# The impacts of almond pollination on honeybee viral dynamics

Nina A. Sokolov[1], Graham R. Northrup[1,2], Lena Wilfert[3], Michael Boots[1]

[1] Department of Integrative Biology, University of California, Berkeley, Berkeley, CA 94720, USA

[2] Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA

[3] Institute of Zoology, University of Regensburg, 93053 Regensburg, Germany

Corresponding author: Nina Sokolov (nsokolov@berkeley.edu)

## Abstract

Honeybees provide critical pollination services, including the remarkable annual mass human-mediated intercontinental 'migration' where over half of all commercial honeybees in the United States are shipped to California's Central Valley to pollinate most of the world's almond supply. Clearly, this 'mass mixing event' with hives from across the country has the risk of spreading highly virulent pathogens, including viruses. In addition to their economic benefits, almonds also provide the first major food source for the honeybees emerging from winter. It is essential to weigh the advantages of the almond bloom against the risks of disease in honeybees, which may also impact native pollinators. We conducted an observational longitudinal RNAseq study of viral diversity across multiple honeybee colonies from a California-based beekeeping operation as they participated in almond pollination, compared to control hives that did not. Viral diversity in honeybee colonies increased transiently during almond pollination but returned to pre-bloom levels afterward, indicating a short-lived bloom-associated effect. Directly after the bloom, viral communities became more homogeneous across colonies, consistent with enhanced viral exchange through shared floral resources and inter-colony interactions. Viruses exhibited seasonal patterns in prevalence and abundance, yet no single species dominated the community. Instead, viral composition varied with site location and time, with hives in closer proximity displaying more similar viral dynamics, suggesting local mixing within orchards through drift. These findings indicate that the resource superabundance during almond bloom increases viral diversity and homogenization, but with a short-term effect as diversity and community structure return to pre-bloom conditions once hives leave the orchards.

Keywords: Virome, honeybee, virus, almond, pollination

## Introduction

Migrations are likely to enhance the geographical spread of pathogens, potentially increasing interspecies transmission and spillover risks (Altizer et al., 2011). Each season, the majority of commercially managed honeybees (*Apis mellifera,* Linneaus 1758*)* in the United States are transported to California's Central Valley to pollinate the almond orchards (*Prunis dulcis* (Mill.) DA Webb) (Ferrier et al., 2018; USDA-NASS, 2018), described as one of the largest pollination events in the world (vanEngelsdorp and Meixner, 2010). The almonds require animal pollination but also begin blooming in February when few native pollinators have emerged from overwintering diapause (Klein et al., 2012). Therefore, the industry is currently

heavily dependent on honeybees, with each acre of almonds requiring two honeybee hives for pollination (Carman, 2011). Honeybees will regularly fly over three km away from their hives to forage (Beekman and Ratnieks, 2000), meaning bees in a large orchard could potentially share forage and interact with hives from across the country. These interactions pose a heightened transmission risk of parasites and pathogens, including the ectoparasitic mite *Varroa destructor* (Anderson & Trueman, 2000) and RNA viruses.

Although there is concern about how migratory practices impact honeybee health, relatively few studies follow the consequences of commercial hive movement for almond pollination on disease emergence (Cavigli et al., 2016; Glenny et al., 2017; Alger et al., 2018; Faurot-Daniels et al., 2020). Nevertheless, this mixing of hives from all over the US runs the risk of spreading the most transmissible pathogen strains throughout the honeybee population. Theory suggests long-distance global transmission events can select for higher levels of pathogen infectivity and virulence (Boots and Sasaki, 1999). Some focal bee viruses appear to be generalists, being found in a wide range of insect species, some of which show active replication (McMahon et al., 2015; Tehel et al., 2016). Whereas other studies find that only a small fraction of viruses screened have evidence of active replication across multiple species (Doublet et al., 2025). Therefore, this poses a considerable hazard not only to managed honeybees but also has the potential to increase the risk of spillover to cohabiting native bees at stopover sites. While it seems likely that this mass migration would lead to considerable mixing, we have little information on how honeybees from different hives interact within mass flowering crops (MFCs) and what that means for viral transmission.

Horizontal transmission of viruses can occur indirectly through the environment on contaminated floral resources (Alger et al., 2019; Burnham et al., 2021) or through direct transmission from natal hive members (Chen et al., 2006) and non-natal honeybees through inter-colony drift and robbing (Fries and Camazine, 2001). MFCs can decrease disease incidence by diluting the indirect interactions on flowers due to the sheer abundance of floral resources (Figueroa et al., 2020; Piot et al., 2021). Alternatively, MFCs have also been shown to increase disease incidence, likely via increased contact between susceptible pollinator host abundance on the same flowering resources (Cohen et al., 2021; Piot et al., 2019).This pattern is likely pathogen-specific, with previous work showing decreased deformed wing virus (DWV) prevalence associated with increased non-crop bloom, but not for acute bee paralysis virus (ABPV) (Manley et al., 2023).

Drifting is the movement of individuals from one hive to another, and can occur from orientation errors after foraging flights (Neumann et al., 2000). The rate of drifting depends on both the misorientation as well as the acceptance level of the new hive (Bordier et al., 2017). Rates of drifting can vary based on the density and configuration of the apiary (Bordier et al., 2017; Seeley and Smith, 2015), as well as virus infection (Geffre et al., 2020; Iqbal and Mueller, 2007), or varroa infestation (Kralj and Fuchs, 2006; Kulhanek et al., 2021). Guard bees at the hive entrance only check a small percentage of incoming bees into the hive, and acceptance level has been shown to increase during times of abundant forage (Bordier et al., 2017). In this study, the abundance of floral resources and the high density of hives in a mass monoculture landscape with limited visual landmarks would suggest an increase in hive acceptance levels and bee orientation errors. Therefore, honeybee virus patterns may show spatial heterogeneity, with hives closer together sharing a more similar viral community than

hives further away in the orchard, due to honeybees drifting into non-natal hives (Pfeiffer and Crailsheim, 1998).

Longitudinal studies that monitor individual colonies are rare, with most assessing health at the total apiary level, examining viruses across all hives rather than looking at individual hive dynamics and interactions. A handful of studies have looked at honeybee viruses through time in the almonds (Alger et al., 2018; Faurot-Daniels et al., 2020; Glenny et al., 2017), but have focused on specific viruses and not looked at the entirety of the viral community. To date, fine-scale temporal tracking of viruses throughout the almond bloom has not been carried out, and in general, there are few studies monitoring individual commercial honeybee hive viruses at a transcriptomic scale through time. As such, there is a clear need for enhanced viral monitoring through surveillance transcriptomic testing. This resolution is necessary to understand the spread of viruses between hives and to monitor viral dynamics associated with the mass movements of honeybee hives for crop pollination. This work is currently especially prudent, with the 2024-2025 season boasting the highest losses of commercial honeybees in the United States on record, with over 60% losses amounting to over 1.7 million hives lost. Early work by Lamas et al. (2025) suggests that viruses are playing a role in these record-breaking losses, with highly virulent strains of DWV and ABPV being present in dead and dying hives.

Understanding the impact of mass pollination events on the virome of managed honeybees is key to the long-term management of pollinator health. In this study, we investigated the impact of participating in almond pollination on the virome composition in honeybees in a single beekeeping operation based in California. The temporal changes in the honeybee virome were quantified throughout the almond bloom via a longitudinal survey following honeybee hives across four time points in the Spring of 2021. Sampling occurred before, during, and after participation in the almond bloom in California, in comparison to control samples that were collected from hives that did not participate in pollination. Treatment hives were sampled across six spatially distinct ‘blocks’ to determine if viral composition was more similar between hives that were physically closer to one another via honeybee colony drift. A viral time series was generated, and individual hive viromes were examined across time to determine which viruses were associated with almond pollination. We hypothesized that hives participating in the almond bloom would experience higher viral pressures, leading to increases in viral richness and diversity. If contaminated almond pollen was the potential source of viruses for honeybees, this would only be seen at the final time point in the series, as there is a time delay between when pollen is collected, stored as bee bread, fed to larvae, and for those bees to emerge as foraging adults (Figure 1). We predict that if direct interaction is the source of these viruses, then we would see the highest values at the second and potentially third time point, during and after the almond bloom, when the largest number of honeybees from different hives would be interacting. We also predict that hives that were closer together spatially (in the same block, same location) would have more similar viral compositions.

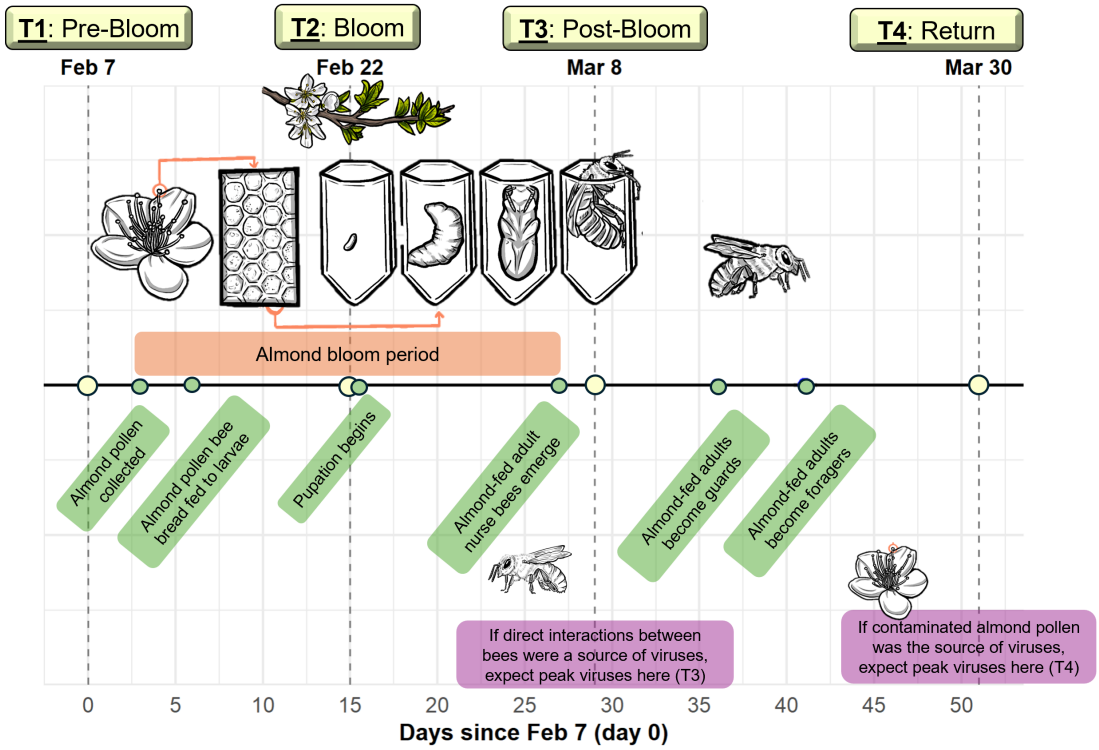


**Figure 1. Timeline with days since the first sampling point depicting the almond blooming period and honeybee development.** This illustrates the earliest time by which almond pollen would be collected, processed into bee bread, fed to larvae, which would then develop into pupae, nurse bees, guard bees, and finally forager bees. This depicts the earliest point at which purely almond pollen-fed adult forager bees would emerge. If contaminated almond pollen were a primary source of viral transmission to honeybees, then we would expect the highest viral abundance to be at the final time point (T4). However, if direct interactions between honeybees were the primary source of viruses, we would expect the highest viral abundance to be at the time point after the totality of almonds finished blooming (T3).

