



OPEN Metagenomic and gene expression patterns in declining commercial honey bee colonies

Anthony Nearman¹✉, Zachary S. Lamas^{1,2}, Elina L. Niño³, Julia Fine⁴, Christopher Mayack⁵, Arathi Seshadri⁶, Dawn Boncristiani¹, Wei-Fone Huang⁷, Jay D. Evans¹ & Yan Ping Chen¹

Managed honey bee colonies (*Apis mellifera*) in the US continue to experience high overwinter loss rates driven by parasites, pathogens, poor nutrition, and pesticides. To mitigate these losses, inspection and monitoring are critical for identifying traits of colonies in decline and potential causal factors. In this study, we apply molecular methods to associate potential causative agents with colonies in various stages of decline. Initially, we investigated in-hive bee metagenomic RNA isolated from 15 colonies across seven managed operations in California whose adult bee and brood populations were classified as Strong, Medium, or Weak in strength. We discovered that Weak colonies harbored 2.2- and 3.6- fold more viral species than Medium and Strong colonies, respectively, as well as larger viral read pools despite similar library sizes. They also displayed higher nucleotide variation in *Varroa*-vectored viruses, indicating associations with high mite populations. When investigating differences in host gene expression, we discovered an upregulation of immune-related pathways in Weak colonies relative to Strong. Specifically, Weak colonies upregulated genes related to wound healing, phagocytosis, oxidative stress resistance, apoptosis, and RNA interference. Most antimicrobial peptides were upregulated in Weak colonies, although *defensin1* was significantly higher in Strong colonies, along with several detoxification enzymes and the royal jelly peptide *apisimin*. Weak colonies also showed an upregulation of transcripts tied to abnormal protein digestion. The low levels of viral replication and fewer species of mite-vectored viruses in Strong colonies may be due to successful *Varroa* management. Strong colonies also displayed upregulated levels of nine different ubiquinone transcripts, arguably reflecting increasing longevity or a younger in-hive population compared to Weak colonies. Overall, these results provide a detailed account of viral metagenomics and associated host responses, providing new insights into the mechanisms underlying honey bee colony decline under comparable management conditions.

Keywords Agriculture, Gene expression, Metagenomics, Pollination, RNA-sequencing, Virus

Managed honey bees in the US continue to experience high rates of colony loss^{1–3}, contributing to economic concern among stakeholders^{4,5}. Researchers are focused on investigating the underlying drivers of loss and developing approaches to monitor colony health in real time. While existing health management measures have contributed to reducing losses⁶, additional diagnostic tools are needed to further improve these results. Further, continued monitoring remains essential to detect emerging pathogens and more virulent strains of existing pathogens⁷.

Metagenomic analysis represents a critical first step in developing new diagnostic tools^{8–10}. This approach facilitates the discovery and characterization of novel pathogens^{11,12}, reveals the combinations of viruses and parasites present^{8,9}, and maps sequence divergence among viral quasi-species¹³. With improved sequencing technologies¹⁴, genomic¹⁵ and hologenomic (host plus metagenome)^{16–18} resources, researchers are better positioned to reveal causal factors, a key step toward understanding and potentially mitigating colony losses.

¹USDA-ARS Bee Research Lab, BARC-East Bldg. 306, Beltsville, MD 20705, USA. ²Department of Biology, University of Maryland, Baltimore County, Baltimore, MD, USA. ³Department of Entomology and Nematology, University of California Davis, One Shields Avenue, Davis, CA 95616, USA. ⁴Invasive Species and Pollinator Health Research Unit, USDA-ARS, 3026 Bee Biology Rd, Davis, CA 95616, USA. ⁵Department of Biology, William Paterson University, Wayne, NJ 07470, USA. ⁶USDA-ARS, Pollinator Health in Southern Crops Ecosystems Research Unit, Stoneville, MS 38776, USA. ⁷School of Agriculture and Natural Resources, Kentucky State University, 400 E. Main St, Frankfort, KY 40601, USA. ✉email: anthony.j.nearman@gmail.com

Another promising approach to improving colony health diagnostics involves analyzing the host's RNA (transcriptome)^{19,20}. As viral titers alone are not always predictive of colony loss^{21,22}, drawing associations between combinations of specific viruses and changes to a host's transcriptome may reveal gene targets related to tolerance, resistance, or susceptibility to infection. Incorporating such targets into existing rapid diagnostics of viruses could increase predictive power of colony health assessment. This could include, but not limited to, genes specific to immune function.

In this study, we assess viral and transcriptomic differences across honey bee colonies classified in the field as Weak, Medium, or Strong based on population strength. To do so, 15 colony-level samples from seven different commercially managed beekeeping operations were collected and RNA sequencing was performed. The resulting libraries were aligned to host and known cellular parasite genomes and quantified. Remaining sequenced reads were assembled into contigs and identified. Accession sequences of identified viral contigs were downloaded, and libraries were aligned to measure diversity and quantify read counts. The results indicate combinations of viruses and host transcript expression that are associated with colony strength. This study details key biological differences among colonies in varying stages of decline.

