Predictors of virus prevalence and diversity across a wild bumblebee community

David J. Pascall[1,2], Matthew C. Tinsley[3], Bethany L. Clark[4,5], Darren J. Obbard[6] and Lena Wilfert[2,7]

1. Institute of Biodiversity, Animal Health and Comparative Medicine, Boyd Orr Centre for Population and Ecosystem Health, University of Glasgow, Glasgow, United Kingdom
2. College of Life and Environmental Science, University of Exeter, Penryn Campus, Cornwall, United Kingdom
3. Biological and Environmental Sciences, Stirling University, Stirling, United Kingdom
4. Environment and Sustainability Institute, University of Exeter, Penryn Campus, Cornwall, United Kingdom
5. BirdLife International, The David Attenborough Building, Pembroke Street, Cambridge, United Kingdom
6. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom
7. Institute of Evolutionary Ecology and Conservation Genomics, University of Ulm, Ulm, Germany

Abstract

Viruses are a key regulator of natural populations. Despite this, we have limited knowledge of the diversity and ecology of viruses that lack obvious fitness effects on their host. This is even the case in wild host populations that provide ecosystem services, where small fitness effects may have major ecological and financial impacts in aggregate. One such group of hosts are the bumblebees, which have a major role in the pollination of food crops and have suffered population declines and range contractions in recent decades. In this study, we used a multivariate generalised linear mixed model to investigate the ecological factors that determine the prevalence of four recently discovered bumblebee viruses (Mayfield virus 1, Mayfield virus 2, River Liunaeg virus and Loch Morlich virus), and two previously known viruses that infect both wild bumblebees and managed honeybees (Acute bee paralysis virus and Slow bee paralysis virus). We show that the recently discovered bumblebee viruses were more genetically diverse than the viruses shared with honeybees, potentially due to spillover dynamics of shared viruses. We found evidence for ecological drivers of prevalence in our samples, with relatively weak evidence for a positive effect of precipitation on the prevalence of River Luinaeg virus. Coinfection is potentially important in shaping

prevalence: we found a strong positive association between River Liunaeg virus and Loch Morlich virus presence after controlling for host species, location and other relevant ecological variables. This study represents a first step in the description of predictors of bumblebee infection in the wild not driven by spillover from honeybees.

Introduction

Viruses are among the most abundant and diverse groups of organisms on Earth (Wommack et al., 2015); wherever they are looked for, they are found in other species as obligate pathogens. Despite this, viral ecology in natural populations remains understudied. In the wild, infection is generally only recorded when clear symptoms of the underlying disease are present, such as discolouration, aberrant tissue structures, or an increase in mortality. However, these symptoms are rarely detected in natural infections (Mackenzie & Jeggo, 2013). By focusing only on those viruses that cause obvious symptoms in well-studied host species, we are likely to be underestimating the diversity of viruses and their ecological importance in regulating natural populations. For example, the dynamics of algal blooms are strongly driven by density dependent regulation of the algae through viral infections (Bratbak et al., 1991; Brussaard et al., 1996). The development of relatively cheap and easily applied molecular techniques has allowed the detection and identification of potentially pathogenic organisms within both the host and the environment, enabling the systematic study of viral ecology in wild populations (e.g. (Webster et al., 2015)). This is especially important for threatened host species, where understanding the viral burden may have conservation implications (Gordon et al., 2015).

Pollinators are economically important and so their viruses are important in turn. Over 50 viruses have now been described in bees, and their importance to survival is well recognised (e.g. (McMahon et al., 2018)). However, the majority of this work has been performed in the European honeybee, *Apis mellifera*, thus the knowledge of the viral ecology of bumblebees is more limited. Some honeybee work is transferable to bumblebees; for instance, viruses known from honeybees have pathogenic effects in the buff-tailed bumblebee, *Bombus terrestris* (Fürst et al., 2014; 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). However, few predictors of infection in bumblebees other than the presence of sympatric honeybees have been described in any depth.

In the wild, differences in viral 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 & Schmid-Hempel, 1994; Graystock et al., 2015; McArt et al., 2014) and so factors that reduce contamination of floral structures may be predicted to reduce the rate of infection in the general bumblebee population (Adler et al., 2020); obvious candidates are UV exposure, rainfall, temperature and humidity. High UV levels may deactivate virus particles rapidly (Lytle & Sagripanti, 2005), an effect thought to be highly important in the regulation of viral populations in oceanic waters (Suttle & Chen, 1992). Rainfall may physically clean the flowers. Heat can increase the rate of viral deactivation, both independently and through an interaction with relative humidity (Mbithi et al., 1991). Furthermore, bees must physically reach the flowers where infection can occur, so factors that change the rate of contact of workers with heavily contaminated flowers may also modify viral prevalence. Wind speed affects the relative rates of pollen and nectar collection (Peat & Goulson, 2005), which may alter flower visitation and the energetic costs of foraging (T. J. Wolf et al., 1999), consequently affecting susceptibility to infection. However, environmental conditions would only be expected to lead to interspecific prevalence differences locally through species-specific effects on be behaviour.

Another factor that can explain infection in populations is the presence of other infectious agents. Co-occurrences of particular pathogens have been observed in many species and can drive infection dynamics with certain combinations being over- or under-represented relative to chance expectations (Johnson et al., 2015; Tollenaere et al., 2016). This may be due to synergistic or antagonistic interactions between the pathogens. For example, mechanical damage to tissues by a primary infection allowing easier access to secondary pathogens (Joseph et al., 2013) or due to competitive exclusion of pathogens with the same niche within the host (Amaku et al., 2013). Very few studies in pollinators have looked for these between pathogen interactions in a statistically rigorous manner.

Given the complexity of the pollinator system, it is also unclear how much genetic diversity will be present in the viral populations. As some viruses such as the re-emerging Deformed wing virus represent spillovers from honeybee populations into wild pollinator populations (Fürst et al., 2014; Manley et al., 2019; Wilfert et al., 2016), the genetic diversity in wild pollinators for these viruses will likely represent a potentially non-random sample of viral diversity in its maintenance host. For viruses that are maintained in bumblebees, however,

the diversity would be expected to be impacted by the normal evolutionary patterns of mutation, selection and drift. Depending on the population size and the historical selection regime, this could lead to very high diversity, or almost none if a major selective event or bottleneck had recently occurred.

Here we present an exploratory investigation into the determinants of viral prevalence and genetic diversity in wild bumblebee populations consisting of 13 species from nine sites across Scotland. We explore the effect of differences in UV radiation, precipitation, temperature and wind speed on the prevalences of four recently discovered bumblebee viruses Mayfield virus 1 (MV1), Mayfield virus 2 (MV2), River Liunaeg virus (RLV), and Loch Morlich virus (LMV), where fitness effects have not yet been tested. We also explore whether or not there are observable statistical interactions between the pathogens that might indicate facilitative or suppressive interactions. We consider the genetic diversity in these viruses and contrast this with two viruses known from honeybees, Acute bee paralysis virus (ABPV) and Slow bee paralysis virus (SBPV). We show that the viruses only described in bumblebees are universally more diverse than ABPV and SBPV and that there is a strong positive association between LMV and RLV infection.

Methods

Sampling Regime and Molecular Work

Samples were derived from the field collections described in Pascall et al. (2018). Briefly, we collected a total of 759 bumblebees of 13 species from 9 sites across Scotland, UK (Supp Table 1; Fig. 1). The Ochil Hills, Glenmore, Dalwhinnie, Stirling, Iona, Staffa and the Pentlands were sampled in 2009, while Edinburgh and Gorebridge were sampled in 2011. 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.