# Material and Methods

## Sampling and RNA extraction

Sampling took place across two almond orchards in the northern region of the Central Valley in California. All hives belonged to a single commercial beekeeping operation. Beekeeping operations and management practices act as significant contributors to variation in disease dynamics (Glenny et al., 2017; Martínez-López et al., 2022). To control for this variation, we focused sampling on a single beekeeping operation that was based in California. Honeybee foragers were collected from 'treatment' hives that were in an orchard within Arbuckle, CA, and sampled between [02/09/21 – 04/20/21]. A total of 30 treatment hives were sampled during this time, all from the same beekeeping operation. The spatial composition of the orchard consisted of approximately 25 hives per 500 meters in clustered "blocks". Treatment hives were selected randomly by numbering each of the hives in the block and using a random number generator to select focal hives across six blocks that were followed for the remainder of the study. All the treatment hives were within a three km radius of one another. Control hives were isolated from direct interactions with almond orchards and sampled in parallel at

the beekeeping operation's home yard in the foothills of the Sierra (Grass Valley, CA) and were also within a three-km radius of each other (Figure 2a). Due to constraints with the beekeeping operation, the control hive sample size was low. In total, only three strong hives were left behind from the almonds that could be used as comparable controls. These hives were part of a breeding management regime to select for varroa resistance. This resulted in samples across four time points: two weeks before the bloom (T1), during the bloom (T2), two weeks after the bloom (T3), and then a final time point when the almond and control hives were brought together three weeks after the bloom (T4) (Figure 2b). Foragers were sampled at the hive entrance using an insect vacuum. Approximately 50 workers were sampled per hive at each time point and placed into 50 ml vials. Samples were frozen on dry ice during transport, with a sub-sample of five foragers per hive placed into cryovials and flash frozen in liquid nitrogen. Samples were then stored at -80 degrees Celsius on the same day for further processing.

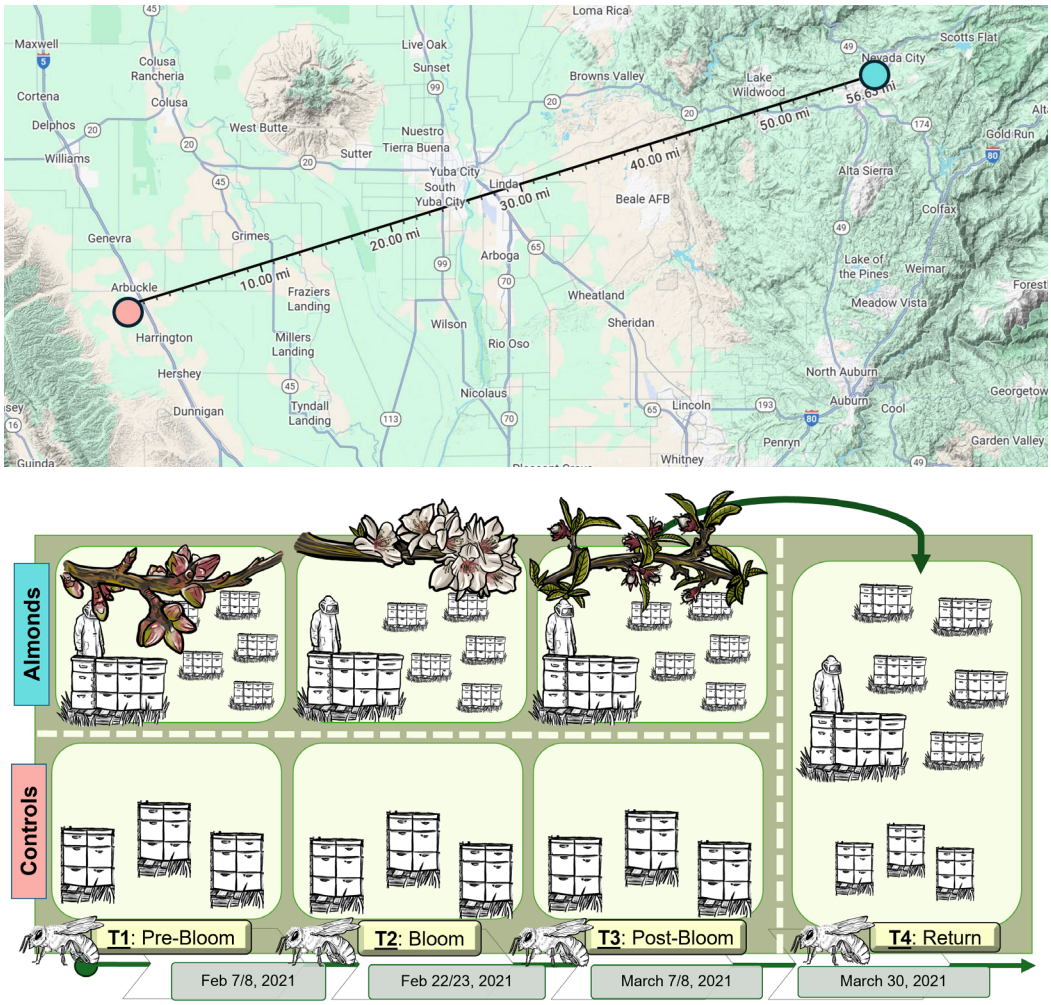





**Figure 2. (a) A map of northern California depicting the distance between the almond sites** (blue) and control sites (orange) at approximately a 56-mile distance from one another. **(b) The sampling scheme involved sampling hives in the almonds (top) or controls (bottom) across four time points (February – March 2021).** The first time point was before the almonds bloomed (T1), the second time point was during the almond bloom (T2), and the third time point was after the almonds finished blooming (T3). The final time point involved sampling once the almond and control hives came back together in their home yard.

## Pools for RNA-seq

Five foragers per hive were pooled together by homogenizing tissue in bead-beating tubes containing a single steel ball and zirconia beads in an MPBio FastPrep-24 Classic Bead Mill (speed 4 m/second for 30 seconds; 3 cycles; 20-second dwell) (Runckel et al., 2011). RNA was extracted using standard Trizol protocols (Manley et al., 2020). RNA was extracted using 1.3 ml TRI Reagent (Invitrogen) and 100 µl of 1-Bromo-3-chloropropane (Sigma-Aldrich), precipitated with 500 µl of isopropanol, washed with 1000 µl of 70% EtOH twice, and eluted in 150 µl of ultra-pure water. DNA was removed from the samples using TURBO DNA-free DNAse Treatment (Thermo-Fisher) and cleaned using an RNA Clean and Concentrator kit (Zymo). RNA concentration and quality were measured using a spectrophotometer, Nanodrop (OD 260/280 ≥ 2; OD 260/230 ≥ 2). RNA integrity of sample pools was tested using a BioAnalyzer Agilent) (s ≥ 4.0, with flat baseline). Samples that passed quality control checks were sequenced on the Illumina NovaSeq PE150 machines with Novogene (Davis, CA). A poly-A tail enrichment step was done during library prep as a way to preferentially sequence RNA viruses that often have long 3' poly (a) tails, along with host mRNA.

## Bioinformatics pipeline

In total, 34 treatment hives and three control hives were sampled over time. A total of 131 mRNA-seq libraries were generated for the samples that successfully passed quality controls. CZ ID was used for the bioinformatic quantification methods (Kalantar et al., 2020). Sequence adapter trimming and low-quality read removal were done using 'fastp' (Chen, 2023). Honeybee and any contaminant human reads were removed using 'Bowtie2' (Langmead & Salzberg, 2012) and 'HISAT2' (Zhang et al., 2021). After filtration, between 2.5 - 19 M reads remained. Taxonomic classification was carried out by aligning and querying the reads against the NCBI nucleotide reference database using 'Minimap2' and 'DIAMOND' (Kadlečková et al., 2022). Contigs were assembled using 'SPAdes' (Bankevich et al., 2012), and reads were mapped against contigs using 'Bowtie2' to assess genomic coverage. From the alignments, a taxon report was generated that showed taxonomically identified viral taxa of interest, the number of nucleotide total reads (NT r), and the reads per million (RPM). Cumulative nucleotide coverage (cumulative_NT), was calculated as:

$$\text{Cumulative_NT} = \text{Reference length} \times \text{Coverage breadth} \times \text{Coverage depth}$$

If contig information was not available, the cumulative nucleotide coverage was instead calculated as the number of reads multiplied by the average alignment length (L), though this is an overestimate. Samples were excluded from further analysis if they had fewer than five total reads and had a cumulative NT < 100 (Chen et al., 2023). Low confidence samples were defined as having greater than or equal to 10 reads, cumulative NT > 100, and no contig data, resulting in no available coverage breadth or depth estimates. Moderate confidence samples were defined as read count greater than or equal to ten, breadth of at least 2%, depth greater than 0.1, and cumulative NT greater than 100. High-confidence samples were defined as having greater than or equal to 50 reads, with coverage breadth of at least 2%, cumulative NT greater than 250, and depth greater than 0.5. As viruses were not being specifically amplified, and our study aimed to capture the full breadth of the viral community, we sought to assess low, moderate, and high confidence detections across different analyses, as detailed below. Low-confidence detections were excluded from quantitative analyses to avoid inflation due to the absence of contig support, thereby minimizing the potential for false positives. Detections

212 between 10–49 reads were reported but not included in the community-level analysis. A
213 taxonomy table was generated to describe the taxonomic classification of each of the focal
214 viruses with the taxonomic ranks, which include Kingdom, Phylum, Class, Order, Family,
215 Genus, and Species. To normalize across differences in sequencing effort amongst samples,
216 we applied a hybrid normalization strategy. If contig-level data were available with coverage
217 depth indicating the average number of reads mapped relative to the viral genome, we
218 calculated normalized depth (avg_norm_depth) as:

219 Normalized depth = coverage depth x library size / average library size

220 For viruses that lacked contig coverage estimates, we normalized with raw read counts as:

221 Normalized reads = reads aligned x library size /average library size

222 From there, both normalized read counts were additionally converted to relative abundance
223 metrics by dividing the abundance of each of the viruses by the total normalized abundance of
224 all viruses in each sample. To calculate prevalence, we created presence/absence matrices by
225 scoring any non-zero abundance as present and zeroes as absent. For richness and diversity
226 analyses, low-confidence detections with stronger support (≥50 reads and average alignment
227 length ≥80 bp) were retained along with moderate and high confidence detections. As both
228 high and moderate confidence detections were supported by contig-level data, both were used
229 to calculate prevalence and quantitative abundance estimates.

## 230 Virus abundance and virome diversity

231 Analyses were performed in R v4.4.0 (R Core Team, 2014). BioManager was used to install
232 Bioconductor packages (Morgan and Ramos, 2024). Viral species richness was quantified as
233 the number of different viral OTUs/species found in each hive at each time point. Richness was
234 modeled using a generalized linear model (GLM) with a Poisson error distribution and log link
235 function. The model included fixed effects of sampling time point, location, and their
236 interaction (richness ~ Time_point × Location). Model residuals were examined to confirm
237 assumptions of the Poisson distribution. Additionally, to assess temporal changes of individual
238 viruses within Almond colonies, we fit separate GLMs with binomial error distributions for
239 each virus, using presence/absence of detection as the response variable and sampling time
240 point as the predictor. Estimated marginal means were calculated using the 'emmeans' package
241 and pairwise contrasts were assessed to test the differences in viral richness between locations
242 and time points and viral prevalence between time points in the almonds. P-values were
243 adjusted using the Benjamini-Hochberg procedure to account for multiple comparisons.

244 All plots were visualized using ggplot2 (Wickham, 2016). To begin assessing community
245 composition differences between times and locations, mean viral abundance was calculated
246 and plotted for different taxa across samples (Koch et al., 2018). Absolute abundance was
247 graphed across the viral species with the most amount of variation and representative data to
248 compare temporal patterns between locations. Viral community diversity was quantified at
249 each time point using relative abundance measures (Chem et al., 2023) to calculate alpha
250 diversity metrics. The diversity of species within a single sample was assessed using Shannon's
251 diversity index and Simpson's diversity index (Magurran and McGill, 2011) and calculated
252 using 'phyloseq' (McMurdie and Holmes, 2013). Shannon diversity gives weight to rare
253 species, whereas Simpson's diversity index weighs evenness more heavily and is therefore
254 more sensitive to common species (Roswell et al., 2021).