Results

Colonies observed to have Strong populations presented fewer viruses, viral transcripts, and *Nosema ceranae* transcripts compared to Weak and Medium Strength colonies. A declining trend was observed for the within-library proportion of mapped virus reads for Weak ($1.7 \times 10^{-3} \pm 4.3 \times 10^{-4}$), Medium ($9.3 \times 10^{-4} \pm 3.7 \times 10^{-4}$), and Strong colonies ($9.7 \times 10^{-5} \pm 5.0 \times 10^{-5}$) (Fig. 1). Accounting for the number of unique viruses with > 25% genome coverage revealed the same trend, with Weak colonies containing an average of 10.8 ± 0.9 unique viral alignments, Medium colonies 6.8 ± 1.2 , and Strong colonies 4.6 ± 1.3 (Fig. 1). Among the 22 unique viral accessions detected across all colonies, the odds of detection among Weak colonies were 3.6 times (95%CI = 1.8–7.4, z-value = 4.3, $p < 0.001$) higher compared to Strong colonies on average, and 2.2 times more (95%CI = 1.1–4.2, z-value = 2.7, $p < 0.02$) compared to Medium colonies. No significant difference was observed between Medium and Strong colonies in the number of viruses. Transcripts mapping to the *Nosema (Vairimorpha) ceranae* genome displayed a similar declining pattern across colony strengths (Fig. 1). The mean number of reads per library did not differ substantially among colony strength groups, with values falling within one standard error of each other (Fig. S1).

On a per-virus per-library basis, sequencing results indicate trends with colony population strength. Among the number of polymorphic sites (variant calls with > 1% read proportion), Weak colonies displayed an average of 252.4 ± 39.6 , Medium colonies 225.7 ± 37.7 , and Strong colonies 135.8 ± 29.6 per viral kilobase per library (Fig. 2a–b). The proportion of reads containing polymorphisms was lower in Strong colonies (Weak = 0.134 ± 0.022 , Medium = 0.121 ± 0.025 , Strong = 0.034 ± 0.005 , Fig. 2a and c). Mean sequencing depth per base (Weak = 927.9 ± 409.5 , Medium = 952.1 ± 332.0 , Strong = 345.8 ± 213.5) and the mean proportional length of coverage (Weak = 0.894 ± 0.024 , Medium = 0.893 ± 0.027 , Strong = 0.764 ± 0.037) also followed similar trends (Fig. 2d–f). Specific viruses over-represented in weak colonies include two variants of Deformed wing virus, Israeli acute paralysis virus, and *Apis mellifera* filamentous virus (Fig. 2, Fig. S2). A standard panel of viruses, including multiple Lake Sinai variants, indicate differences in viral titers for Black Queen Cell virus and three Lake Sinai variants, with most viruses present in all strength categories except Israeli acute paralysis virus and Kashmir bee virus (Fig. S3, Table S1).

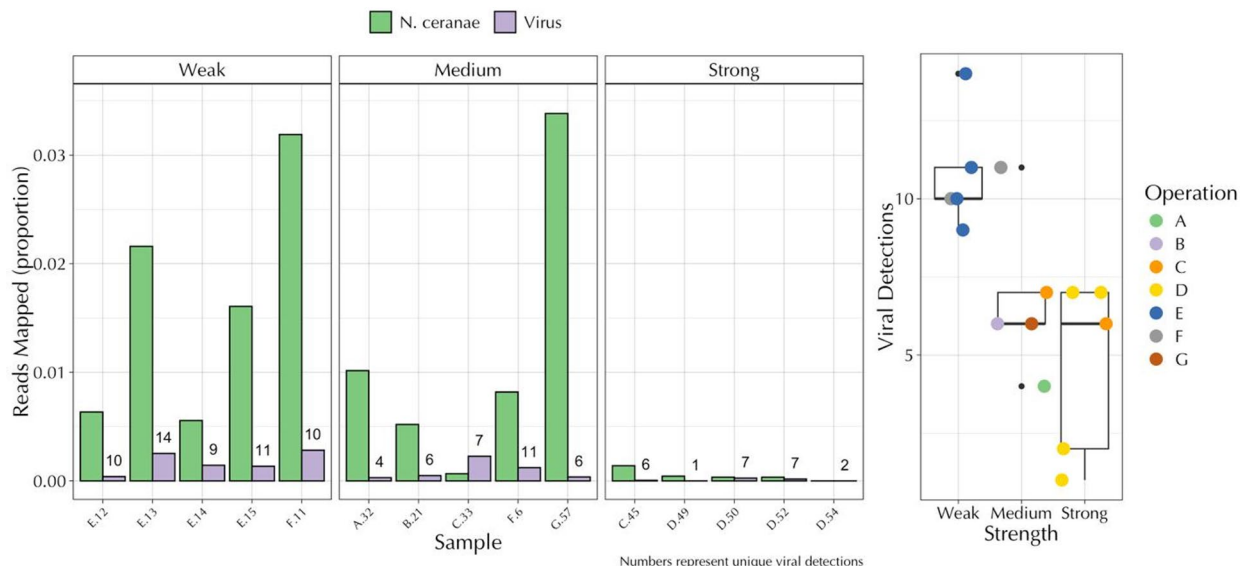


Fig. 1. Proportion of sequenced reads mapped to viruses and *Nosema ceranae* by colony strength. Numbers and box plot represent unique viral detections. Viral detections with less than 25% genome coverage were removed when counting the number of viruses. Colonies are organized by their adult bee and brood frame population strength (Weak, Medium, Strong) and source beekeeping operation (A: G).