In this study, we assayed the prevalence of Mayfield virus 1 (MV1), Mayfield virus 2 (MV2), River Luinaeg virus (RLV), Loch Morlich virus (LMV) at the individual level by RTPCR (Supp Table 2). We tested a subset of the samples for Slow bee paralysis virus (SBPV) (n = 544) and Acute bee paralysis virus (ABPV) (n = 385).

Community Similarity

To estimate host community similarity between sampling sites, we estimated the underlying sampling probability of each bumblebee species at each site by treating the observed samples as being drawn from a multinomial distribution with 24 categories, corresponding to the 24 bumblebee species in the United Kingdom. We use a Dirichlet prior with these 24 categories and a concentration parameter of 1 for each category, implying complete uncertainty about the underlying probability. This has the advantage that the posterior has a known analytical form. 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 bumblebee species at each site, which we assume to be roughly equivalent to the frequency of that bumblebee species at that site. For each of the 10,000 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.

Prevalence and Climatic Association

Climatic data for each of the nine sites at which bees were collected was taken from the WorldClim database at 1km resolution (Fick & Hijmans, 2017). Predictions for July and August derived from data from 1960-2010 were extracted for mean daily maximum temperature, mean precipitation, mean solar radiation and mean wind speed at the grid reference for the sites with a buffer area of 2km to account for the fact that bumblebees forage over approximately that distance (Wolf & Moritz, 2008). All values were averaged to generate a consensus value for that site and then mean-centred and scaled to unit standard deviation. At this point, we tested the correlation between the variables. Mean solar radiation and mean daily maximum temperature were highly correlated (Pearson correlation: 0.78), so only mean maximum temperature was carried forward. All remaining variables had low Pearson correlations between them in the range of -0.4 to 0.4.

We tested associations between individual prevalence and climate data using Stan version 2.18.2 (Carpenter et al., 2017) via the rstan interface (Stan Development Team, 2016) in R version 3.6 (R Core Development Team, 2016). A multivariate probit model was fitted, with random host, location and host-location effects, and mean maximum temperature, precipitation, and wind speed as fixed effects for each virus. The usage of the multivariate probit allows us to test for excess co-infection between viruses in the study, i.e. coinfection

beyond random expectations after the effect of shared covariates has been removed. As the number of sampling locations was small, we expected our ability to accurately determine the size and direction of effects caused by ecological covariates would be limited. To reduce the effect of drawing spurious conclusions due to our small number of sites, we applied regularisation as recommended by Lemoine et al. (2016), using a regularising prior distribution. The global intercept for each virus was given a Gaussian (mu=0, sigma=10) prior, which does not substantially penalise low probabilities. Each fixed effect coefficient was given a Gaussian (mu=0, sigma=1) prior, which, given that the fixed effects act at the site level, should dominate the likelihood if the effect is small. Random effects were drawn from normal distributions centred at 0 with estimated standard deviations. In all cases, the standard deviations were given Exponential (lambda=2) hyperpriors, which are only weakly informative on the logit scale when the data is informative for the standard deviation. The correlation in residuals for the multivariate normal was given a near flat prior using a Lewandoski, Kurowicka and Joe (eta=1) prior. While the Stan code used can regularly give outputs with divergent transitions, the presented model had no divergent transitions over 24000 samples, tail and bulk effective sample sizes of over 400 for all parameters and no Bayesian fraction of missing information warning. We did not perform model selection given our regularising priors, and statements are made based on estimates from the full model.

Diversity Analysis

To analyse sequence diversity, we used the raw reads from the RNA sequencing described in Pascall et al. (2018). 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. MV1, MV2, RLV, LMV, SBPV Rothamsted (EU035616.1) and ABPV (AF486072.2) sequences were aligned on the TranslatorX server (Abascal et al., 2010), using its MAFFT setting (Katoh & Standley, 2013). Post-alignment, we manually trimmed sequences to the conserved region of the RdRp gene, minus eight codons, owing to the shortness of the RLV 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, 1536, 1501, 1519 and 1522 base pairs for MV1, MV2, RLV, LMV, SBPV and ABPV respectively. Raw bioinformatic reads were trimmed in sickle version 1.33 using the default parameters (Joshi & Fass, 2011). Overlapping mate reads were combined by FLASH version 1.2.11 using the default settings (Magoč & Salzberg, 2011). Reads were aligned to the RdRp sequences generated above

using MOSAIK version 2.1.73 (Lee et al., 2014). Both merged reads 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, MV1 and MV2 duplicate sequences were not marked, as per best practice (McKenna et al., 2010). Variants were called using the default settings in LoFreq* version 1.2.1 (Wilm et al., 2012). Base quality scores were 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 (a total of four recalibrations). Once the score distribution stabilised, variant calling was performed 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. If the median depth over called differences from the consensus was less than 20, species-virus combinations were removed from the variant analysis. *B. lucorum* was analysed for SBPV, ABPV and MV1. *B. terrestris* was analysed for SBPV, MV1 and MV2. *B. pascuorum* was analysed for ABPV, SBPV and MV2. The mixed *Bombus* pool was analysed for all six viruses.

The number of polymorphic sites was calculated for each virus. Variants with allele frequencies of 1 were removed as these represent fixed differences from the underlying reference sequence. To convert these counts to approximate Watterson's theta (Watterson, 1975) for each host-virus combination, a measure of the genetic diversity, we had to account for the fact that the make-up of the pools was not precisely the same as the samples that were tested by PCR. We predicted the status of each of the untested individuals in the pools from the model discussed above and took the median and a 90% central credible interval of the number of additional positives (over those confirmed by PCR). We then used these numbers to give bounds on the approximation to Watterson's theta estimator, by looking at the value of the estimator at those three estimates of the number of infected individuals. The equation used was:

$$\theta = \frac{n}{l \sum_1^i \frac{1}{i}}$$

where θ is the approximation to Watterson's theta, n is the number of variants, l is the length of the sequence and i is the number of PCR positives. This method makes two strong assumptions: 1) there is no mixed infection of viral variants in individuals (i.e. that one extra

individual represents a single extra count for the harmonic partial sum in the denominator of Watterson's theta) and 2) all variants present are detectable. The impact of deviation from the first assumption is likely to be small. The marginal change in the partial sum in the denominator decreases with every extra count, so a few missed counts will result in little change to the resultant estimate. The second assumption is more impactful, given the larger impact that a missing variant has on the generated number. Given this, we acknowledge that our presented estimates may be conservative.

Results

Bumblebee Community Similarity

There were obvious differences in bumblebee community structure between our Scottish sampling sites. The locations we sampled in the south had *B. terrestris*, *B. pascuorum* and *B. lucorum* dominated communities, whereas those further north had *Bombus jonellus* and *Bombus hortorum* dominated communities (Figure 1, Table 2). Even with its small sample size of 13 bees, the Pentlands, a range of hills in southern Scotland (Figure 1), appeared to represent a third type of community: 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 the other community types. This is potentially due to its higher elevation and habitat being similar to the north with large numbers of heathland plants while being situated in the south.

Prevalence

There were large differences in prevalence of the viruses (all +ssRNA picorna-like viruses) between sites (Figure 2). 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.