We used a linear mixed effects (LMM) model with Shannon/Simpson diversity index as the dependent variable, with fixed effects being "Time" (Levels: T1 = Before bloom; T2 = During bloom, T3 = After bloom; T4 = All back together), "Location" (Levels: Almonds versus Control), their interaction (Time x Location), and "Hive_ID" as a random effect to account for repeated measurements using the 'lme4' package. A type III ANOVA was used to test for statistical significance with corrections for multiple testing. The Shapiro-Wilk Test was used to plot the distribution of Shannon diversity values to assess the distribution's normality. The Levene's Test for homogeneity of variances between groups was performed using the "car" package (Zhou et al., 2023). Assumptions and goodness of fit for the model were quantified and visualized as Q-Q, Residual vs Fitted plots, Scale-location, and residuals vs leverage plots.

Beta diversity was calculated using each hive's normalized abundance data and Bray-Curtis dissimilarity with the 'phyloseq' and 'vegan' packages (Okansen et al., 2024). To assess variation in viral community structure across space and time, we conducted a non-metric multidimensional scaling (NMDS) ordination of the Bray–Curtis dissimilarity values. Before ordination, normalized read counts were log-transformed to reduce the influence of highly abundant viruses. Samples with zero total counts were removed. For visualization of beta diversity patterns, samples containing only a single detected virus were also excluded, but the data were retained in the overall dataset for other analyses. The resulting ordination was visualized with viral communities color-coded by block, yard, time point, and location to assess clustering. To test for significant differences in community composition, we used Permutational multivariate analysis of variance (PERMANOVA) in the 'adonis2' package on the Bray–Curtis distance matrix, with time point, location, and block as predictors (9999 permutations). To evaluate effects based solely on community composition rather than abundance, data were converted to binary presence–absence matrices, and Jaccard dissimilarities were computed before using a similar PERMANOVA model structure.

To assess whether our variables of interest had inherent differences in dispersion values that could be driving differences in beta diversity metrics, we used a permutation test for homogeneity of multivariate dispersions (PERMADISP). The Bray–Curtis dispersion was calculated as the distance of each sample to its hive centroid in multivariate space. Distances were calculated using the 'vegan' package and the Bray–Curtis dissimilarity matrix. Dispersion was visualized with box plots by hive, over time, across blocks, and between locations. To summarize variability across hives, we calculated the standard deviation and range of dispersion values across time points for each hive.

To identify which viruses contributed the most to differences in community composition through time, we performed a SIMilarity PERcentage (SIMPER) analysis using the 'vegan' package. Pairwise dissimilarities were calculated using Bray–Curtis distance, and grouping was based on time point and location. For each pairwise time point comparison, we extracted the average contribution of each virus to Bray–Curtis dissimilarity and computed the cumulative contribution to identify the top drivers of compositional change. Viruses contributing to the top 90% of total dissimilarity for each contrast were retained and visualized using bar plots faceted by differences in viral community between time points.

To determine potential co-infection dynamics between viruses, we assessed patterns of viral co-occurrence across hive samples by converting abundance data into a binary presence or absence matrix for viral detection across hives. Diagonal values were set to zero to exclude

self-co-occurrence. To statistically test for significant patterns of virus co-occurrence, we applied a probabilistic model on the virus matrix data using the 'cooccur' package. Observed co-occurrence frequencies were then compared to the expected values to determine whether viruses co-occurred more or less often than expected by chance. For each significant pair, we calculated the $\log_2$ ratio of observed to expected co-occurrence and plotted the values on a heatmap using 'pheatmap' (Kolde, 2018). Finally, normalized viral coverage data were analyzed using the packages 'dplyr', viridis', and 'gridExtra'. Normalized values were log-transformed ($\log_{10}[x + 1]$) before visualization. A heatmap was produced using 'pheatmap' with hierarchical clustering applied to viral taxa, with grouping by location and time point.

# Results

## Virus prevalence and abundance

We examined the virome composition of 37 hives of honeybees *(Apis mellifera)* from a single commercial operation across four time points in 2021 that either participated in the almond pollination (n=34) or stayed behind at their home yard and acted as controls (n=3). Of those hives, we successfully sequenced and have longitudinal data tracking viral community changes across at least three time points for 29 treatment hives and all three control hives. The remaining eight hive samples did not have sufficient quality for sequencing each time point. We found 28 known insect-associated viruses that passed our filters with more than 2% genome coverage in at least one sample. These viruses mostly belonged to viral families with positive-sense single-stranded RNA genomes (Dicistroviridae, Iflaviridae, and Sinhaliviridae) but also included three Rhabdoviridae with negative-sense single-stranded RNA genomes, one unclassified dsDNA genome, and other unclassified groups. In each hive library, we found between 0 – 8 viruses.

Viral prevalence was assessed for the almond hives and the control hives, as the number of hives found infected with that focal virus over the total number of hives sampled (Figure S1). The viruses with the highest normalized abundance throughout all time points Apis mellifera filamentous virus (AMFV), Apis rhabdovirus, Black queen cell virus (BQCV), Hubei partiti-like virus 34, deformed wing virus (DWV), Lake Sinai virus complex (LSV), sac brood virus (SBV), and uncultured virus 2, which is discussed more below (Figure S2). Some of these viruses in almond hives had significant changes in viral prevalence through time, hinting at seasonal patterns. We describe the rest of the non-signifcant patterns qualitatively. AMFV prevalence increased during and after the bloom before dropping to pre-bloom levels at the final time point and was found at relatively high abundance in almond and control hives throughout the entirety of the study. AMFV prevalence was significantly higher at T2 compared to T4 ($p = 0.021$) and at T3 compared to T4 ($p = 0.036$). Acute bee paralysis virus (ABPV) had the highest prevalence before the bloom and was not present in any of the hives by the final time point. ABPV was only found in the almond hives and was at a consistently lower abundance than many of the other viruses of interest.

BQCV was found in both almond and control hives, with prevalence peaking in almond hives after the almonds bloomed, as nearly all hives in the apiary tested positive, but dropped down to pre-bloom levels at the final time point. BQCV prevalence increased from T1 to T2 ($p = 0.041$), and was significantly higher at T2 compared to T4 ($p = 0.034$) and at T2 compared to T4 more strongly ($p = 0.002$). BQCV abundance stayed steady in the almond hives through

time and increased at the final time point. In control hives, BQCV was found consistently lower abundance than the almond hives and peaked in prevalence at the third time point along with the almond hives. DWV-A and DWV-B were both found at high abundance in the almond hives, with prevalence dropping from the beginning to the end of the study. Both strains' abundance stayed nearly consistent before and after the bloom; however, DWV-B abundance levels dropped sharply at the third time point before rebounding when hives returned to the home yard. DWV-A was only found as low-confidence reads in control hives and only for the first two time points and was therefore not included in this analysis.

The most abundant strains of the LSV-complex were LSV, LSV-2, and LSV-3, with LSV-2 having the highest prevalence. LSV was higher in almonds than in controls and peaked right after the almond bloom but returned to pre-bloom levels at the final time point. The control hives had LSV-2 from the beginning at high prevalence, with LSV and LSV-3 absent until the final time point, where prevalence increased to similar levels as the almond hives. LSV-2 was found at a similar high abundance in both almonds and controls pre-bloom but then decreased in the controls consistently until the end of the study. LSV was at lower abundance in the controls than in the almond hives at the final time point, whereas LSV-3 was at similar levels.

SBV was at relatively low abundance in the almond hives pre-bloom before increasing during the bloom, dropping again once the bloom was finished, and increasing at the final time point when both sets of hives came back together. Prevalence in the almond hives increased consistently through the study. SBV was not found in controls. Uncultured viruses 1 and 2 were found in both almond and control hives at relatively high abundance. Uncultured virus 1 was high in controls pre-bloom, whereas it was low in almond hives, then almond hives increased substantially in prevalence after the almonds bloomed before dropping to pre-bloom levels at the end of the study. Uncultured virus 1 also showed increased prevalence from T1 to T2 ($p = 0.034$). Uncultured virus 2 peaked in abundance in almond hives after the almonds bloomed and peaked in prevalence in the control hives during the bloom. Although these viruses have yet to be cultured, previous studies have found uncultured virus 1 associated with genomic surveys of varroa (Cornman et al., 2010) and the uncultured virus 2 to be associated with collapsing hives (Cornman et al., 2012).

## Virome diversity

Viral richness was calculated as the total number of unique viral OTUs per hive per time point and modeled using a GLM including fixed effects of time, location, and their interaction. Model diagnostics (residual plots, overdispersion test, and DHARMa simulations indicated no violations of Poisson assumptions (dispersion $p=0.57$, zero-inflation $p=0.48$). Throughout the time series, the almond hives increased in viral richness at the second and third time point (during and after the bloom) before dropping back to pre-bloom levels in the final time point (Figure 3, top). We found significant differences between Almond and Control samples at time point T3 (Almonds > Control, $p = 0.05$). Within the Almond location, marginally significant changes in richness were detected between T2 and T4 ($p = 0.059$), and a significant decrease between T3 and T4 ($p = 0.049$). The control hives had overall lower viral richness at all time points. Shannon diversity through time was compared between the almond pollinators and control sites (Figure 3, bottom). The distribution of Shannon Diversity values showed a positive skew. A Shapiro-Wilk Test confirmed the data were not normally distributed ($p < 0.001$). However, when testing model residual diagnostics (Q–Q, residuals vs. fitted, scale–location,

and residuals vs. leverage), we saw residuals were approximately normally distributed, with Levene's test indicating approximate normality and homogeneous variances across groups (F=0.9, p=0.5), supporting the validity of the mixed model (Figure S3). We used the following linear mixed effects (LMM) model to test the significance of the effect of time, location, and their interaction, along with hive number as a random effect on Shannon Diversity:

```
Model_1 <- m_shannon <- lmer(Shannon ~ Time_point * Location + (1 | Hive_number),
            data = diversity_summary)
```

The model revealed a significant effect of location, but not time, on Shannon diversity (Table S1). Residual variance was larger (0.166) than the between-hive variance (0.010), indicating that most variation occurred within hives rather than among them. Post hoc analyses showed that Shannon diversity was significantly higher in almond hives compared to controls during mid-season sampling periods (Table S2). Specifically, diversity increased in almond hives from pre-bloom to the peak of the almond bloom at the second time point (estimate = 0.70 ± 0.25 SE, p = 0.006). Diversity remained elevated immediately following bloom at the third time point, during petal drop and early almond fruit development (estimate = 0.52 ± 0.25 SE, p = 0.042). In contrast, early-season (time 1) and late-season (time 4) samples showed no significant differences between almond and control hives (p > 0.6). These results indicate that Shannon diversity peaked in almond hives during and immediately after bloom, while controls remained comparatively low, reflecting a transient divergence in viral community structure between locations. When looking at viral diversity changes within hives, we can see that there is variation between hives as to whether Shannon diversity increases or decreases during and after the bloom, but most return to similar values as their pre-bloom (Figure S4). No significant interactions were detected at other time points or for the spatial block.

The Simpson's diversity index showed similar patterns; however, the control hives at the first time point show low Shannon value but high Simpson values (Figure S5). This indicates that although the viral richness is low, the high evenness suggests no one virus is dominating. This flips at time point 2 for controls, where there are very few viruses present, and one species dominates in each sample, therefore leading to a drop in the Simpson index. In one of the controls, this sample was dominated by 'uncultured virus 2', whereas the other sample was dominated by LSV-2. Overall, the control hives had variable diversity levels, likely due to the low sample size, but remained low through the duration of the study. The second time point boasted the lowest viral diversity, before increasing at the final time point when the almond pollinating hives returned.