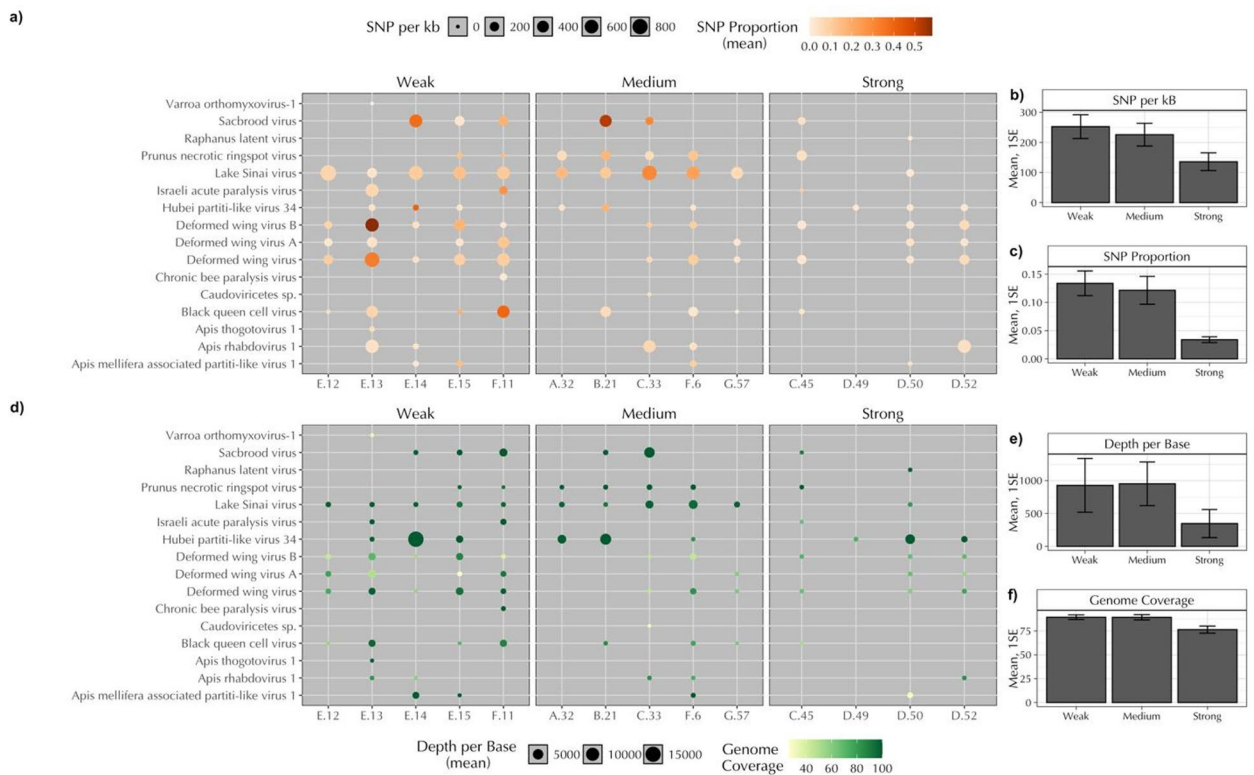


Fig. 2. Measures of viral presence in a host. Single Nucleotide Polymorphism (SNP) count per kilobase of viral genome and their proportional representation in the aligned read pool (a); Mean and 1SE of overall SNPs per kilobase (b); Mean and 1SE proportional SNP representation in the read pool (c); Mean alignment depth per base and proportion of genome coverage for aligned viruses (d); Mean and 1SE read depth per virus per library (e); and Mean and 1SE genome coverage per virus per library (f). Detections with less than 25% genome coverage for all libraries were removed. *Apis mellifera* filamentous virus (AmFV) was excluded from SNP analysis due to extreme relative counts. Colonies are organized by their relative population strength (Weak, Medium, Strong) and source beekeeping operation (A: G). Numbers represent unique colony IDs.

Clustering colonies and their associated viruses by β -actin normalized read counts indicates relationships to colony strength. Hierarchical clustering reveals three distinct clades of colonies and two distinct clades of viruses. The first and third clades of colonies, consisting of all Weak, Medium, and 1 Strong colony, are similar in terms of co-infection levels that span both clades of viruses. One clade with striking differences between Strong colonies and others contains three variants of Deformed wing virus as well as *Apis mellifera* filamentous virus. The second clade of colonies, consisting of all Strong colonies, showed few viruses overall (Fig. S2).

Lake Sinai viruses (LSV) assembled contigs reveal associations to multiple strains, particularly among Medium strength colonies. To better understand the variation and existing evidence regarding LSV, reads that aligned to any known variant were filtered from their libraries and assembled into contigs. RNA-dependent RNA polymerase genes were then identified and extracted from contigs > 5 kb and aligned to closely related NCBI Genbank accessions. The resulting phylogenetic tree reveals three major clades between assembled contigs and previously published sequences. Six of the eleven assembled RdRp genes are nearly identical and closely related to LSV-4, while the remaining contigs indicate close relationships to LSV, LSV-1, LSV-2, LSV-3, LSV-NE, and LSV-TO. Four of the contigs that diverge from the LSV-4 clade are additional contigs from those same colonies, indicating a multi-strain infection or rapid mutation within host colonies (Fig. S4).

Differential gene expression results suggest relationships to colony population strength. Using an adjusted p-value threshold of <0.01, comparisons between Weak and Strong colonies identified 776 differentially expressed transcripts, including 386 upregulated and 390 downregulated. Hierarchical clustering of colonies by their gene expression profiles indicates distinct separation by colony population strength (Fig. S5). Further, principal component analysis of the top 100 differentially expressed transcripts also reveals this separation, with 40% of the variation explained along the primary component axis (Fig. S6).