River Luinaeg virus (RLV) was detected 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*. 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 (LMV) appears to exhibit much higher species specificity with 13/16 detections being in *B. jonellus*. It was also strongly associated with RLV, with

13/16 detections being coinfections. No species other than *B. jonellus* was detected with LMV infection in the absence of RLV coinfection. Mayfield virus 1 (MV1) appears to be a true generalist virus, with frequent infections across bumblebee species. Its prevalence data showed large differences between the degree of infection of different bumblebee species between sites. Edinburgh and Gorebridge, two sites around 15km apart with large sample sizes, have dramatically different MV1 prevalences in *B. terrestris*, *B. pratorum* and *B. pascuorum*, being between 30-60% in Edinburgh, and below 15% in all species in Gorebridge. Mayfield virus 2 (MV2) shows a similar pattern but without obvious differences in infection levels between sites. The prevalence of MV2 is generally lower than that of MV1, but beyond that, the range of species infected is largely similar. Acute bee paralysis virus (ABPV) was found at intermediate modal prevalences of above 10% in all species apart from *B. terrestris* and *B. lucorum*. The prevalence of SBPV was universally high.

Factors Influencing Infection

Viral prevalence was related to environmental covariates in some cases (Table 3; Figure 3). Higher levels of precipitation had a high posterior probability of being associated with higher prevalences of River Luinaeg virus (posterior probability: 95%), and there was some evidence that less precipitation was associated with reduced prevalence of Mayfield virus 1 (posterior probability: 92%) and higher maximum temperatures and wind speeds were associated with higher prevalences of Mayfield virus 1 (posterior probability: 94% and 93% respectively). For most covariates, however, the bulk of the posterior distribution lay close to zero and did not shift considerably from the prior indicating a lack of between site resolution.

We also tested for excess co-infection beyond random between viruses using multivariate probit models that allowed us to calculate 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 affecting infection risk. In this case, these measure 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. Some viruses exhibited excess coinfection (Table 4). RLV and LMV showed strong and positive correlation (mean correlation: 0.73), consistent with the high levels of coinfection noted above; the error correlation between these two viruses was the only one where the bulk of the posterior was not close to zero. There was also some indication of a negative association between MV2 and SBPV, with the bulk of the posterior

supporting a correlation of below zero (posterior probability: 93%), but the variance of the posterior was such that this cannot be stated with great certainty.

Diversity

Over homologous genomic regions within the RdRp gene, there were large differences between viruses in our approximation to Watterson's theta (see methods; Table 4). RLV, LMV, MV1, and MV2 all exhibited more diversity than SBPV and ABPV, with SPBV itself being considerably more diverse than ABPV. 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, implying that the variants present in an area are stable over short periods. Additionally, as would be expected, most variation was in 3rd codon positions leading to either no amino-acid replacements or replacements with similarly charged amino acids.

Discussion

In this study, we explored the ecological factors influencing the genetic diversity and distribution of the viruses of wild bumblebees. We found that all the viruses only detected in bumblebees have considerably higher genetic diversity than the viruses shared with honeybees. Additionally, we found evidence of a positive association between River Luinaeg virus and Loch Morlich virus. Finally, there is some evidence for an effect of environmental variables on viral prevalence with the strongest evidence being that higher levels of precipitation increase the prevalence of River Luinaeg virus.

Diversity

Both Acute bee paralysis virus (ABPV) and Slow bee paralysis virus (SBPV) show considerably less diversity than Mayfield virus 1 (MV1), Mayfield virus 2 (MV2), River Luinaeg virus (RLV) and Loch Morlich virus (MLV) within the study region, even though our estimates of Watterson's theta may be conservative due to the strong assumptions on detecting viral variants in individuals. ABPV and SBPV are viruses that were initially described in honeybees (Bailey et al., 1963; Bailey & Gibbs, 1964), while the other four viruses in the study were found in bumblebees and have not been recorded in honeybees at this point (Pascall et al., 2018). Both the Sanger sequences and Watterson's theta calculations show that the diversity in ABPV and SBPV remains low in bumblebees across the short period between 2009 to 2011. This difference could have multiple causes. It is likely that much of the observed variation in the 'bumblebee viruses' LMV, MV1, MV2 and RLV 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 on the virus. Given the frequent bottlenecking that occurs during transmission (Zwart & Elena, 2015), such 3rd codon position mutations would be inefficiently selected against.

The lack of diversity in ABPV and SBPV could be due to them infecting both honeybees and bumblebees. In multihost systems, species can differ in their susceptibility and response to infection (Ruiz-González et al., 2012), thus different species have very different levels of importance for the maintenance of a virus in a population. As such, single heavily infected host species can act as sources for infection in other sympatric species. In honeybees, these viruses interact with the mite *Varroa destructor,* a parasite that can vector viruses and is associated with the prevalence of ABPV (Mondet et al., 2014) and SBPV (Manley et al., 2020). This vector can lead to reduced viral genetic diversity in honeybees, as shown for Deformed wing virus (Martin et al., 2012), resulting in a limited pool of virus able to spillover into other hosts, which could explain the reduction in variation we observed in the viruses known to infect honeybees relative to those only described from bumblebees.

Bumblebee-limited viruses necessarily undergo multiple independent bottlenecking events when the viruses can only survive in overwintering queens, potentially maintaining diversity through reduced selection efficiency. This would initially seem to apply to SBPV, as McMahon et al. (2015) and Manley et al. (2020) report that SBPV is often found at higher prevalences in bumblebees than in sympatric honeybees during in summer. However, over winter, bumblebee populations are reduced to individual queens while honeybee colonies are maintained at population sizes in the thousands. This can maintain a level of virus in the honeybees that can then spillover into the bumblebees in the spring, which may lead to honeybee-derived SBPV variants dominating even in bumblebees. This could represent a more general effect where the fitness landscapes of viruses infecting managed species are systematically different from those infecting wild species. The hypothesised mechanism is that in a genetically homogenous, densely-packed managed population, one optimal viral genotype could easily achieve dominance as the fitness landscape will be relatively constant. On the other hand, in more genetically heterogeneous wild populations with fluctuating population sizes, the fitness landscape will be less constant, and thus selection of variants on this changing fitness landscape may lead to the maintenance of more genotypes.

Factors Influencing Infection

We found evidence that the prevalence of River Luinaeg virus was positively associated with increased precipitation, though our ability to estimate the precise size of this effect was limited. 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 floral structures 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 increases the severity of bumblebee virus infections (Manley et al., 2017) and has been shown to increase infection risk in mammals (reviewed in (França et al., 2009; Pedersen et al., 2002; Schaible & Kaufmann, 2007)). River Luinaeg virus is predominantly associated with *Bombus jonellus*, a bee species found at significantly higher frequencies in our sampling locations with higher precipitation, while the inclusion of our host random effect should account for this somewhat, it is possible that this is also driving the effect. We also found weak evidence of a negative effect of precipitation on MV1. Given that only nine sites were sampled in this study, we are limited in the between-site conclusions we can draw. Despite using a regularising prior, the risk of erroneously identifying effects can never be fully excluded in studies with small sample sizes. Thus, given the importance of wild pollinators, a study with more sites would be valuable future work to more accurately assess the effect of the environment on viral infection risk.

Coinfection

River Luinaeg virus and Loch Morlich virus were rarely found separately in this study. They are distinct, not different segments of the same virus, as, while whole genomes are available for neither, both partial genomes include an RdRp sequence (Pascall et al., 2018). The mechanism of their transmission is unknown, but we assume that, as with most other reported bee viruses, transmission occurs at flowers.