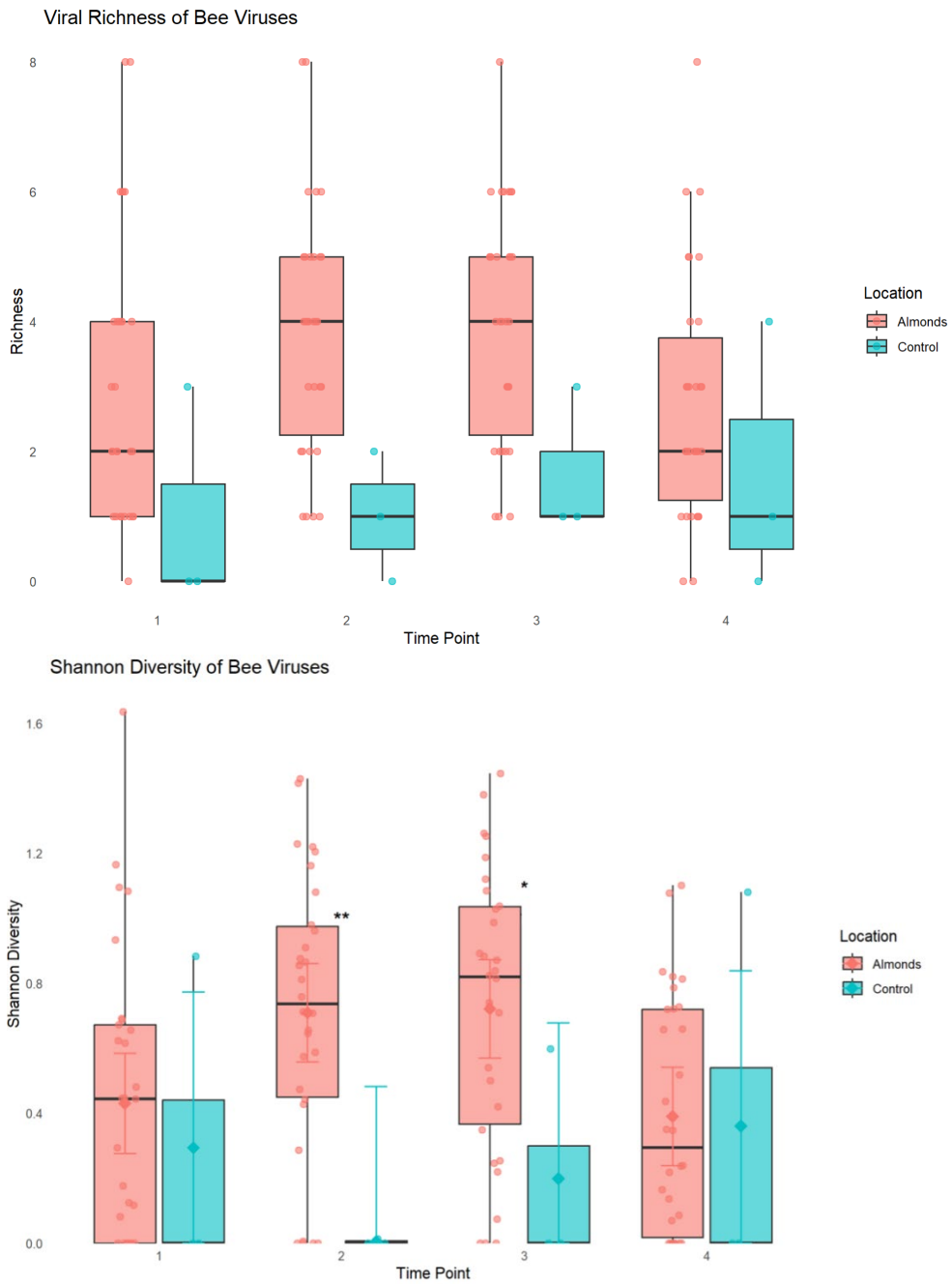


**Figure 3.** (Top) **Viral richness across time in almond pollinating and control honeybee colonies.** Boxplots show the number of distinct viral taxa detected per sample across four time points in almond pollinating hives (pink) and control hives (blue). Individual points indicate viral richness values for each sample. (Bottom) **Shannon diversity of bee viruses across time points and locations.** Boxplots of normalized Shannon diversity indices of viral communities in hives sampled across four time points.

***Beta Diversity Metrics***

To test the differences in viral community diversity between samples, we calculated beta diversity indexes and accounted for differences in species abundance using Bray-Curtis dissimilarity. From there, NMDS plots were generated for hives with ordination, and visual inspection of ordination plots revealed no strong clustering of viral communities by block, yard,

time point, or location (Figure S6). The almond hives in this study were sampled across six distinct spatial "blocks" of approximately 25 hives clustered closely together throughout the orchard. To better assess patterns of viral abundance across locations and through time, we generated a heatmap of relative abundance for each of the hive's viruses (Figure 4). We can see a clustering of a main core group of viruses, which includes LSV, LSV-2, BQCV, LSV-3, DWV-A/B, and AMFV. We used PERMANOVA to test the interaction effect between time and location, along with spatial block, on viral composition using Bray-Curtis dissimilarity. We found no significant effect of time point, location, or block on community composition ($R^2 = 0.035$, $p = 0.36$). This suggests that when viral abundance information is retained, overall community structure does not differ substantially among groups. In contrast, the Jaccard-based PERMANOVA revealed significant effects of both time point and block, and a marginal effect of location. Time point explained 5.7% of the variance ($R^2 = 0.057$, $p = 0.0003$), with community composition differing across sampling periods. Block also explained a notable portion of variation ($R^2 = 0.078$, $p = 0.0027$), indicating hive-to-hive or block-level heterogeneity. Location (almonds vs. controls) approached significance ($R^2 = 0.016$, $p = 0.0503$). This suggests that presence–absence patterns (Jaccard) are more sensitive to temporal and block-level differences than relative abundances (Bray–Curtis).

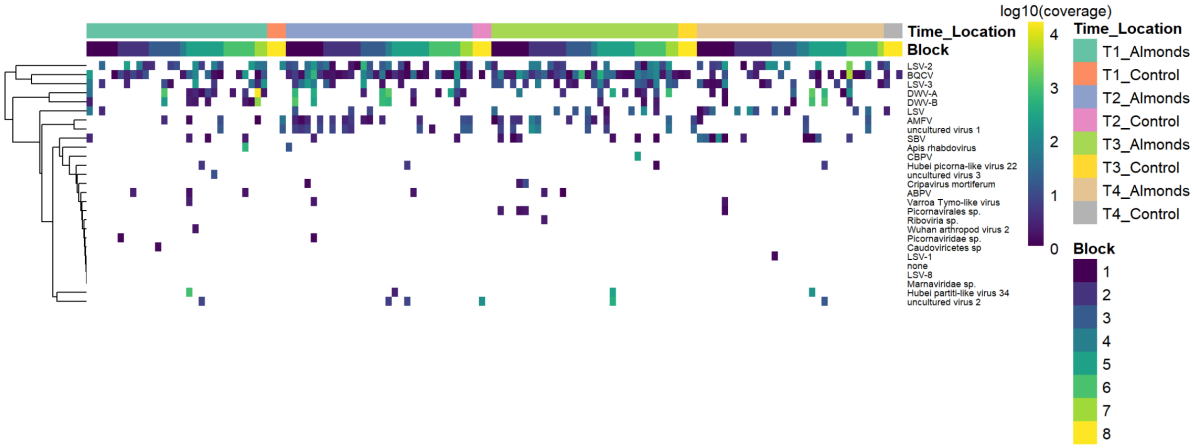


**Figure 4. Heatmap of viral community composition across time points, locations, and blocks.** Relative abundance values (normalized coverage, $\log_{10}$-transformed) are shown for detected viruses (rows) across all samples (columns). Viruses are clustered by similarity in abundance patterns, while samples are ordered by metadata (Time point × Location × Block). The top annotation bars indicate sample grouping by Time_Location and Block. Warmer colors represent higher relative abundance.

We calculated beta dispersion per block and ran a PERMANOVA test for homogeneity of multivariate dispersion. We found that there was no significant difference in the variation of viral community between block, time, or location ($p >> 0.05$). This means that viral community composition showed similar within-group variability across these factors, and PERMANOVA assumptions were met. However, when focusing specifically on almond hives through time, pairwise PERMDISP comparisons revealed a significant decrease in dispersion between Time Point 1 and Time Point 3 ($F = 4.12$, $p = 0.037$) (Figure 5). This finding suggests almond hives became more similar to one another, or homogenous, in their viral community composition after the bloom. But dispersion returns to similar levels as the first two time points at the end of the study, suggesting the viral community became more heterogeneous.

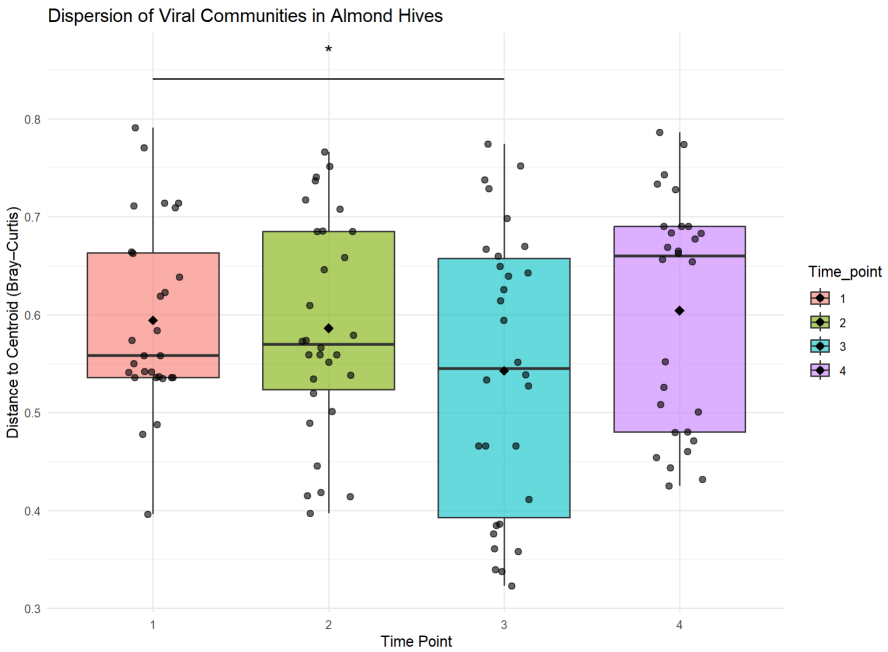


**Figure 5. Beta dispersion of viral communities in almond hives across time points to assess group-level heterogeneity in community composition.** Boxplots depicting within-group variability of distances to centroids illustrate the variation in viral community compositions based on Bray-Curtis dissimilarity. Dispersion is significantly different between time points 1 and 3 in the almonds (p_adjusted = 0.03).

SIMPER analysis showed which viruses were the most important contributors to differences in beta diversity between samples. Only viruses contributing to the top 90% of cumulative dissimilarity were shown. It revealed that a small subset of viruses accounted for the majority of dissimilarity between time points (Figure 6). In almond hives, BQCV and LSV-2 consistently contributed the most to dissimilarities between consecutive time points. For example, between time points 1 and 2, BQCV (~18%) and LSV-2 (~16%) explained the largest fraction of community turnover, alongside DWV-A and several low-abundance viruses. A similar pattern was observed between time points 2 and 3, where BQCV again dominated (17.5% contribution), followed by LSV-2, LSV, and DWV-A. In contrast, control hives showed limited temporal change, with only the comparison between time points 3 and 4 yielding notable dissimilarity. This shift was again dominated by BQCV and LSV-2, with uncultured virus 1 and LSV-3 also making contributions.