Immune, GO, and KEGG pathway analyses support disease-related trends related to colony strength. Isolating genes by their associated immune gene pathways reveal multiple targets for identifying colonies in various stages of decline (Fig. 3, Fig. S7). Several additional genes of interest displayed patterns related to colony strength. This includes nine transcripts for Coenzyme Q10 (ubiquinone) components (Fig. S8), three heat shock proteins (Fig. S9), six Cytochrome P450 genes (Fig. S10), and one major royal jelly protein (Fig. S11). Further, KEGG gene set enrichment and over-representation analysis suggest upregulated pathways within Weak colonies relative to Strong associated with increases in immune responses such as wound healing, phagocytosis, oxidative stress

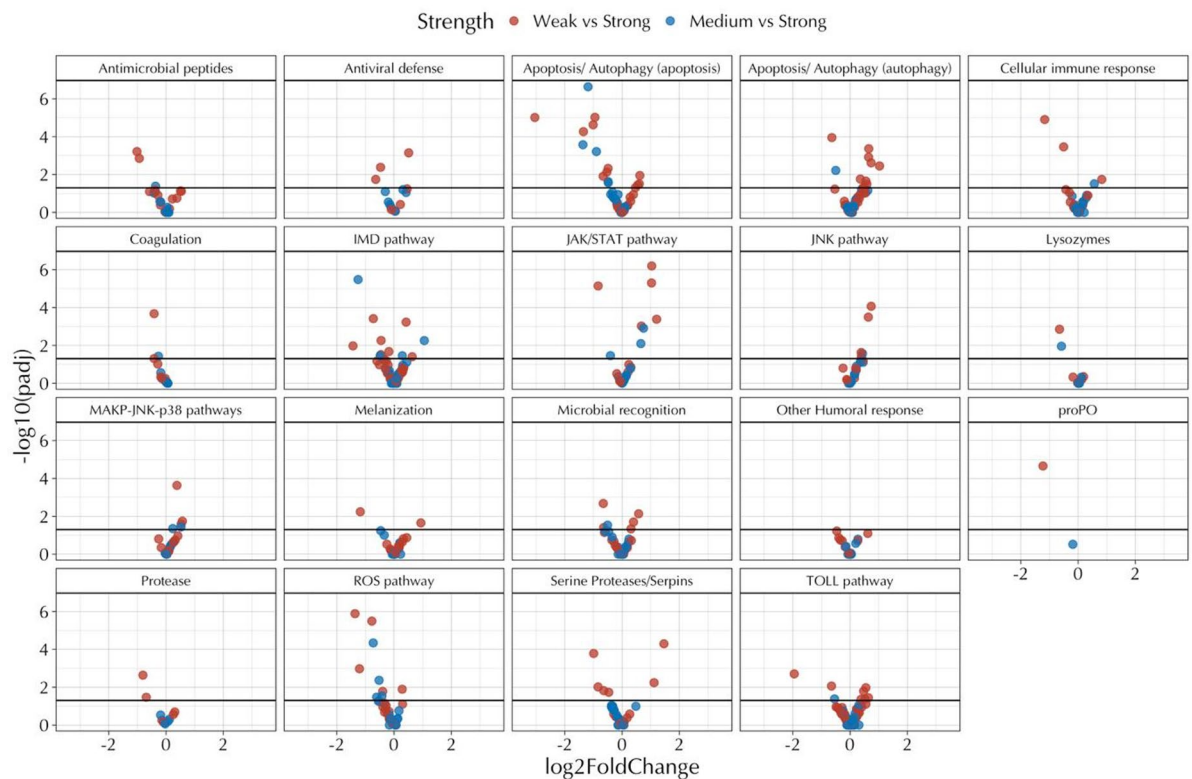


Fig. 3. Volcano plot of differentially expressed genes curated into immune pathways. The black line represents a p-value of $-\log_{10}(0.05)$. Results are displayed for pathways where at least one differentially expressed gene was present.

resistance, and apoptosis (Fig. S12–S13). These results are further confirmed through Gene Ontology gene set enrichment (Fig. S14–S16).

Discussion

Since the outbreak of Colony Collapse Disorder (CCD) in 2006, metagenomic approaches have been used to associate causative agents with colony loss^{8,22–26}. These studies tend to show an increase in viral diversity for colonies in poor health^{8,9,22,27–29}. Here, we show that colonies identified as being Weak³⁰ show a greater number of unique viral species and higher viral nucleotide diversity within those species (Fig. 1). Specifically, Weak colonies were more likely to contain viruses compared to Medium and Strong colonies, while Weak and Medium colonies showed higher nucleotide diversity and sequencing depth than Strong colonies (Figs. 1 and 2).

In this study, we demonstrate relationships between honey bee colony population strength and viral communities using a metagenomic approach. Weak and Medium strength colonies had more diverse viral communities compared to Strong colonies when their RNA was aligned to *de novo* assembled honey bee viruses (Fig. 1, Fig. S2). Similar relationships were observed when quantifying the overall number of polymorphic sites, sequencing depth, and genome coverage among identified viruses (Fig. 2). The prominent virus in all Medium strength colonies was Lake Sinai virus (LSV). Phylogenetic analysis of RdRp genes from LSV assembled contigs and closely related GenBank sequences indicates infection with multiple LSV strains or rapid within-host divergence, with most contigs being derived from Medium Strength colonies (Fig. S3). Last, a host differential gene expression analysis revealed relationships to colony strength that contained differences across multiple curated pathways for response to disease and detoxification (Fig. 3, Fig. S7–S16).