One potential explanation for this strong association 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 et al., 1980). However, this seems unlikely as both virus species are observed separately, though the possibility of false negatives in the PCR reactions cannot be ruled out. 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 in a manner analogous to HIV's synergism with TB though immune suppression (Kwan & Ernst, 2011) or influenza virus' changing of the environment of the

nasopharynx, allowing secondary bacterial invasion (Joseph et al., 2013). Viral coinfections are ubiquitously reported in prevalence studies in bees (Anderson & 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; Manley et al., 2020; McMahon et al., 2015; Mouret et al., 2013; Nielsen et al., 2008; Roberts et al., 2017; Thu et al., 2016), but to our knowledge, only McMahon et al. (2015) and Manley et al. (Manley et al., 2020) tested for a departure from random expectations of infection, and no departure was found. However, non-random associations between parasites appear common, having been reported in, among other taxa including mammals (Behnke et al., 2005; Griffiths et al., 2011; Jolles et al., 2008), birds (Clark et al., 2016), arthropods (Hajek & van Nouhuys, 2016; Václav et al., 2011) and plants (Biddle et al., 2012; Seabloom et al., 2009). Thus, while the cause is uncertain, the strength of this association makes it highly unlikely to be artefactual.

Conclusion

Here we describe the ecological and genetic characteristics of a set of viruses of bumblebees. Of the six viral species studied, genetic diversity is higher in the viruses we only detected in bumblebees. This could be explained if viruses in species managed for food production, such as honeybees, are less diverse than those in wild species, and outbreaks of these viruses in wild species are predominantly due to spillover. Further studies in this and other systems would be valuable to answer the question of whether there is a significant difference in diversity in viruses in managed species and those shared between managed and wild species versus those limited to wild species. Viral genetic diversity could be a factor in determining the risk of both disease emergence and spillover.

Acknowledgements

We thank Dave Goulson, Claire Webster, Floh Bayer, Steph O'Connor, Penelope Whitehorn, Gillian Lye and Mario Vallejo-Marin for assistance with bumblebee collection and species identification. Additionally, we thank Angus Buckling and Andy Fenton for constructive suggestions for improvements throughout.

Data availability statement

All code and raw infection data is available at https://github.com/dpascall/bumble_epi_diversity. All raw sequence data has been deposited on GenBank.

Bibliography

Abascal, F., Zardoya, R., & Telford, M. J. (2010). TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, *38*(Supplement 2), W7–W13. https://doi.org/10.1093/nar/gkq291

Adler, L. S., Barber, N. A., Biller, O. M., & Irwin, R. E. (2020). Flowering plant composition shapes pathogen infection intensity and reproduction in bumble bee colonies. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.2000074117

Amaku, M., Coutinho, F. A. B., Chaib, E., & Massad, E. (2013). The Impact of Hepatitis A Virus Infection on Hepatitis C Virus Infection: A Competitive Exclusion Hypothesis. *Bulletin of Mathematical Biology*, *75*(1), 82–93. https://doi.org/10.1007/s11538-012-9795-0

Anderson, D. L., & Gibbs, A. J. (1988). Inapparent Virus Infections and their Interactions in Pupae of the Honey Bee (Apis mellifera Linnaeus) in Australia. *Journal of General Virology*, *69*(7), 1617–1625. https://doi.org/10.1099/0022-1317-69-7-1617

Bacandritsos, N., Granato, A., Budge, G., Papanastasiou, I., Roinioti, E., Caldon, M., Falcaro, C., Gallina, A., & Mutinelli, F. (2010). Sudden deaths and colony population decline in Greek honey bee colonies. *Journal of Invertebrate Pathology*, *105*(3), 335–340. https://doi.org/10.1016/j.jip.2010.08.004

Bailey, L., Ball, B. V., Carpenter, J. M., & Woods, R. D. (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), 149–155. https://doi.org/10.1099/0022-1317-46-1-149

Bailey, L., & Gibbs, a. J. (1964). Acute infection of bees with Paralysis Virus. *Journal of Insect Pathology*, *6*(4), 395–407.

Bailey, L., Gibbs, A. J., & Woods, R. D. (1963). Two viruses from adult honey bees (Apis mellifera Linnaeus). *Virology*, *21*(3), 390–395. https://doi.org/10.1016/0042-6822(63)90200-9

Behnke, J. M., Gilbert, F. S., Abu-Madi, M. A., & Lewis, J. W. (2005). Do the helminth parasites of wood mice interact? *Journal of Animal Ecology*, *74*(5), 982–993. https://doi.org/10.1111/j.1365-2656.2005.00995.x

Biddle, J. M., Linde, C., & Godfree, R. C. (2012). Co-infection patterns and geographic distribution of a complex pathosystem targeted by pathogen-resistant plants.

Ecological Applications, *22*(1), 35–52. https://doi.org/10.1890/11-0341.1

Blažytė-Čereškienė, L., Skrodenytė-Arbačiauskienė, V., Radžiutė, S., Čepulytė-Rakauskienė, R., Apšegaitė, V., & Būda, V. (2016). A three-year survey of honey bee viruses in Lithuania. *Journal of Apicultural Research*, *55*(2), 176–184. https://doi.org/10.1080/00218839.2016.1211389

Bratbak, G., Heldal, M., & Egge, J. K. (1991). Termination of algal blooms: viral mortality of the marine coccolithophorid Emiliania huxleyi. *Nature - Submitted*, *93*, 39–48.

Brussaard, C., Kempers, R., Kop, A., Riegman, R., & Heldal, M. (1996). Virus-like particles in a summer bloom of *Emiliania huxleyi* in the North Sea. *Aquatic Microbial Ecology*, *10*(1992), 105–113. https://doi.org/10.3354/ame010105

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Chen, Y., Zhao, Y., Hammond, J., Hsu, H. T., Evans, J., & Feldlaufer, M. (2004). Multiple virus infections in the honey bee and genome divergence of honey bee viruses. *Journal of Invertebrate Pathology*, *87*(2–3), 84–93. https://doi.org/10.1016/j.jip.2004.07.005

Choe, S. E., Nguyen, L. T. K., Noh, J. H., Koh, H. B., Jean, Y. H., Kweon, C. H., & Kang, S. W. (2012). Prevalence and distribution of six bee viruses in Korean Apis cerana populations. *Journal of Invertebrate Pathology*, *109*(3), 330–333. https://doi.org/10.1016/j.jip.2012.01.003

Clark, N. J., Wells, K., Dimitrov, D., & Clegg, S. M. (2016). Co-infections and environmental conditions drive the distributions of blood parasites in wild birds. *Journal of Animal Ecology*, *85*(6), 1461–1470. https://doi.org/10.1111/1365-2656.12578

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Durrer, S., & Schmid-Hempel, P. (1994). Shared Use of Flowers Leads to Horizontal Pathogen Transmission. *Proceedings of the Royal Society B: Biological Sciences*, *258*(1353), 299–302. https://doi.org/10.1098/rspb.1994.0176

Evans, J. D. (2001). Genetic Evidence for Coinfection of Honey Bees by Acute Bee Paralysis and Kashmir Bee Viruses. *Journal of Invertebrate Pathology*, *78*(4), 189–193. https://doi.org/10.1006/jipa.2001.5066

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*. https://doi.org/10.1002/joc.5086

França, T., Ishikawa, L., Zorzella-Pezavento, S., Chiuso-Minicucci, F., da Cunha, M., & Sartori, A. (2009). Impact of malnutrition on immunity and infection. *Journal of Venomous Animals and Toxins Including Tropical Diseases*, *15*(3), 374–390. https://doi.org/10.1590/S1678-91992009000300003

Fürst, M. a., McMahon, D. P., Osborne, J. L., Paxton, R. J., & Brown, M. J. F. (2014). Disease associations between honeybees and bumblebees as a threat to wild pollinators. *Nature*, *506*(7488), 364–366. https://doi.org/10.1038/nature12977