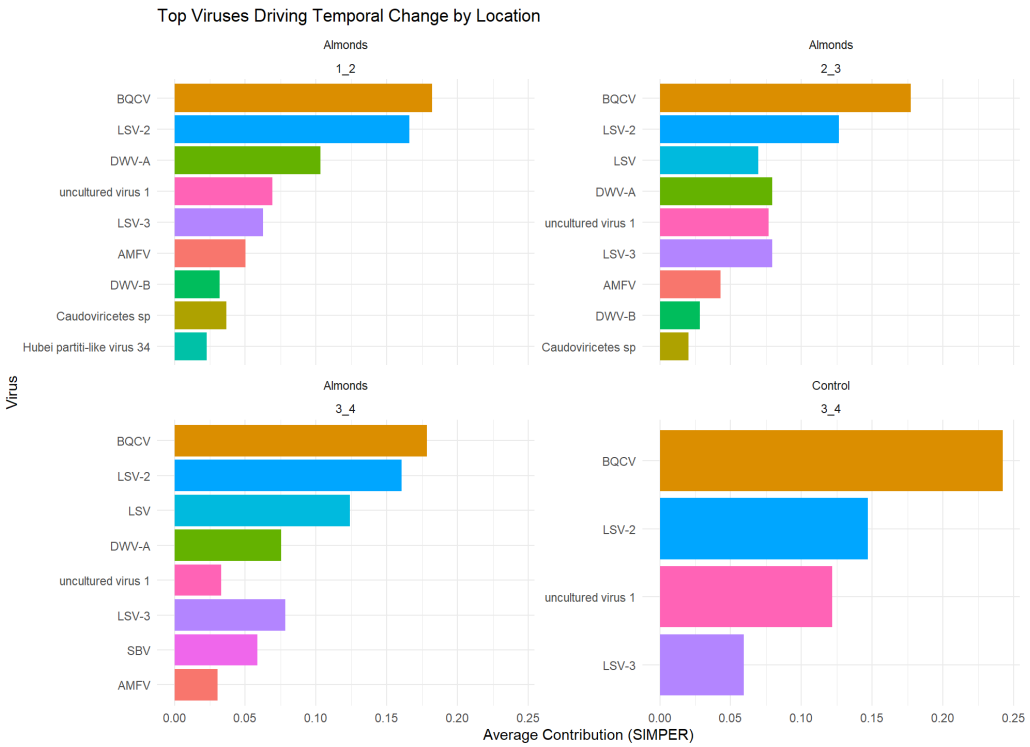


**Figure 6: Top viral contributors to dissimilarity in viral community composition between time points.** Bar plots show the top viral taxa that contribute to differences in beta diversity between all possible time point contrasts using SIMPER. Each panel represents pairwise comparisons between two time points, with the bars representing the contribution of that virus to the overall Bray-Curtis dissimilarity. The most influential viruses consistently contributed to the community differences across time point contrasts.

Viral pair co-occurrence patterns were compared against the null expectation using probabilistic modeling. Significant positive or negative associations were identified after accounting for prevalence. The resulting heatmap (Figure 7) displays the $\log_2$ enrichment values for each virus pair, representing how much more (red) or less (blue) frequently viruses co-occurred than expected by chance. Several significant positive co-occurrences were detected, with DWV-A and DWV-B showing the strongest association ($\log_2(obs/exp) > 1.5$, $p < 0.001$). AMFV also strongly co-occurred with uncultured virus 1 ($p < 0.001$) and with BQCV and LSV-3 ($p < 0.05$). Similarly, LSV-2 exhibited multiple co-occurrence relationships, including with DWV-A, DWV-B, and LSV-3 ($p < 0.05$–$0.001$).

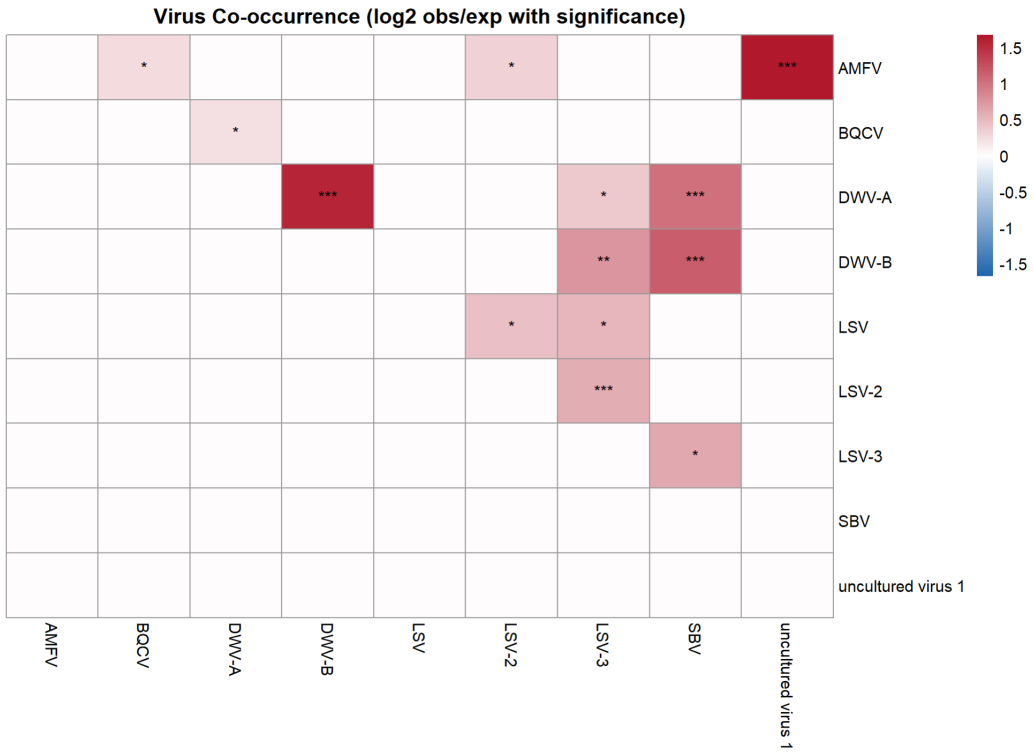


**Figure 7. Viral co-occurrence patterns across honeybee hive samples.** Heatmap of the $\log_2$-transformed enrichment of observed versus expected co-occurrence frequencies for virus pairs from probabilistic co-occurrence modeling. Viruses that occurred more often than expected by chance are in red, while co-occurrences that occurred less often than expected by chance were blue. Only viruses that are part of at least one significant pair are plotted. Color intensity is scaled by effect size, and asterisks indicate significance levels ($p < 0.05$ = *, $p < 0.01$ = **, $p < 0.001$ = ***).

# Discussion

Almond orchards provide a potentially high-risk environment for the emergence of virulent pathogens due to the mass mixing of over half of all commercial honeybees in the US, providing a wealth of susceptible hosts. Honeybees can theoretically interact with thousands of other honeybees through indirect interactions and environmental transmission on flowers, or through drift between colonies. At the same time, almond pollen is known to be nutritious and thus could bolster honeybee immune function through the bloom as the first natural food source coming out of winter (Loper et al., 1980). The almonds are a unique case in that there is minimal interaction with wild native bees that are not yet active during the peak bloom in February, and pollination is limited based on the number of hives that are imported to orchards. Although honeybees would not be directly interacting with many wild species of bees during the bloom itself (Reilley et al., 2020), the impact of the bloom on the viral dynamics of honeybees has the risk of impacting native bees when the honeybees are moved to their next pollination location. Therefore, it is important to study what viruses are emerging in honeybees from participating in this pollination event.

We generated an observational time series of viral diversity from a longitudinal RNAseq study of honeybee hives that were shipped to California's Central Valley to participate in almond

pollination in comparison to hives that did not participate in pollination. Our results show a high diversity of viruses, including 28 known insect-associated RNA and DNA viruses. Additionally, there was also evidence of plant viruses and bacteriophages, which can be the focus of future studies. Virus prevalence for focal viruses showed that almond hives had higher viral prevalence and abundance than controls, with some evidence of seasonality. Our study has shown which viruses increase and decrease in their abundance during the bloom and demonstrates that there is significant hive-specific variation in viral dynamics. Viral diversity increased during and after the almonds bloomed at the second and third time points. The highest range of viral diversity values was seen at the third time point, which is where we hypothesized this would be caused from direct interaction between bees, rather than consuming contaminated almond pollen. To get at this more concretely, warrants further investigation specifically looking at the viruses found on almond pollen.

Overall, we saw that viral diversity dropped to pre-bloom levels when almond hives returned to their home yard after finishing with almond pollination. There was no evidence of specific viruses emerging and dominating the viral community for the almond pollinating bees. The effect of increased viral diversity was transient, which means the benefits of honeybees pollinating the almond bloom may outweigh the disease impacts, at least in the studied beekeeping operation. Although Shannon diversity peaked after almonds bloomed, beta dispersion simultaneously decreased. This suggests that while viral communities within hives became more diverse, they also became more similar to one another across the apiary. Such a pattern is consistent with a bloom-associated homogenization effect, where shared floral resources and increased inter-hive contact during pollination promote the spread of a common set of viruses, leading to convergence in community composition across colonies.

When looking at changes in the viral community, we saw that the structure remains relatively stable across almond and control hives when relative abundances are considered, as evidenced by the lack of significant differences in Bray–Curtis dissimilarity. However, the significant effects detected using the Jaccard metric highlight that the presence or absence of viral taxa, rather than abundance, varies across time points and amongst spatial blocks. The strong temporal effect indicates that community membership shifts over time, potentially reflecting seasonal dynamics or changes in virus transmission. The effect of location also shows a difference in community assemblage between the almond and control hives. The significant block effects show that viral community composition is more similar amongst hives that are closer together in the orchard. This suggests viral interactions between hives through inter-colony drift. It appears that while the overall viral community structure is robust to location and time when abundances are considered, the presence of rare or transient taxa contributes to detectable shifts in beta diversity over time and space. The SIMPER analysis showed which viruses were significant contributors to changes in the viral community through time. BQCV and LSV-2 showed up as consistent drivers in these hives. Notably, earlier time points in control hives lacked sufficient viral diversity to generate meaningful contrasts, underscoring the consistently lower viral heterogeneity relative to almond hives. The overall abundance-weighted community structure, however, did not vary significantly among groups. These results indicate that while the relative abundance of these dominant viruses fluctuates over time, the shifts occur within a broadly stable community framework. The divergence between these metrics underscores that rare taxa drive much of the turnover in viral presence–absence structure, even as the dominant viral assemblage remains relatively consistent in abundance.

When looking at the viral co-occurrence patterns, results showed that viral infections in honeybee colonies are structured by non-random associations, with DWV strains and LSV variants frequently co-occurring, potentially reflecting shared transmission routes or interactions within hosts. All together, these results suggest that honeybee viral communities experience subtle temporal and hive-level shifts in membership, yet maintain stable abundance profiles dominated by a few persistent, co-occurring viruses.

BQCV, AMFV, LSV-2 and DWV-A all explained the most change in beta diversity across all time points. Lake Sinai virus (LSV) is a highly prevalent virus during the almond blooms that boasts high strain diversity, with unknown consequences on honeybees or other bee species (Daughenbaugh et al., 2015). ABPV was only found in the almond hives, but at a lower abundance than many of the other viruses of interest. This virus is of particular concern for potential risks of viral transmission to wild bees, as this is one of the examples of a definitive multi-host virus, as seen in United Kingdom bee communities (Doublet et al., 2025). Several of the viruses present in this study have been found in hives across a wide geographic range, but with little information available so far on their biology. Hubei partiti-like virus 34 has been found through metagenomic studies in honeybee hives across the world, including the United States (McKeown et al., 2025), Korea (Kwon et al., 2023), and Uzbekistan (Kwon et al., 2024). Uncultured virus 2 and 3 were found in a previous genomic survey of hives experiencing colony collapse disorder, but little is known about them (Cornman et al., 2012). We found individual viruses mostly in the almonds, not the controls. However, this could be explained by low sample sizes. This pattern becomes more pronounced when looking at the high-confidence-only data. The control hives that were left behind were the only strong hives that did not go to pollinate because they were being used in a breeding regimen, focusing on traits for varroa resistance. The low disease incidence in controls could then also be a function of their ability to maintain low varroa numbers without chemical treatment. This was particularly notable for the surprisingly low DWV prevalence and abundance, which is normally a virus commonly found in commercial hives.