Viruses vectored by *V. destructor* (i.e., DWV-A, DWV-B, and IAPV) were detected at higher prevalence and reproduced at higher levels in Weak colonies compared to Medium or Strong colonies (Fig. 2, Fig. S2). This pattern likely reflects heavier mite infestation in Weak colonies. Mite stress and viral infection weaken immune responses in honey bees^{31,32}, which may aid in secondary infections through non-vectored transmission routes^{33–35}. In addition, prior studies on colony strength demonstrate that bees from Strong colonies better resist DWV infection and replication compared to Weak colonies³⁶. The link between mite stress, vectored viruses and secondary infection is further supported by our finding that Weak colonies showed 2.2 and 3.6 times the number of viral species on average compared to Medium and Strong colonies, respectively. Weak colonies also showed higher levels of *Apis mellifera* filamentous virus and greater proportions of transcripts mapped to *N. ceranae* (Figs. 1 and 2, Fig. S2).

Our analysis of differentially expressed transcripts separates colonies based on colony strength. Significant pathways and many individual genes in Weak colonies agree with controlled studies that describe upregulated

immune responses when faced with viral infection^{37–41}. Genes in the antimicrobial peptide family, on the other hand, were upregulated in Strong colonies (Fig. 3). Both *apisimin* and *defensin1* are upregulated in the presence of certain neonicotinoids⁴², which coincides with the upregulation of detoxification enzymes⁴³ (Fig. S9). *Apisimin* was previously detected in royal jelly⁴⁴, and at least one major royal jelly protein (*mjrp5*) was upregulated in Strong colonies (Fig. S11). Also upregulated in Strong colonies are the family of ubiquinone enzymes (Fig. S8). Artificial treatment with these proteins has been demonstrated to increase longevity for bees in controlled studies⁴⁵. Further, ubiquinone transcript abundance decreases with chronological age and task performance⁴⁶, suggesting these results reflect either a younger cohort of in-hive bees or a lack of precocious foragers in Strong colonies. Precocious foragers have been observed in bees infected with *N. ceranae*⁴⁷ and Deformed wing virus⁴⁸, which we observed at higher levels and greater prevalence among Weak and Medium strength colonies.

Differential transcript abundance of specific immune pathway members with respect to colony strength may be used to predict colonies at risk of decline without the need for extensive differential gene expression analysis. Alongside qPCR quantification of a standard panel of viruses, these gene targets may improve the accuracy of future colony health diagnostic efforts and indicate where preventative measures can be used to save colonies.

Our analysis reveals an association between field measures of colony strength, viruses, and host gene expression. Differences in viral populations can be seen across all levels of colony strength. Weak and Medium strength colonies also displayed upregulated immune pathways and genes relative to Strong colonies. Strong colonies, however, had fewer viruses and lower viral replication overall, coupled with upregulation of several immune-related genes as well as genes for detoxification enzymes. Despite earlier suggestions, these upregulated responses may be the signs of a successful immune response and/or exposure to certain acaricides⁴⁹, which agrees with the assertion of variation in successful mite control among all colonies. This work provides detailed insights into the viral dynamics that can occur within honey bee populations. It also indicates the need for improved accuracy of health diagnostic techniques beyond a standard panel of viruses, as the number of detections did not differentiate between all levels of decline. Monitoring virus levels and transcript abundances for certain indicators of host health can be a useful tool for enhancing diagnostic accuracy and serve as valuable indicators of host health, supporting efforts to prevent disease and colony loss.

Methods

Honey bees

In February 2023, a field investigation was conducted in California in response to reports from commercial beekeepers regarding significant colony losses. Colonies were field inspected by researchers and categorized as Strong, Medium, or Weak based on two main criteria: the size of the adult bee population and the number of brood frames within each colony³⁰. A brood frame was considered “full” if more than half of it was covered with brood. Additionally, the presence of the queen was confirmed for each colony to ensure that the colony’s scores did not reflect queen failure (Supplemental Table 1).

For each colony, a brood frame covered with adult bees was carefully removed during sampling. Adult bees were gently shaken into an alcohol-cleaned plastic pan to minimize contamination and guided into one corner of the pan to facilitate collection. Bees were then collected using clean 50 mL Falcon tubes, filling each tube completely. Two tubes were collected per colony, with approximately 100–120 worker bees per tube. Collected samples were immediately placed into a container with dry ice. Once all the samples were collected, they were shipped back to the laboratory under dry ice conditions to ensure sample integrity during transit. Upon arrival at the lab, the samples were immediately transferred to a -80°C freezer for later analysis. A graphical abstract depicting the overall experimental design is included in the supplemental figures (Fig. S17).

Virus and mRNA enrichment from collected worker bees

Frozen worker bees were homogenized in liquid nitrogen with ceramic mortars and pestles to yield a fine powder. A total of 0.5 g of ground tissue was transferred into a 2.0 mL microcentrifuge tube containing 750 μL of $0.2\times$ PBS and homogenized using a FastPrep system (Fisher Science) for 40 s. For each sample, four tubes (totaling 2.0 g) were prepared, and the homogenates were pooled into a 15 mL centrifuge tube.

The pooled bee homogenate was centrifuged at $3,000\times g$ for 10 min, and the supernatant was filtered through a $0.8\text{ }\mu\text{m}$ syringe filter to remove debris and large particles. The filtered bee homogenate was concentrated using a 30 kDa ultrafiltration tube (Amicon, Millipore) by centrifugation at $20,000\times g$ for 20 min. The filtrate ($<30\text{ kDa}$ fraction) was discarded, and additional filtered bee homogenate was added to the ultrafiltration tube. This concentration step was repeated until the retentate reached approximately 20-fold concentration.

The ultrafiltration tube was then inverted into a clean centrifuge tube, and the concentrated bee homogenate was recovered by centrifugation at $6,500\times g$ for 5 min. RNA was extracted from the concentrated bee homogenate using the QIAamp Viral RNA Mini Kit (Qiagen) according to the manufacturer’s protocol.

