV-1–DWV) in the Head of the Honey Bee.” *Virology* 1(1): 1–7.

Zwart and Elena. 2015. “Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution.” *Annual Review of Virology* 2(1): 161–79.