Although other studies have characterized honeybee viral prevalence and load before and after almond blooms, it was over a coarser time scale (Alger et al., 2018; Glenny et al., 2017), and only two studies sampled during the pollination event itself (Cavigli et al., 2016; Faurot-Daniels et al., 2020). Runckel et al. (2011) found the dynamics to be sporadic cases of acute infections where hives were able to clear infections after brief periods, with BQCV being the only virus found to persist across time points as a chronic infection. (Alger et al., 2018) also saw BQCV persisting across their experiments. A study that sampled in February and then again in March showed that their hives had the lowest pathogen prevalence during the almond bloom itself, but then the total number of pathogens sharply increased after the bloom (Cavigli et al., 2016). Certain viruses were prevalent before and throughout the bloom, such as BQCV and LSV (Cavigli et al., 2016). Faurot-Daniels et al. (2020) followed a cohort of 50 individual hives before, during, and after the bloom across the span of a year. They found the highest total number of pathogens per colony a month after almond pollination. This correlated with seasonal buildups of hive population numbers and increased springtime foraging in the almond bloom and beyond. Although little is known about LSV, they found a correlation between increased LSV2 abundance in weaker colonies, suggesting it is a pathogen of concern. DWV prevalence and load were highest in the fall, LSV was higher in the spring, and BQCV and SBV were both present throughout.

Work has shown that migratory bee colonies face heightened degrees of oxidative stress levels during migratory events, which in turn could be negatively impacting the immune function of individual bees (Simone-Finstrom et al., 2016). The existing papers tracking viruses and honeybee health in the almonds all followed operations that were based outside of California. The beekeeping operation studied here was in Northern California, only about 100 km away from the almond bloom. This shorter migration distance and potential for local adaptation could mean that this operation had decreased stressors in comparison to out-of-state beekeepers. The location, along with this operation's work to breed for varroa resistance in their honeybees, suggests this is a 'best-case-scenario' instance for almond pollination. More work needs to be done on the greater viral community for larger commercial operations that have to undergo greater migration distances to reach the almonds, and therefore greater risk of potential stress.

The critical pollination services that the commercial honeybee hives in the United States provide to the world's almond crop are commercially very important. The potential risk of the emergence of infectious disease, and in particular, RNA viruses, is clearly of concern with so many bee hives from all over the country potentially interacting. Mass interactions and the spread of viruses from across the country seem to be less likely than we might have predicted. We found that the bloom was associated with a transient increase in viral diversity during and directly after the bloom. Despite bees theoretically being able to interact with hives across the country during this bloom, we see viral diversity returning to pre-bloom levels by the end of March. The early mass almond bloom provides the first major food source for the honeybees emerging from winter, while the pollination services also provide the majority of income for commercial beekeepers. High losses in the winter before the almonds bloom are economically impactful to beekeepers but may also differentially cull hives that harbor the most impactful pathogens. This suggests that the commercial benefits and early food resources that the bloom provides may outweigh population-level infectious disease risks in this local California-based operation.

# Acknowledgements

We thank Jenn Thompson for helping with sample collection and Sara Herrejon Chavez for helping with RNA extractions. Thank you to Priya Pillai for assisting with RNA preparation. The authors thank the beekeeping operation that collaborated with us during this study, along with the almond growers for providing the study sites. We also thank the bees that were sacrificed for this research. The authors acknowledge the funding sources from the National Science Foundation (NSF) Graduate Research Fellowship program and the NSF Ecology and Evolution of Infectious Disease program grant.

# Supplementary

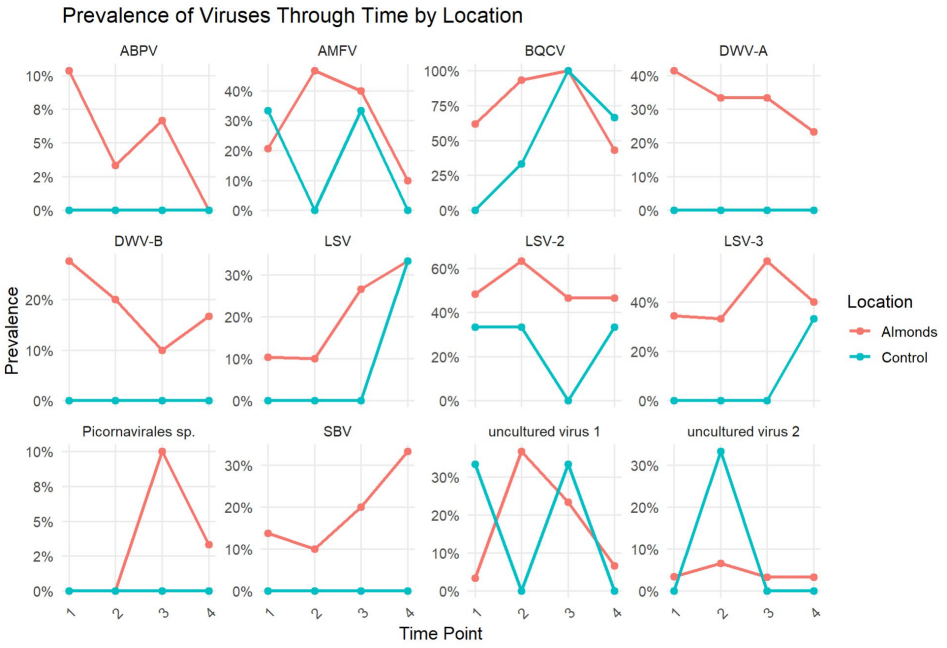


**Figure S1. Viral prevalence over time for honeybee colonies in almond pollination and control groups.** Line plots show the number of hives out of the total at each time point that were found to have each of these viruses across time in almond (pink) versus control (blue) hives to indicate prevalence as a percentage of positive cases for each virus across the apiary.

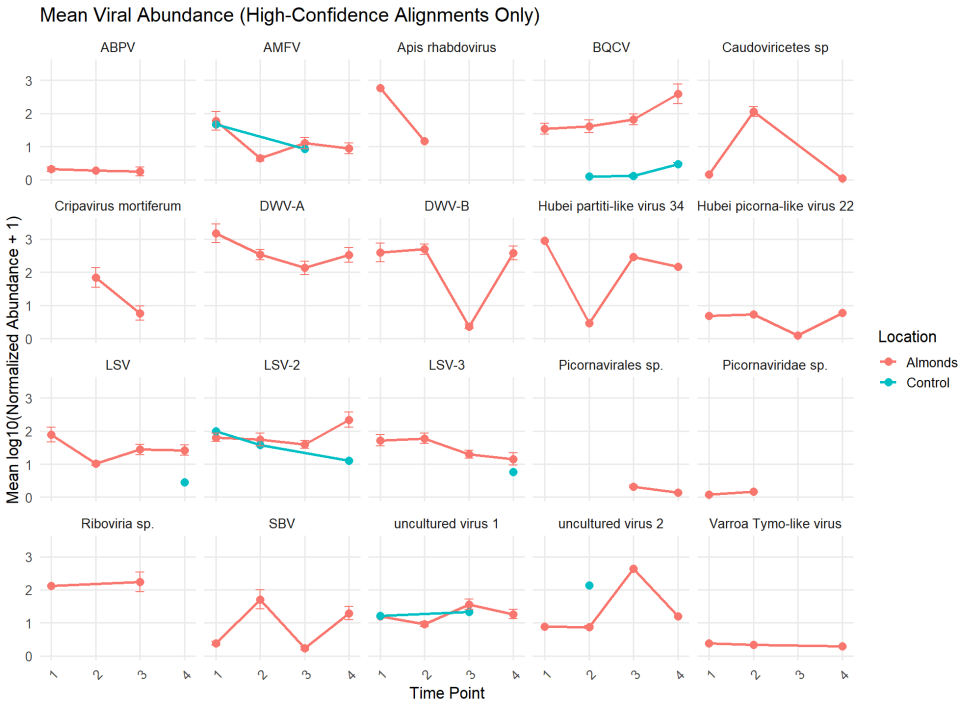


**Figure S2. Mean viral abundance over time for honeybee colonies in almond pollination and control groups.** Line plots display the mean viral abundance (log10-transformed normalized abundance) for individual viruses across time. Each panel represents a virus with temporal dynamics in mean abundance for each group.

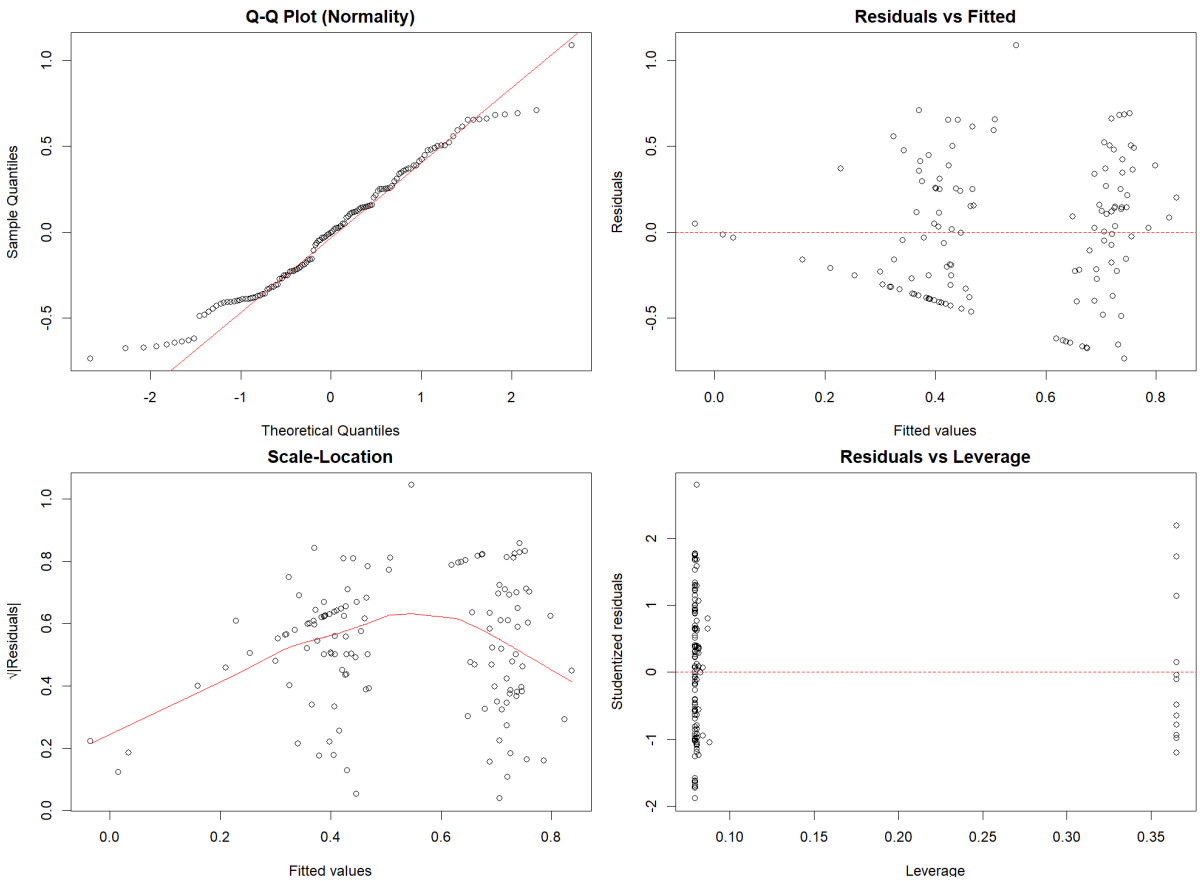


**Figure S3. Diagnostic plots for the linear mixed model (LMM) predicting Shannon diversity.** The Q–Q plot (top left) shows deviation from normality, confirmed by the Shapiro–Wilk test ($p < 0.001$). Residuals vs. fitted values (top right) and the scale–location plot (bottom left) indicate homogeneity of variance across groups, supported by Levene's test ($p = 0.51$). The residuals vs. leverage plot (bottom right) shows no strong evidence of influential observations. Together, these diagnostics suggest that while the assumption of normality is violated, variance homogeneity is satisfied, and no outliers strongly influence the model, therefore the use of the LMM is appropriate.