Library preparation and sequencing

The integrity and quantity of the RNA were assessed with the RNA Nano 6000 Assay Kit on the Bioanalyzer 2100 system (Agilent Technologies). RNA samples were submitted to the University of Maryland Institute for Genome Sciences for whole-metagenome shotgun sequencing. Following DNase I digestion (Qiagen RNase-Free DNase, 1 μL of DNase I to 9 μL of RNA, incubated at 37°C for 30 min, followed by enzyme inactivation) and ribosomal RNA depletion (Illumina Ribo-Zero Plus Microbiome rRNA Depletion Kit), cDNA was synthesized using random hexamer primers and reverse transcriptase (SuperScript™ II Reverse Transcriptase, Thermo Fisher Scientific). Illumina RNA-seq libraries were prepared with the TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA) without the poly-A isolation steps. Adapters were ligated to the double-stranded cDNA, which was then purified, and library size selection was performed using AMPure XT beads (Beckman Coulter Genomics,

Danvers, MA). Paired-end (100 bp) sequencing of each RNA library was conducted on the Illumina HiSeq 2500 platform.

Metagenomic analysis

Short read libraries were analyzed using *FastQC* v0.12.1⁵⁰ and low-quality reads and bases were removed with *trimmomatic* v0.39⁵¹. Reads from each library were aligned to the Amel HAv3.1 (GCA_003254395.2), *N. cerenae* (GCA_000988165.1), and *L. passim* (GCA_037349495.1) genomes, with aligned reads being quantified and removed at each step to generate a set of host and parasite cleaned libraries using *STAR* short read aligner v2.7.11b⁵². Remaining reads were normalized to a target depth of 100 using *bbnorm* from BBMap suite v39.17⁵³. Normalized reads were then assembled using *SPAdes* v4.0 with the `--isolate` parameter⁵⁴. The resulting contigs were filtered to a length > 500 bases and identified with the NCBI non-redundant nucleotide database using *blastn* 2.16.0 and an e-value of 1e-5⁵⁵. BLAST results were filtered to non-viral contig IDs, which were used to generate a concatenated fasta file of non-viral contigs. This fasta of non-viral contigs was then used to remove aligned reads from the host/parasite cleaned libraries using *STAR* aligner. To remove any PCR duplicates, the remaining reads were then deduplicated using *clumpify* from the BBMap suite. These presumed virus-only, deduplicated reads were then assembled into contigs using *SPAdes* with the `--rnaviral` and `--only-assembler` parameters, with the contigs then filtered to sequences with a length > 500 bases. Contigs from the resulting filtered set were identified using *blastn* with an e-value of 1e-5 and validated using *CheckV* v1.0.3⁵⁶. BLAST results were filtered to the top 10 viruses for each contig, and associated accession numbers were used to retrieve a non-redundant FASTA file of viral sequences with Entrez batch download⁵⁷. This virome was then indexed and host/parasite clean libraries were aligned using *STAR* aligner. Statistics per virus were then calculated using *samtools coverage* and *mpileup* v1.22⁵⁸ on the resulting alignment files. All downstream analyses were performed in R v4.5.1 “Great Square Root”⁵⁹, and plots were generated with *ggplot2* v3.5.2⁶⁰ and *ggpubr* v0.6.0⁶¹.

Statistical significance between virus number and colony strength was calculated using the *glm* function in base R with colony strength as the independent variable and the number of detected and undetected viruses as the dependent variable in a binomial model. Multiple comparisons were performed using the *multcomp* package⁶². Actin normalized read counts per virus were clustered and plotted using the *pheatmap* R package⁶³. Actin was chosen as the housekeeping gene used to normalize read counts across libraries.

LSV-aligned reads were assembled into contigs, and their RdRP genes were extracted for multisequence alignment with *RaXML-NG* v1.2.2 under the GTRGAMMA model over 100 bootstrap replicates⁶⁴. The resulting best tree was plotted in R with the *ggtree* and *treeio* packages^{65,66}.

Differential gene expression

Gene counts for host differential gene expression were calculated using *Salmon* v1.10.1⁶⁷ and the Amel HAv3.1 transcriptome. Differential expression was calculated using the *DESeq2* R package v1.48.1⁶⁸. Calculations were performed at both the transcript and gene levels. Gene set enrichment was performed using the *clusterProfiler* R package v4.16.0⁶⁹, with access to Kyoto Encyclopedia of Genes and Genomes (KEGG) databases^{70–72}.

Data availability

The RNA libraries for this study are available on NCBI under project accession PRJNA1364028 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1364028>).