Gajger, I. T., Kolodziejek, J., Bakonyi, T., & Nowotny, N. (2014). Prevalence and distribution patterns of seven different honeybee viruses in diseased colonies: a case study from Croatia. *Apidologie*, *45*(6), 701–706. https://doi.org/10.1007/s13592-014-0287-0

Gordon, C. H., Banyard, A. C., Hussein, A., Laurenson, M. K., Malcolm, J. R., Marino, J., Regassa, F., Stewart, A. M. E., Fooks, A. R., & Sillero-Zubiri, C. (2015). Canine distemper in endangered Ethiopian wolves. *Emerging Infectious Diseases*. https://doi.org/10.3201/eid2105.141920

Graystock, P., Goulson, D., & Hughes, W. O. H. (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. https://doi.org/10.1098/rspb.2015.1371

Graystock, P., Meeus, I., Smagghe, G., Goulson, D., & Hughes, W. O. H. (2016). The effects of single and mixed infections of Apicystis bombi and deformed wing virus in Bombus terrestris. *Parasitology*, *143*, 358–365. https://doi.org/10.1017/S0031182015001614

Griffiths, E. C., Pedersen, A. B., Fenton, A., & Petchey, O. L. (2011). The nature and consequences of coinfection in humans. *Journal of Infection*, *63*(3), 200–206. https://doi.org/10.1016/j.jinf.2011.06.005

Hajek, A. E., & van Nouhuys, S. (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). https://doi.org/10.1098/rspb.2016.0154

Horn, H. S. (1966). Measurement of “Overlap” in Comparative Ecological Studies. *The American Naturalist*, *100*(914), 419–424. https://doi.org/10.1086/282436

Johnson, P. T. J., de Roode, J. C., & Fenton, A. (2015). Why infectious disease research needs community ecology. *Science*, *349*, 1259504. https://doi.org/10.1126/science.1259504

Jolles, A. E., Ezenwa, V. O., Etienne, R. S., Turner, W. C., & Olff, H. (2008). Interactions between macroparasites and microparasites drive infection patterns in free-ranging African buffalo. *Ecology*, *89*(8), 2239–2250. https://doi.org/10.1890/07-0995.1

Joseph, C., Togawa, Y., & Shindo, N. (2013). Bacterial and viral infections associated with influenza. *Influenza and Other Respiratory Viruses*, *7*(SUPPL.2), 105–113. https://doi.org/10.1111/irv.12089

Joshi, N., & Fass, J. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). In *Available at https://github.com/najoshi/sickle*.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kwan, C., & Ernst, J. D. (2011). HIV and tuberculosis: A deadly human syndemic. *Clinical Microbiology Reviews*, *24*(2), 351–376. https://doi.org/10.1128/CMR.00042-10

Lee, W. P., Stromberg, M. P., Ward, A., Stewart, C., Garrison, E. P., & Marth, G. T. (2014). MOSAIK: A hash-based algorithm for accurate next-generation sequencing short-read mapping. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0090581

Lemoine, N. P., Hoffman, A., Felton, A. J., Baur, L., Chaves, F., Gray, J., Yu, Q., & Smith, M. D. (2016). Underappreciated problems of low replication in ecological feld studies. *Ecology*, *97*(10), 2554–2561. https://doi.org/10.1002/ecy.1506

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lytle, C. D., & Sagripanti, J.-L. (2005). Predicted Inactivation of Viruses of Relevance to Biodefense by Solar Radiation. *Journal of Virology*. https://doi.org/10.1128/jvi.79.22.14244-14252.2005

Mackenzie, J. S., & Jeggo, M. (2013). Reservoirs and vectors of emerging viruses. *Current Opinion in Virology*, *3*(2), 170–179. https://doi.org/10.1016/j.coviro.2013.02.002

Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

Manley, R., Boots, M., & Wilfert, L. (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), 305–315. https://doi.org/10.1007/s00442-017-3851-2

Manley, R., Temperton, B., Boots, M., & Wilfert, L. (2020). Contrasting impacts of a novel specialist vector on multihost viral pathogen epidemiology in wild and managed bees. *Molecular Ecology*. https://doi.org/10.1111/mec.15333

Manley, R., Temperton, B., Doyle, T., Gates, D., Hedges, S., Boots, M., & Wilfert, L. (2019). Knock-on community impacts of a novel vector: spillover of emerging DWV-B from Varroa-infested honeybees to wild bumblebees. In *Ecology Letters*. https://doi.org/10.1111/ele.13323

Martin, S. J., Highfield, A. C., Brettell, L., Villalobos, E. M., Budge, G. E., Powell, M., Nikaido, S., & Schroeder, D. C. (2012). Global honey bee viral landscape altered by a parasitic mite. *Science*. https://doi.org/10.1126/science.1220941

Mbithi, J. N., Springthorpe, V. S., & Sattar, S. A. (1991). Effect of Relative Humidity and Air Temperature on Survival of Hepatitis A Virus on Environmental Surfaces. *Applied and Environmental Microbiology*, *57*(5), 1394–1399. http://aem.asm.org/content/57/5/1394.full.pdf

McArt, S. H., Koch, H., Irwin, R. E., & Adler, L. S. (2014). Arranging the bouquet of disease: Floral traits and the transmission of plant and animal pathogens. In J. Gurevitch (Ed.), *Ecology Letters* (Vol. 17, Issue 5, pp. 624–636). https://doi.org/10.1111/ele.12257

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. https://doi.org/10.1101/gr.107524.110

McMahon, D. P., Fürst, M. A., Caspar, J., Theodorou, P., Brown, M. J. F., & Paxton, R. J. (2015). A sting in the spit: Widespread cross-infection of multiple RNA viruses across wild and managed bees. *Journal of Animal Ecology*, *84*(3), 615–624. https://doi.org/10.1111/1365-2656.12345

McMahon, D. P., Wilfert, L., Paxton, R. J., & Brown, M. J. F. (2018). Emerging Viruses in Bees: From Molecules to Ecology. In *Advances in Virus Research*. https://doi.org/10.1016/bs.aivir.2018.02.008

Mondet, F., de Miranda, J. R., Kretzschmar, A., Le Conte, Y., & Mercer, A. R. (2014). On the Front Line: Quantitative Virus Dynamics in Honeybee (Apis mellifera L.) Colonies

along a New Expansion Front of the Parasite Varroa destructor. *PLoS Pathogens*, *10*(8), e1004323. https://doi.org/10.1371/journal.ppat.1004323

Mouret, C., Lambert, O., Piroux, M., Beaudeau, F., Provost, B., & Benet, P. (2013). Prevalence of 12 infectious agents in field colonies of 18 apiaries in western France. *Revue de Médecine Vétérinaire*, *164*(12), 577–582. http://www.revmedvet.com/2013/RMV164_577_582.pdf

Nielsen, S. L., Nicolaisen, M., & Kryger, P. (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*, 310–314. https://doi.org/10.1051/apido:2008007

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2017). vegan: Community Ecology Package. In *R package version 2.4-4*. https://doi.org/10.4135/9781412971874.n145

Pascall, D. J., Tinsley, M. C., Obbard, D. J., & Wilfert, L. (2018). Host evolutionary history predicts virus prevalence across bumblebee species. In *bioRxiv*. https://doi.org/10.1101/498717

Peat, J., & Goulson, D. (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), 152–156. https://doi.org/10.1007/s00265-005-0916-8

Pedersen, S., Saeed, I., Michaelsen, K. F., Friis, H., & Murrell, K. D. (2002). Impact of protein energy malnutrition on Trichuris suis infection in pigs concomitantly infected with Ascaris suum. *Parasitology*, *124*(Pt 5), 561–568. https://doi.org/10.1017/S0031182002001592