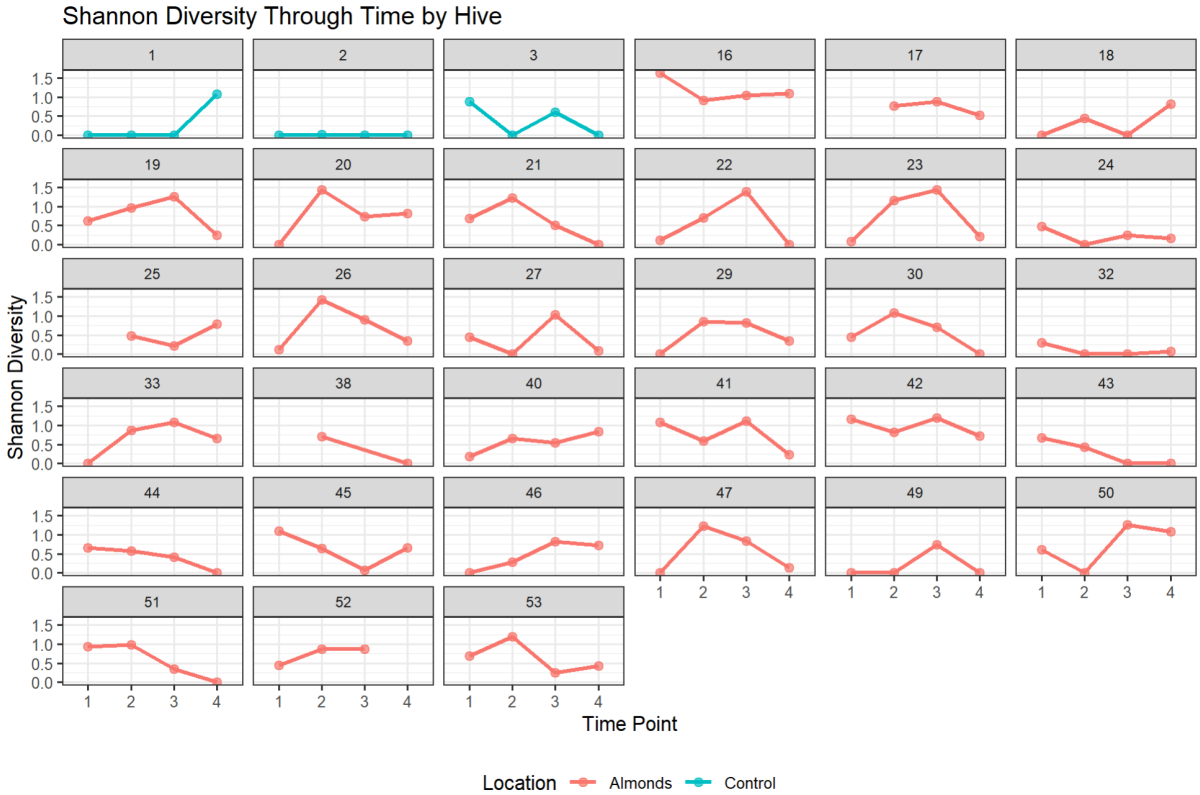


**Fig S4. Shannon diversity index for individual hives plotted through time on line plots.** Blue indicates control hives, whereas pink indicates almond hives. Only hives with viral community data for three out of four time points were plotted.

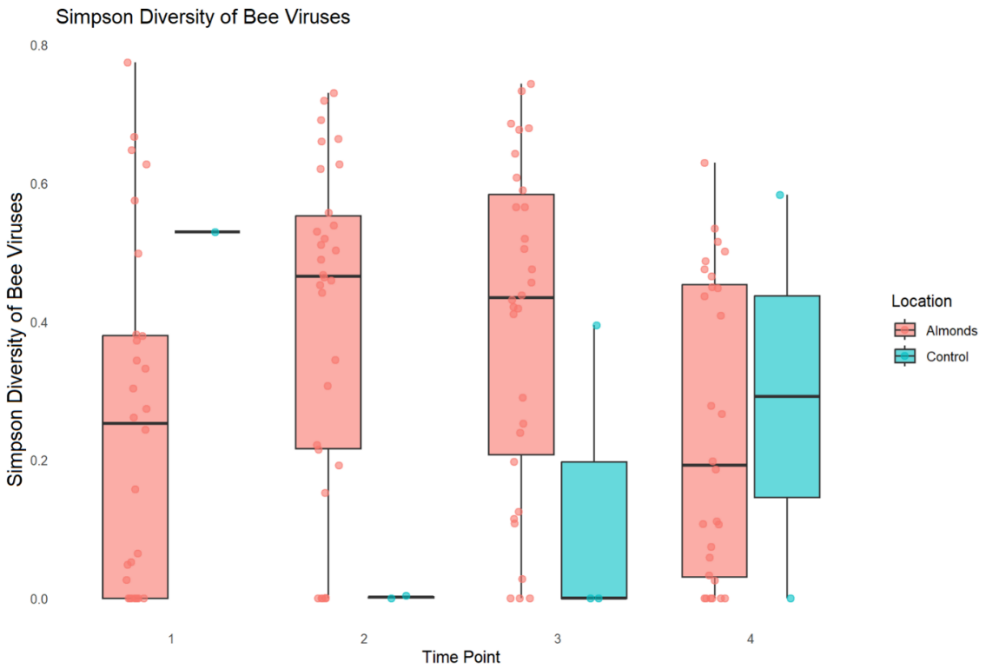


**Figure S5. Simpson diversity of bee viruses across time in almond pollinating and control honeybee colonies.** Boxplots show the normalized Simpson diversity index in almond (pink) and control hives (blue). Points indicate individual hive Simpson diversity values.

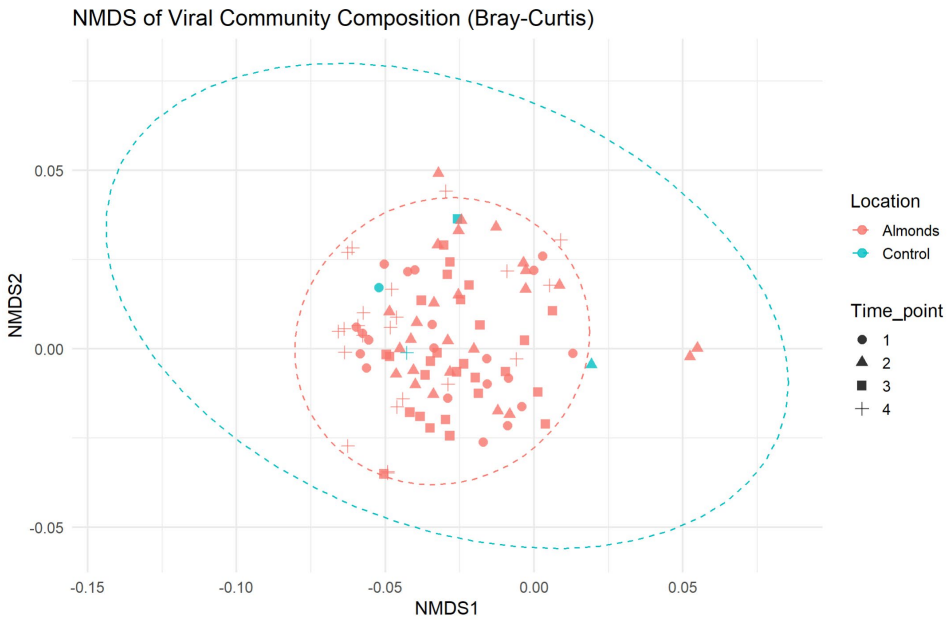


**Figure S6. Non-metric multidimensional scaling (NMDS) of viral community composition across time and location.** Each point represents a hive, colored by the location, and the shape indicates time. Viral community dissimilarities were calculated using Bray-Curtis distance on relative abundance data. Ellipses represent 95% confidence intervals around the centroid.

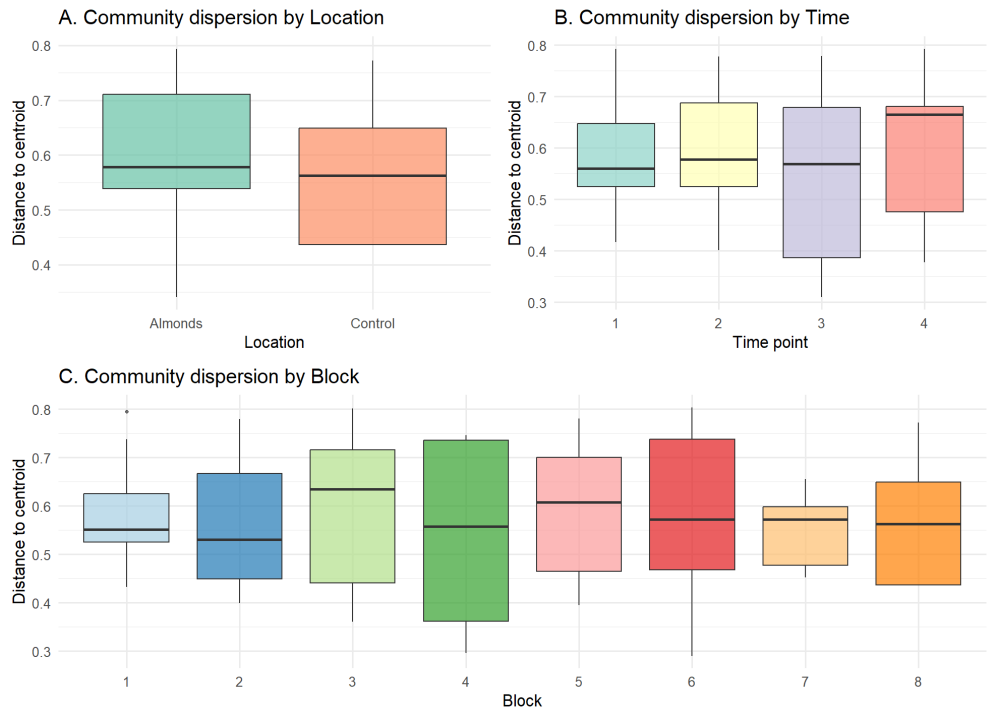


**Fig S7. Multivariate dispersion of viral communities by (A) Location, (B) Time point, and (C) Block, based on Jaccard dissimilarities.** Boxplots show the distance of each sample to its group centroid. PERMDISP tests confirmed no significant differences in dispersion among groups (Location: $p = 0.553$; Time point: $p = 0.248$; Block: $p = 0.991$), indicating comparable within-group variability across factors, and therefore the use of a PERMANOVA model is appropriate.