Received: 17 December 2025; Accepted: 26 February 2026

Published online: 03 March 2026

References

- Aurell, D. et al. A national survey of managed honey bee colony losses in the USA: Results from the Bee Informed Partnership for 2020–21 and 2021–22. *J. Apic. Res.* **63** (1), 1–14 (2024).
- Steinhauer, N. et al. *United States Honey Bee Colony Losses 2022–2023: Preliminary Results from the Bee Informed Partnership*. Link. (2023).
- Nearman, A. et al. Insights from US beekeeper triage surveys following unusually high honey bee colony losses 2024–2025. *Sci. Total Environ.* **1003**, 180650 (2025).
- Lamas, Z. S., Chen, Y. & Evans, J. D. Case report: Emerging losses of managed honey bee colonies. *Biology* **13**(2), 117 (2024).
- Jordan, A. et al. Economic dependence and vulnerability of United States agricultural sector on insect-mediated pollination service. *Environ. Sci. Technol.* **55**(4), 2243–2253 (2021).
- Stephen, F. Epidemiology and biosecurity for veterinarians working with honey bees (*Apis mellifera*). the veterinary clinics of North America. *Food Anim. Pract.* **37**(3), 479–490 (2021).
- Lee, K. et al. Honey bee surveillance: a tool for understanding and improving honey bee health. *Curr. Opin. Insect Sci.* **10**, 37–44 (2015).
- Cornman, R. S. et al. Pathogen webs in collapsing honey bee colonies. *PLoS One* **7**(8), e43562 (2012).
- Kwon, M., Jung, C. & Kil, E. J. Metagenomic analysis of viromes in honey bee colonies (*Apis mellifera*; Hymenoptera: Apidae) after mass disappearance in Korea. *Front. Cell. Infect. Microbiol.* **13**, 1124596 (2023).
- Kadlečková, D. et al. Discovery and characterization of novel DNA viruses in *Apis mellifera*: expanding the honey bee virome through metagenomic analysis. *Msystems* **9** (4), e00088–e00024 (2024).
- Ryabov, E. V. et al. *Apis mellifera* Solinvivirus-1, a novel honey bee virus that remained undetected for over a decade, is widespread in the USA. *Viruses* **15**(7), 1597 (2023).
- Ray, A. M. et al. Distribution of recently identified bee-infecting viruses in managed honey bee (*Apis mellifera*) populations in the USA. *Apidologie* **51**(5), 736–745 (2020).
- Ryabov, E. V. et al. Dynamic evolution in the key honey bee pathogen deformed wing virus: Novel insights into virulence and competition using reverse genetics. *PLoS Biol.* **17** (10), e3000502 (2019).
- Hesketh-Best, P. J. et al. Dominance of recombinant DWV genomes with changing viral landscapes as revealed in national US honey bee and varroa mite survey. *Commun. biology.* **7** (1), 1623 (2024).