R Core Development Team. (2016). R: A language and environment for statistical computing. In *Document freely available on the internet at: http://www. r-project. org*. R Foundation for Statistical Computing. https://doi.org/10.1017/CBO9781107415324.004

Roberts, J. M. K., Anderson, D. L., & Durr, P. A. (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. https://doi.org/10.1038/s41598-017-07290-w

Ruiz-González, M. X., Bryden, J., Moret, Y., Reber-Funk, C., Schmid-Hempel, P., & Brown, M. J. F. (2012). Dynamic transmission, host quality, and population structure in a

multihost parasite of bumblebees. *Evolution*, *66*(10), 3053–3066. https://doi.org/10.1111/j.1558-5646.2012.01655.x

Schaible, U. E., & Kaufmann, S. H. E. (2007). Malnutrition and infection: Complex mechanisms and global impacts. *PLoS Medicine*, *4*(5), 0806–0812. https://doi.org/10.1371/journal.pmed.0040115

Seabloom, E. W., Hosseini, P. R., Power, A. G., & Borer, E. T. (2009). Diversity and Composition of Viral Communities: Coinfection of Barley and Cereal Yellow Dwarf Viruses in California Grasslands. *The American Naturalist*, *173*(3), E79–E98. https://doi.org/10.1086/596529

Stan Development Team. (2016). *RStan: the R interface to Stan* (R package version 2.14.1).

Suttle, C. A., & Chen, F. (1992). Mechanisms and rates of decay of marine viruses in seawater. *Applied and Environmental Microbiology*, *58*(11), 3721–3729. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC183166/pdf/aem00052-0289.pdf

Thu, H. T., Thi, N., Lien, K., Linh, M. T., Le, T. H., Thi, N., Thai, P. H., Reddy, K. E., Yoo, M. S., & Kim, Y.-H. (2016). Prevalence of bee viruses among Apis cerana populations in Vietnam. *Journal of Apicultural Research*, *55*(5), 379–385. https://doi.org/10.1080/00218839.2016.1251193

Tollenaere, C., Susi, H., & Laine, A. L. (2016). Evolutionary and Epidemiological Implications of Multiple Infection in Plants. In *Trends in Plant Science*. https://doi.org/10.1016/j.tplants.2015.10.014

Václav, R., Ficová, M., Prokop, P., & Betáková, T. (2011). Associations Between Coinfection Prevalence of Borrelia lusitaniae, Anaplasma sp., and Rickettsia sp. in Hard Ticks Feeding on Reptile Hosts. *Microbial Ecology*, *61*(2), 245–253. https://doi.org/10.1007/s00248-010-9736-0

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*. https://doi.org/10.1016/0040-5809(75)90020-9

Webster, C. L., Waldron, F. M., Robertson, S., Crowson, D., Ferrari, G., Quintana, J. F., Brouqui, J. M., Bayne, E. H., Longdon, B., Buck, A. H., Lazzaro, B. P., Akorli, J., Haddrill, P. R., & Obbard, D. J. (2015). The discovery, distribution, and evolution of viruses associated with drosophila melanogaster. *PLoS Biology*, *13*(7), e1002210. https://doi.org/10.1371/journal.pbio.1002210

Wilfert, L., Long, G., Schmid-Hempel, P., Butlin, R., Martin, S. J. M., & Boots, M. (2016). Deformed wing virus is a recent global epidemic in honeybees driven by Varroa mites.

Science, *351*(6273), 594–597. https://doi.org/10.1126/science.aac9976

Wilm, A., Aw, P. P. K., Bertrand, D., Yeo, G. H. T., Ong, S. H., Wong, C. H., Khor, C. C., Petric, R., Hibberd, M. L., & Nagarajan, N. (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), 11189–11201. https://doi.org/10.1093/nar/gks918

Wolf, S., & Moritz, R. a. (2008). Foraging distance in Bombus terrestris L. (Hymenoptera: Apidae). *Apidologie*, *39*(4), 419–427. https://doi.org/10.1051/apido:2008020

Wolf, T. J., Ellington, C. P., & Begley, I. S. (1999). Foraging costs in bumblebees: field conditions cause large individual differences. *Insectes Sociaux*, *46*(3), 291–295. https://doi.org/10.1007/s000400050148

Wommack, K. E., Nasko, D. J., Chopyk, J., & Sakowski, E. G. (2015). Counts and sequences, observations that continue to change our understanding of viruses in nature. *Journal of Microbiology*, *53*(3), 181–192. https://doi.org/10.1007/s12275-015-5068-6

Zwart, M. P., & Elena, S. F. (2015). Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annual Review of Virology*, *2*(1), 161–179. https://doi.org/10.1146/annurev-virology-100114-055135

Supporting Table 1 The species-site breakdown of the bumblebees used in the study.

	Locations									
Species	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

Supporting Table 2 The PCR primers for each virus used in the study

Virus	**Primer F**	**Primer R**	**Ref**
Loch Morlich virus	AGTGGTGGAGATGGAGACGA	CCACAGATACCAGTGGCGTA	Pascall et al. 2018
River Luineag virus	ACCAGGTGGAACTCGTGTTT	GTACTCTGGACCTTTGCCGT	Pascall et al. 2018
Mayfield virus 1	TATCCGCCGGCGTAATCTTC	GGATCTGATCCGTAGCGTGG	Pascall et al. 2018
Mayfield virus 2	CGGCTGCGTTGCGTAGTATA	ACCTGCCGTGCTAACAAATA	Pascall et al. 2018
Slow bee paralysis virus	GAGATGGATMGRCCTGAAGG	CATGAGCCCAKGARTGTGAA	Lena Wilfert (*pers comm)*
Acute bee paralysis virus	CYATGGACACACCCTATGTG	CGCCATTTTGCTACTTCTCC	Lena Wilfert (pers comm)

Table 1 The Morisita-Horn dissimilarities of the bumblebee compositions of the different sampling 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

Table 2 The posterior means and 90% shortest posterior intervals of the coefficients of the effect of each environmental covariate on each virus on the link scale from the multivariate probit model.

	Precipitation	Maximum Temperature	Wind Speed
River Luinaeg virus	0.60 (0.00, 1.19)	-0.36 (-0.99, 0.23)	-0.22 (-0.79, 0.33)
Loch Morlich virus	0.37 (-0.46, 1.18)	-0.18 (-1.01, 0.70)	-0.02 (-0.87,0.83)
Mayfield virus 1	-0.50 (-1.11, 0.11)	0.66 (-0.11, 1.39)	0.56 (-0.08, 1.17)
Mayfield virus 2	-0.13 (-0.90, 0.63)	-0.10 (-0.86, 0.69)	-0.20 (-0.95, 0.53)
SBPV	-0.03 (-0.70, 0.66)	0.11 (-0.64, 0.88)	-0.64 (-1.40, 0.08)
ABPV	0.01 (-0.88, 0.91)	-0.15 (-1.13, 0.80)	-0.04 (-1.00, 0.91)

Table 3 The posterior correlations of the errors of each virus from the multivariate probit model, measuring the degree of co-occurance. Positive numbers represent excess co-occurrence beyond that predicted by covariates and negative numbers represent a dearth of double infections. 90% shortest posterior intervals for each correlation are shown in brackets.