**Table S1**. Type III ANOVA results for a linear mixed model assessing the effect of time, location, and the interaction between time and location on Shannon diversity with hive number as a random factor.

| Factor | Degrees of freedom | Den DF | F value | P value |
|---|---|---|---|---|
| Time point | 3 | 92.08 | 0.15 | 0.927 |
| Location | 1 | 31.23 | 6.47 | 0.016 * |
| Time × Location | 3 | 92.08 | 1.66 | 0.18 |

**Table S2.** Post hoc comparisons of Shannon diversity at Almonds in comparison to controls at each time point with a Benjamini–Hochberg correction for multiple tests.

| Time point | Contrast | Estimate | SE | df | t ratio | p value |
|---|---|---|---|---|---|---|
| 1 | Almonds – Control | 0.136 | 0.254 | 122 | 0.54 | 0.593 |
| 2 | Almonds – Control | 0.704 | 0.253 | 122 | 2.78 | 0.006 ** |
| 3 | Almonds – Control | 0.522 | 0.253 | 122 | 2.06 | 0.042 * |
| 4 | Almonds – Control | 0.03 | 0.253 | 122 | 0.12 | 0.906 |

**Table S3.** Results of PERMANOVA analyses testing the effects of time point, location, and block on viral community composition using Bray–Curtis and Jaccard dissimilarities.

| Model + Factors | df | Sum of Squares | $R^2$ | F | P value |
|---|---|---|---|---|---|
| **Bray–Curtis (interaction model)** | | | | | |
| Time_point × Location + Block | 4 | 1.571 | 0.03475 | 1.0712 | 0.3564 |
| Residual | 119 | 43.638 | 0.96525 | | |
| Total | 123 | 45.209 | 1 | | |
| **Jaccard (marginal effects)** | | | | | |
| Time_point | 1 | 0.644 | 0.01754 | 2.2171 | 0.0293 * |
| Location | 1 | 0.585 | 0.01595 | 2.0156 | 0.0503 * |
| Block | 1 | 0.728 | 0.01983 | 2.5066 | 0.0180 * |
| Residual | 120 | 34.848 | 0.94941 | | |
| Total | 123 | 36.705 | 1 | | |

# References


Alger, S.A., Burnham, P.A., Brody, A.K., 2019. Flowers as viral hot spots: Honey bees (Apis mellifera) unevenly deposit viruses across plant species. PLoS ONE 14, e0221800. https://doi.org/10.1371/journal.pone.0221800

Alger, S.A., Burnham, P.A., Lamas, Z.S., Brody, A.K., Richardson, L.L., 2018. Home sick: impacts of migratory beekeeping on honey bee ( *Apis mellifera* ) pests, pathogens, and colony size. PeerJ 6, e5812. https://doi.org/10.7717/peerj.5812

Altizer, S., Bartel, R., Han, B.A., 2011. Animal Migration and Infectious Disease Risk. Science 331, 296–302. https://doi.org/10.1126/science.1194694

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

Beekman, M., Ratnieks, F.L.W., 2000. Long-range foraging by the honey-bee, *Apis mellifera* L. Functional Ecology 14, 490–496. https://doi.org/10.1046/j.1365-2435.2000.00443.x

Boots, M., Sasaki, A., 1999. 'Small worlds' and the evolution of virulence: infection occurs locally and at a distance. Proc. R. Soc. Lond. B 266, 1933–1938. https://doi.org/10.1098/rspb.1999.0869

Bordier, C., Pioz, M., Crauser, D., Le Conte, Y., Alaux, C., 2017. Should I stay or should I go: honeybee drifting behaviour as a function of parasitism. Apidologie 48, 286–297. https://doi.org/10.1007/s13592-016-0475-1

Burnham, P.A., Alger, S.A., Case, B., Boncristiani, H., Hébert-Dufresne, L., Brody, A.K., 2021. Flowers as dirty doorknobs: Deformed wing virus transmitted between *Apis mellifera* and *Bombus impatiens* through shared flowers. Journal of Applied Ecology 58, 2065–2074. https://doi.org/10.1111/1365-2664.13962

Carman, H., 2011. The Estimated Impact of Bee Colony Collapse Disorder on Almond Pollination Fees. University of California Giannini Foundation of Agricultural Economics 14, 9–11.

Cavigli, I., Daughenbaugh, K.F., Martin, M., Lerch, M., Banner, K., Garcia, E., Brutscher, L.M., Flenniken, M.L., 2016. Pathogen prevalence and abundance in honey bee colonies involved in almond pollination. Apidologie 47, 251–266. https://doi.org/10.1007/s13592-015-0395-5

Chen, Y., Evans, J., Feldlaufer, M., 2006. Horizontal and vertical transmission of viruses in the honey bee, Apis mellifera. Journal of Invertebrate Pathology 92, 152–159. https://doi.org/10.1016/j.jip.2006.03.010

Cohen, H., Smith, G.P., Sardiñas, H., Zorn, J.F., McFrederick, Q.S., Woodard, S.H., Ponisio, L.C., 2021. Mass-flowering monoculture attracts bees, amplifying parasite prevalence. Proc. R. Soc. B. 288, 20211369. https://doi.org/10.1098/rspb.2021.1369

Cornman, R.S., Schatz, M.C., Johnston, J.S., Chen, Y.-P., Pettis, J., Hunt, G., Bourgeois, L., Elsik, C., Anderson, D., Grozinger, C.M., Evans, J.D., 2010. Genomic survey of the ectoparasitic mite Varroa destructor, a major pest of the honey bee Apis mellifera. BMC Genomics 11, 602. https://doi.org/10.1186/1471-2164-11-602

Cornman, R.S., Tarpy, D.R., Chen, Y., Jeffreys, L., Lopez, D., Pettis, J.S., vanEngelsdorp, D., Evans, J.D., 2012. Pathogen Webs in Collapsing Honey Bee Colonies. PLoS ONE 7, e43562. https://doi.org/10.1371/journal.pone.0043562

Doublet, V., Doyle, T.D., Carvell, C., Brown, M.J.F., Wilfert, L., 2025. Host ecology and phylogeny shape the temporal dynamics of social bee viromes. Nat Commun 16, 2207. https://doi.org/10.1038/s41467-025-57314-7

Faurot-Daniels, C., Glenny, W., Daughenbaugh, K.F., McMenamin, A.J., Burkle, L.A., Flenniken, M.L., 2020. Longitudinal monitoring of honey bee colonies reveals dynamic nature of virus abundance and indicates a negative impact of Lake Sinai virus 2 on colony health. PLoS ONE 15, e0237544. https://doi.org/10.1371/journal.pone.0237544

Figueroa, L.L., Grab, H., Ng, W.H., Myers, C.R., Graystock, P., McFrederick, Q.S., McArt, S.H., 2020. Landscape simplification shapes pathogen prevalence in plant-pollinator networks. Ecology Letters 23, 1212–1222. https://doi.org/10.1111/ele.13521

Fries, I., Camazine, S., 2001. Implications of horizontal and vertical pathogen transmission for honey bee epidemiology. Apidologie 32, 199–214. https://doi.org/10.1051/apido:2001122

Geffre, A.C., Gernat, T., Harwood, G.P., Jones, B.M., Morselli Gysi, D., Hamilton, A.R., Bonning, B.C., Toth, A.L., Robinson, G.E., Dolezal, A.G., 2020. Honey bee virus causes context-dependent changes in host social behavior. Proc. Natl. Acad. Sci. U.S.A. 117, 10406–10413. https://doi.org/10.1073/pnas.2002268117

Glenny, W., Cavigli, I., Daughenbaugh, K.F., Radford, R., Kegley, S.E., Flenniken, M.L., 2017. Honey bee (Apis mellifera) colony health and pathogen composition in migratory beekeeping operations involved in California almond pollination. PLoS ONE 12, e0182814. https://doi.org/10.1371/journal.pone.0182814

Iqbal, J., Mueller, U., 2007. Virus infection causes specific learning deficits in honeybee foragers. Proc. R. Soc. B. 274, 1517–1521. https://doi.org/10.1098/rspb.2007.0022

Klein, A., Brittain, C., Hendrix, S.D., Thorp, R., Williams, N., Kremen, C., 2012. Wild pollination services to California almond rely on semi-natural habitat. Journal of Applied Ecology 49, 723–732. https://doi.org/10.1111/j.1365-2664.2012.02144.x

Kolde, R., 2018. pheatmap: Pretty Heatmaps. R package version 1.0.12.

Kralj, J., Fuchs, S., 2006. Parasitic *Varroa destructor* mites influence flight duration and homing ability of infested *Apis mellifera* foragers. Apidologie 37, 577–587. https://doi.org/10.1051/apido:2006040

Kulhanek, K., Garavito, A., vanEngelsdorp, D., 2021. Accelerated Varroa destructor population growth in honey bee (Apis mellifera) colonies is associated with visitation from non-natal bees. Sci Rep 11, 7092. https://doi.org/10.1038/s41598-021-86558-8

Kwon, M., Jung, C., Kil, E.-J., 2023. Metagenomic analysis of viromes in honey bee colonies (Apis mellifera; Hymenoptera: Apidae) after mass disappearance in Korea. Front. Cell. Infect. Microbiol. 13, 1124596. https://doi.org/10.3389/fcimb.2023.1124596

Kwon, M., Nilufar Zakhidovna, R., Boymaxmat Abdiazizovich, K., Jung, C., Kil, E.-J., 2024. First metagenomic analysis of virome in Uzbekistan honey bee (Apis mellifera): Investigating basic information on honey bee viruses. Journal of Invertebrate Pathology 206, 108171. https://doi.org/10.1016/j.jip.2024.108171

Magurran, A.E., McGill, B.J., 2011. Biological diversity: frontiers in measurement and assessment. Oxford Univ. Press.

Manley, R., Doublet, V., Wright, O.N., Doyle, T., Refoy, I., Hedges, S., Pascall, D., Carvell, C., Brown, M.J.F., Wilfert, L., 2023. Conservation measures or hotspots of disease transmission? Agri-environment schemes can reduce disease prevalence in pollinator communities. Phil. Trans. R. Soc. B 378, 20220004. https://doi.org/10.1098/rstb.2022.0004

McKeown, D.A., Evans, E., Helgen, J., Warner, J., Zimmern, R., Masterman, R., Berrington, A., Nemecek, M., Costello, C., Bernstein, E., Hesketh-Best, P.J., Mendel, B., Spivak,

M., Schroeder, D.C., 2025. Distinct virome compositions and lack of viral diversification indicate that viral spillover is a dead-end between the western honey bee and the common eastern bumblebee. Commun Biol 8, 926. https://doi.org/10.1038/s42003-025-08351-x

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, 615–624. https://doi.org/10.1111/1365-2656.12345

Neumann, P., Moritz, R.F.A., Mautz, D., 2000. Colony evaluation is not affected by drifting of drone and worker honeybees ( *Apis mellifera* L.) at a performance testing apiary. Apidologie 31, 67–79. https://doi.org/10.1051/apido:2000107

Pfeiffer, K.J., Crailsheim, K., 1998. Drifting of honeybees. Insectes Sociaux 45, 151–167. https://doi.org/10.1007/s000400050076

Piot, N., Eeraerts, M., Pisman, M., Claus, G., Meeus, I., Smagghe, G., 2021. More is less: mass-flowering fruit tree crops dilute parasite transmission between bees. International Journal for Parasitology 51, 777–785. https://doi.org/10.1016/j.ijpara.2021.02.002

Piot, N., Meeus, I., Kleijn, D., Scheper, J., Linders, T., Smagghe, G., 2019. Establishment of wildflower fields in poor quality landscapes enhances micro-parasite prevalence in wild bumble bees. Oecologia 189, 149–158. https://doi.org/10.1007/s00442-018-4296-y

Seeley, T.D., Smith, M.L., 2015. Crowding honeybee colonies in apiaries can increase their vulnerability to the deadly ectoparasite Varroa destructor. Apidologie 46, 716–727. https://doi.org/10.1007/s13592-015-0361-2

Tehel, A., Brown, M.J., Paxton, R.J., 2016. Impact of managed honey bee viruses on wild bees. Current Opinion in Virology 19, 16–22. https://doi.org/10.1016/j.coviro.2016.06.006

vanEngelsdorp, D., Meixner, M.D., 2010. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. Journal of Invertebrate Pathology 103, S80–S95. https://doi.org/10.1016/j.jip.2009.06.011

# Data Accessibility and Benefit-Sharing

**Data Accessibility Statement**

All sequence reads generated in this study will be uploaded onto the NCBI Sequence Read Archive (SRA) database upon publication. Sample metadata will also be available in the SRA submission. The assembled viral genomic sequences will be deposited in GenBank.

**Benefit-Sharing Statement**

The Nagoya Protocol does not apply to this study, as all honeybee samples were collected in the United States. However, benefits were still generated, including the sharing of these results with the beekeeping community, such as at beekeeping association meetings and conferences.

# Author Contributions

NS and MB designed and performed the research. NS conducted the laboratory work, analyzed the data, and wrote the manuscript GRN assisted with the analysis and statistical testing. LW assisted with analysis and manuscript editing.