15. Wallberg, A. et al. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nat. Genet.* **46** (10), 1081–1088 (2014).
16. Evans, J. D., Schwarz, R. & Childers, A. *HoloBee Database v 1*. 2016. (2016).
17. Daisley, B. A. & Reid, G. BEEexact: A metataxonomic database tool for high-resolution inference of bee-associated microbial communities. *mSystems* **6**(2), 10.1128/msystems.00082–21 (2021).
18. Engel, P. et al. The bee microbiome: Impact on bee health and model for evolution and ecology of host-microbe interactions. *mBio* <https://doi.org/10.1128/mBio.02164-15> (2016).
19. Galbraith, D. A. et al. Parallel epigenomic and transcriptomic responses to viral infection in honey bees (*Apis mellifera*). *PLoS Pathog.* **11**(3), e1004713 (2015).
20. Badaoui, B. et al. RNA-sequence analysis of gene expression from honeybees (*Apis mellifera*) infected with *Nosema ceranae*. *PLoS One* **12**(3), e0173438 (2017).
21. Genersch, E. et al. The German bee monitoring project: a long term study to understand periodically high winter losses of honey bee colonies. *Apidologie* **41** (3), 332–352 (2010).
22. Cox-Foster, D. L. et al. A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318** (5848), 283–287 (2007).
23. Remnant, E. J. et al. A Diverse Range of Novel RNA Viruses in Geographically Distinct Honey Bee Populations. *J. Virol.* **91** (16), 19 (2017).
24. Galbraith, D. A. et al. Investigating the viral ecology of global bee communities with high-throughput metagenomics. *Sci. Rep.* **8** (1), 1–11 (2018).
25. Tozkar, C. Ö. et al. Metatranscriptomic analyses of honey bee colonies. *Front. Genet.* **6**, 100 (2015).
26. Thaduri, S. et al. Global similarity, and some key differences, in the metagenomes of Swedish varroa-surviving and varroa-susceptible honeybees. *Sci. Rep.* **11** (1), 23214 (2021).
27. Gebremedhn, H. et al. Metagenomic approach with the NetoVIR enrichment protocol reveals virus diversity within Ethiopian honey bees (*Apis mellifera simensis*). *Viruses* **12**(11), 1218 (2020).
28. Runckel, C. et al. Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, *Nosema*, and *Crithidia*. *PLoS. One* **6**(6), e20656 (2011).
29. Chen, Y. P. et al. Multiple virus infections in the honey bee and genome divergence of honey bee viruses. *J. Invertebr. Pathol.* **87** (2/3), 84–93 (2004).
30. Delaplane, K. S., van der Steen, J. & Guzman-Novoa, E. Standard methods for estimating strength parameters of *Apis mellifera* colonies. *J. Apic. Res.* **52**(1), 1–12 (2013).
31. Nazzi, F. et al. Synergistic parasite-pathogen interactions mediated by host immunity can drive the collapse of honeybee colonies. *PLoS. Pathog.* **8**(6), e1002735 (2012).
32. Gregory, P. G. et al. Conditional immune-gene suppression of honeybees parasitized by Varroa mites. *J. Insect Sci. (Tucson)*. **5** (20), 57 (2005).
33. Nazzi, F. & Pennacchio, F. Disentangling multiple interactions in the hive ecosystem. *Trends Parasitol.* **30** (12), 556–561 (2014).
34. Belaid, M. et al. Bacterial contamination of haemolymph in emerging worker honeybee (*Apis mellifera* L.) parasitized by Varroa destructor. (2018).
35. Annoscia, D. et al. Mite infestation during development alters the in-hive behaviour of adult honeybees. *Apidologie* **46**, 306–314 (2015).
36. Prisco, G. D. et al. Dynamics of persistent and acute Deformed wing virus infections in honey bees, **Apis mellifera**. *Viruses* **3**(12), 2425–2441 (2011).
37. Brutscher, L. M., Daughenbaugh, K. F. & Flenniken, M. L. Virus and dsRNA-triggered transcriptional responses reveal key components of honey bee antiviral defense. *Sci. Rep.* **7** (1), 6448 (2017).
38. Ryabov, E. V. et al. The Iflaviruses Sacbrood virus and Deformed wing virus evoke different transcriptional responses in the honeybee which may facilitate their horizontal or vertical transmission. *PeerJ* **4**, e1591 (2016).
39. Pizzorno, M. C. et al. Transcriptomic responses of the honey bee brain to infection with Deformed wing virus. *Viruses* **13**(2), 287 (2021).
40. Chen, Y. P. et al. Israeli Acute Paralysis Virus: Epidemiology, pathogenesis and implications for honey bee health. *PLoS Pathog.* <https://doi.org/10.1371/journal.ppat.1004261> (2014).
41. Deng, Y. et al. Chronic bee paralysis virus exploits host antimicrobial peptides and alters gut microbiota composition to facilitate viral infection. *ISME J.* **18** (1), wrac051 (2024).
42. Li, Z. et al. Differential physiological effects of neonicotinoid insecticides on honey bees: A comparison between *Apis mellifera* and *Apis cerana*. *Pestic. Biochem. Physiol.* **140**, 1–8 (2017).
43. Hou, M. S. et al. Effects of imidacloprid on immune detoxification-related gene expression and immune detoxification enzymes activity in nurse bees of *Apis mellifera ligustica*. (2020).
44. Biliková, K. et al. Apisimin, a new serine–valine-rich peptide from honeybee (*Apis mellifera* L.) royal jelly: purification and molecular characterization. *FEBS Lett.* **528** (1–3), 125–129 (2002).
45. Strachecka, A. et al. Coenzyme Q10 treatments influence the lifespan and key biochemical resistance systems in the honeybee, *Apis mellifera*. *Arch. Insect Biochem. Physiol.* **86** (3), 165–179 (2014).
46. Menail, H. A. et al. Age-related flexibility of energetic metabolism in the honey bee *Apis mellifera*. *FASEB J.* **37** (11), e23222 (2023).
47. Dussaubat, C. et al. Flight behavior and pheromone changes associated to *Nosema ceranae* infection of honey bee workers (*Apis mellifera*) in field conditions. *J. Invertebr. Pathol.* **113** (1), 42–51 (2013).
48. Benaets, K. et al. Covert deformed wing virus infections have long-term deleterious effects on honeybee foraging and survival. *Proc. R. Soc. Lond. B Biol. Sci.* **284** (1848), 20162149–p (2017).
49. Boncristiani, H. et al. Direct effect of acaricides on pathogen loads and gene expression levels in honey bees *Apis mellifera*. *J. Insect Physiol.* **58** (5), 613–620 (2012).
50. Brown, J., Pirrung, M. & McCue, L. A. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* **33** (19), 3137–3139 (2017).
51. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30** (15), 2114–2120 (2014).
52. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29** (1), 15–21 (2013).
53. Bushnell, B. *BBTools software package*. e, (2014).
54. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19** (5), 455–477 (2012).
55. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 1–9 (2009).
56. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39** (5), 578–585 (2021).
57. Schuler, G. D. Entrez: Molecular biology database and retrieval system. In *Methods in enzymology* 141–162 (Elsevier, 1996).
58. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**(16), 2078–2079 (2009).
59. Team, R. C. R: *A language and environment for statistical computing* (R Foundation for Statistical Computing, 2025).
60. Wickham, H. & Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
61. Kassambara, A. *ggpubr: ggplot2-based publication ready plots*. R package version, : p. 2. (2018).

62. Hothorn, T. et al. *Multcomp: simultaneous inference for general linear hypotheses*. URL <http://CRAN.R-project.org/package=multcomp>, R package version, : pp. 1–2. (2012).
63. Kolde, R. & Kolde, M. R. Package ‘pheatmap’. *R package*. **1** (7), 790 (2015).
64. Kozlov, A. M. et al. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35** (21), 4453–4455 (2019).
65. Yu, G. et al. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8** (1), 28–36 (2017).
66. Wang, L. G. et al. Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. *Mol. Biol. Evol.* **37** (2), 599–603 (2020).
67. Patro, R. et al. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*. **14** (4), 417–419 (2017).
68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
69. Xu, S. et al. Using clusterProfiler to characterize multiomics data.. *Nat. Protoc.* <https://doi.org/10.1038/s41596-024-01020-z> (2024).
70. Kanehisa, M. et al. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53** (D1), D672–D677 (2025).
71. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28** (11), 1947–1951 (2019).
72. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1), 27–30 (2000).

Acknowledgements

We would like to thank the beekeepers whose colonies and operations are represented in this study.

Author contributions

AN Wrote the main manuscript, methodology, formal analysis, and prepared figures; ZSL, ELN, JF, CM, and AS performed sample acquisition and field colony inspections; DB and WFH prepared samples and collected data; JDE and YPC led the conceptualization and manuscript editing.

Funding

This work is supported in part by the USDA Animal and Plant Health Inspection Service (APHIS) fund (8130–0960; 8130–0990) and USDA Farm Service Agency (FSA) fund (#FSA25IRA0012292).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-42605-w>.

Correspondence and requests for materials should be addressed to A.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2026