	River Luinaeg virus	Loch Morlich virus	Mayfield virus 1	Mayfield virus 2	Slow bee paralysis virus	Acute bee paralysis virus
River Luinaeg virus	1					
Loch Morlich virus	**0.73** (0.57, 0.70)	1				
Mayfield virus 1	-0.16 (-0.60, 0.27)	-0.17 (-0.69, 0.36)	1			
Mayfield virus 2	-0.18 (-0.71, 0.35)	-0.09 (-0.66, 0.47)	0.05 (-0.13, 0.23)	1		
Slow bee paralysis virus	0.07 (-0.22, 0.33)	-0.06 (-0.36, 0.24)	0.12 (-0.18, 0.41)	-0.24 (-0.51, 0.01)	1	
Acute bee paralysis virus	-0.35 (-0.82, 0.10)	-0.31 (-0.81, 0.20)	0.09 (-0.10, 0.28)	0.05 (-0.15, 0.25)	0.15 (-0.12, 0.43)	1

Table 4 Approximations to Watterson's theta for sites over a homologous genomic region within the RdRp gene by host species. Combinations that weren't tested due to low numbers of mapping reads are marked with "-". Entries with 0.000 had no observed variation over the region under study. The point estimate is at the median predicted number of infected individuals and uncertainty corresponds to the values at the 5th and 95th percentile of the predicted number of infected individuals (see methods).

	Mixed *Bombus* pool	*Bombus terrestris* pool	*Bombus lucorum* pool	*Bombus pascuorum* pool
River Luinaeg virus	0.029 (0.029-0.030)	-	-	-
Loch Morlich virus	0.026 (0.025-0.027)	-	-	-
Mayfield virus 1	0.010 (0.009-0.010)	0.015 (0.014-0.015)	0.024 (0.023-0.025)	-
Mayfield virus 2	0.035 (0.034-0.036)	0.014 (0.012-0.016)	-	0.029 (0.028-0.030)
Acute bee paralysis virus	0.000 (0.000-0.000)	-	0.000 (0.000-0.000)	0.000 (0.000-0.000)
Slow bee paralysis virus	0.008 (0.008-0.008)	0.002 (0.002-0.002)	0.003 (0.003-0.003)	0.006 (0.006-0.006)

Figure 1 The locations of the sampling sites and species distributions and sample sizes of the bumblebees caught at them.

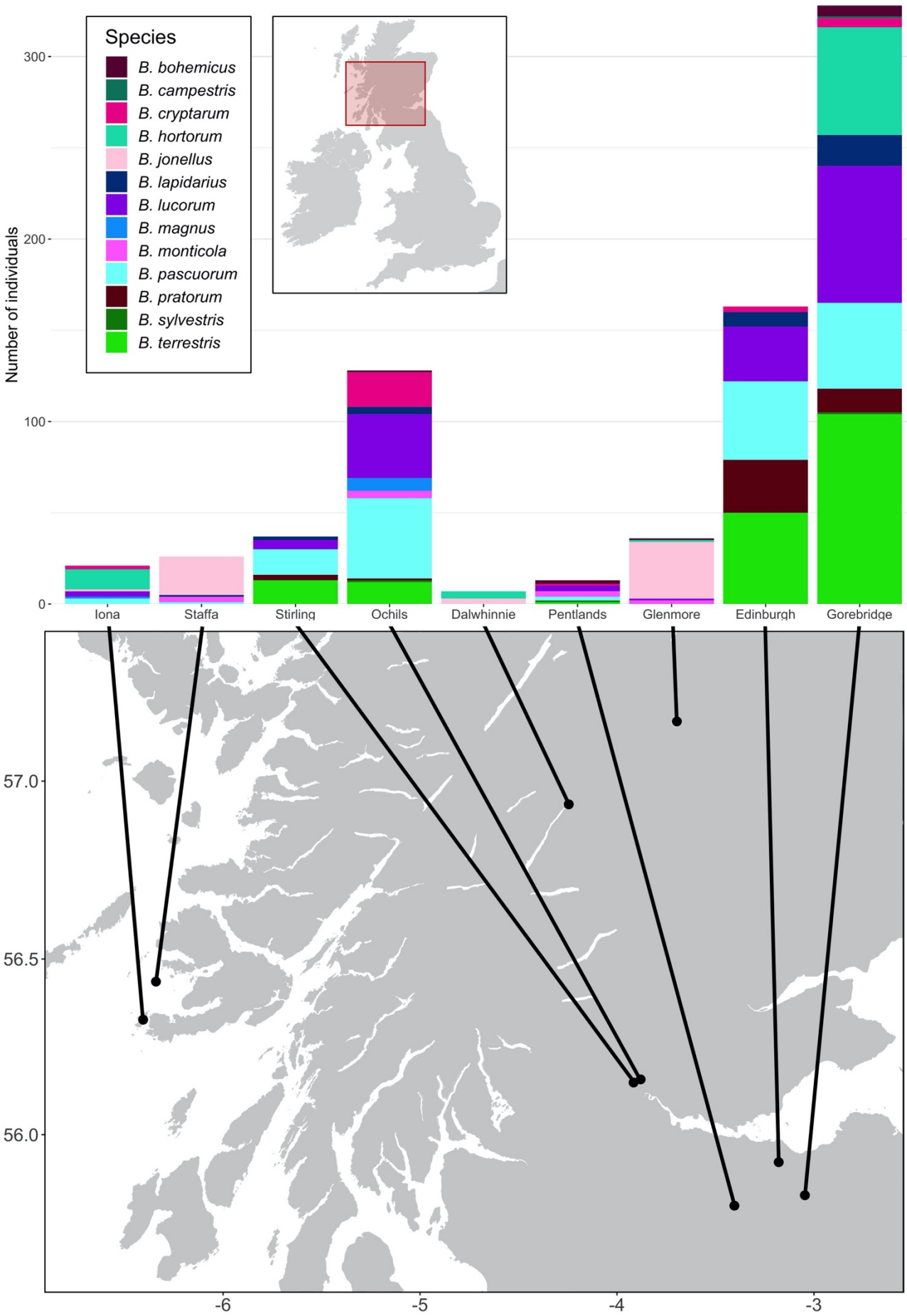
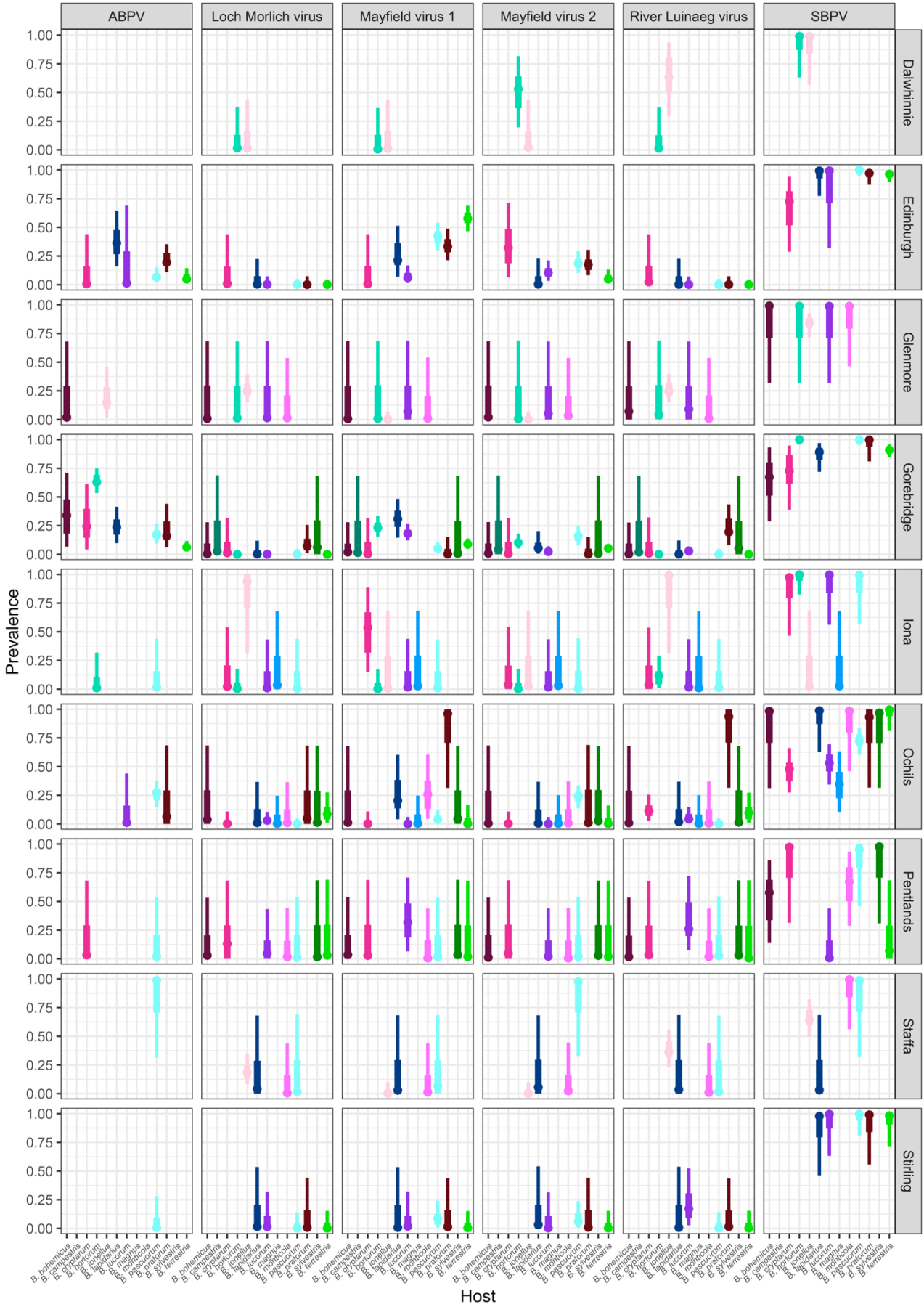

Species
B. bohemicus
B. campestris
B. cryptarum
B. hortorum
B. jonellus
B. lapidarius
B. lucorum
B. magnus
B. monticola
B. pascuorum
B. pratorum
B. sylvestris
B. terrestris
Number of individuals
300
200
100
0
Iona
Staffa
Stirling
Ochils
Dalwhinnie
Pentlands
Glenmore
Edinburgh
Gorebridge
57.0
56.5
56.0
-6
-5
-4
-3

Figure 2 The prevalence of Acute bee paralysis virus, Loch Morlich virus, Mayfield virus 1, Mayfield virus 2, River Luinaeg virus and Slow bee paralysis virus 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. Species are coloured by their corresponding colour in Figure 1 for ease of reading.


ABPV
Loch Morlich virus
Mayfield virus 1
Mayfield virus 2
River Luinaeg virus
SBPV
Dalwhinnie
Edinburgh
Glenmore
Gorebridge
Iona
Ochils
Pentlands
Staffa
Stirling
Prevalence
1.00
0.75
0.50
0.25
0.00
B. bohemicus
B. campestris
B. cryptarum
B. hortorum
B. ionellus
B. lapidarius
B. lucorum
B. magnus
B. monticola
B. pascuorum
B. pratorum
B. sylvestris
B. terrestris
Host

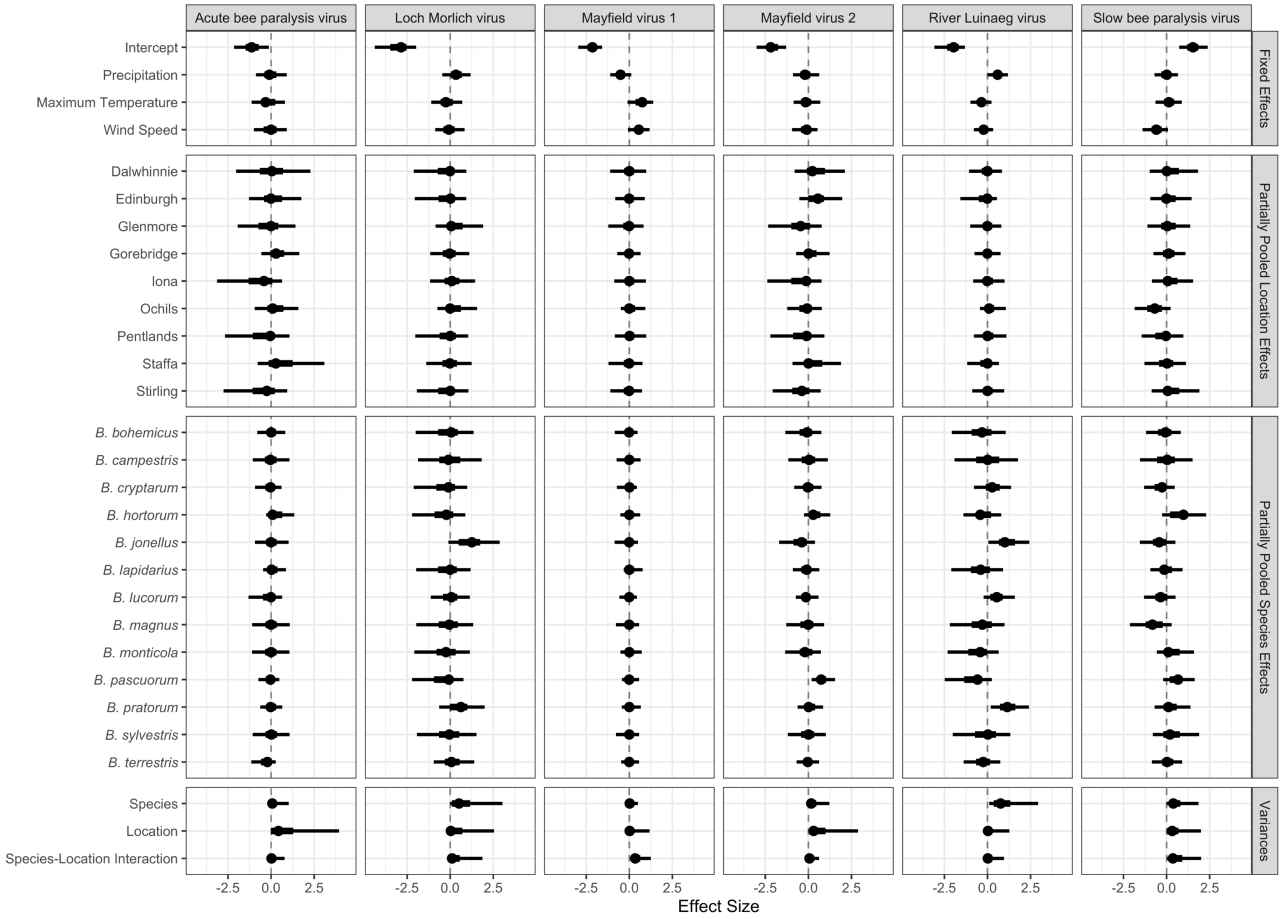



Figure 3 The estimates for each parameter in each virus from the multivariate probit model. The point estimate is the posterior mode, with 50% shortest posterior intervals represented by the thick lines and 90% shortest posterior intervals represented by the thin lines.