The Honey Bee Gene *Bee Antiviral Protein-1* Is a Taxonomically Restricted Antiviral Immune Gene

Alexander J. McMenamin [1,2,3], Laura M. Brutscher [3], Katie F. Daughenbaugh [1,3] and Michelle L. Flenniken [1,2,3]*

[1] Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, MT, United States, [2] Department of Microbiology and Immunology, Montana State University, Bozeman, MT, United States, [3] Pollinator Health Center, Montana State University, Bozeman, MT, United States

Insects have evolved a wide range of strategies to combat invading pathogens, including viruses. Genes that encode proteins involved in immune responses often evolve under positive selection due to their co-evolution with pathogens. Insect antiviral defense includes the RNA interference (RNAi) mechanism, which is triggered by recognition of non-self, virally produced, double-stranded RNAs. Indeed, insect RNAi genes (e.g., *dicer* and *argonaute-2*) are under high selective pressure. Honey bees (*Apis mellifera*) are eusocial insects that respond to viral infections *via* both sequence specific RNAi and a non-sequence specific dsRNA triggered pathway, which is less well-characterized. A transcriptome-level study of virus-infected and/or dsRNA-treated honey bees revealed increased expression of a novel antiviral gene, GenBank: MF116383, and *in vivo* experiments confirmed its antiviral function. Due to *in silico* annotation and sequence similarity, MF116383 was originally annotated as a probable *cyclin-dependent serine/threonine-protein kinase.* In this study, we confirmed that MF116383 limits virus infection, and carried out further bioinformatic and phylogenetic analyses to better characterize this important gene—which we renamed *bee antiviral protein-1* (*bap1*). Phylogenetic analysis revealed that *bap1* is taxonomically restricted to Hymenoptera and *Blatella germanica* (the German cockroach) and that the majority of *bap1* amino acids are evolving under neutral selection. This is in-line with the results from structural prediction tools that indicate Bap1 is a highly disordered protein, which likely has relaxed structural constraints. Assessment of honey bee gene expression using a weighted gene correlation network analysis revealed that *bap1* expression was highly correlated with several immune genes—most notably *argonaute-2*. The coexpression of *bap1* and *argonaute-2* was confirmed in an independent dataset that accounted for the effect of virus abundance. Together, these data demonstrate that *bap1* is a taxonomically restricted, rapidly evolving antiviral immune gene. Future work will determine the role of *bap1* in limiting replication of other viruses and examine the signal cascade responsible for regulating the expression of *bap1* and other honey bee antiviral defense genes, including coexpressed *ago-2*, and determine whether the virus limiting function of *bap1* acts in parallel or in tandem with RNAi.

Keywords: honey bee, antiviral, MF116383, *bee antiviral protein-1 (bap1)*, honey bee virus, honey bee immune system, ago2

INTRODUCTION

Conflict is the engine of evolution by natural selection and ecological diversification. Inter-organism conflict is most apparent when one organism parasitizes another, thus pitting the interests of two organisms against one another. Conflict between hosts and parasites drives the evolution of strategies to distinguish self from non-self and defend against invading pathogens, which are critical functions of immune systems (1). Metazoan immune systems are subdivided into the innate and adaptive responses which are further classified as humoral or cellular responses (1–6). Both humoral and cellular responses have been implicated in antiviral immunity in insects, including *Apis mellifera*—the western honey bee, a eusocial, cavity-nesting insect that is an important plant pollinator (7–9). The innate antibacterial and antifungal humoral responses in insects are induced by Pathogen Recognition Receptors (PRRs) which recognize Pathogen Associated Molecular Patterns (PAMPs) (10). The recognition of PAMPs by PRRs activates immune cascades—such as the Toll (NF-kB/Dorsal), Imd (NF-kB/Relish), and Jak/STAT pathways—which regulate the production and secretion of soluble antimicrobial peptides (AMPs) into the hemolymph [reviewed in (4–6, 10–13)]. These pathways are also involved in insect antiviral immunity, although the primary antiviral immune response is the RNA interference (RNAi) response [reviewed in (5)]. The RNAi response is a sequence-specific post-transcriptional gene silencing mechanism induced upon the recognition of double stranded RNAs (dsRNAs), including the replicative forms of single-stranded positive sense RNA viruses, by Dicer endoribonucleases. Dicer cleaves long dsRNAs into 21-22 bp short interfering RNAs, one strand of which (the guide strand) is retained by Argonaute-2 (Ago2), which is part of the RNAi-induced silencing complex (RISC). The RISC surveils the cell for complementary target RNAs, including viral genomes and transcripts, that are recognized, bound, and cleaved—which in turn limits virus infections (14–19) [reviewed in (4, 5, 20)].

Historically the invertebrate immune system was considered strictly innate (i.e., hosts respond naïvely to each encounter with a pathogen), while it was thought that the adaptive immune system, which includes pathway-specific and memory responses, was limited to jawed vertebrates [reviewed in (2, 11, 21)]. However, from an ecological and evolutionary perspective, it would be expected that insect hosts would have evolved immune systems with plasticity and specificity to respond to repeated exposure to pathogens within and across generations. Indeed, recent studies indicate that *Drosophila melanogaster* exhibits an adaptive immune response to viral infection (22–25). This response is predicated on the production of virus genome-derived DNA (vDNA) by reverse transcriptase, followed by vDNA transcription, and cleavage of the RNA transcripts into short interfering RNAs (siRNAs) (23, 24). Because RNAi proteins, including *dicer* and *ago2,* are involved in the conflict between hosts and viruses one would hypothesize that they are under positive selection and evolving rapidly (26, 27). In fact, analysis of dN/dS ratios of RNAi genes across multiple insect/arthropods indicate that these proteins have higher rates of

adaptive evolution and selective sweeps compared to paralogous "housekeeping" genes, including in fruit flies and honey bees (28, 29).

Similar to other insects, RNA interference is antiviral in honey bees, as co-administration of a virus with virus-sequence specific dsRNA results in a reduction of virus abundance (30–37). For example, larval and adult honey bees fed diets with deformed wing virus (DWV) and DWV-specific dsRNA had greater longevity and lower DWV levels than bees exposed to DWV alone (32). In addition to sequence-specific RNAi, honey bees and bumble bees mount a virus-limiting response when exposed to dsRNA of any sequence (33, 36, 38–40). In contrast, this non-sequence specific dsRNA-triggered immune response has not been observed in highly investigated, solitary insects, including fruit-flies and mosquitos. Although the details of this non-sequence specific dsRNA-triggered antiviral mechanism have yet to be fully elucidated, the transcriptional level response to dsRNA both independently and in the context of virus infection have been well-characterized in honey bees, and the expression of select genes, including those involved in RNAi have also been examined in bumble bees (33, 36, 39–44). Specifically, a honey bee transcriptome sequencing study identified hundreds of genes differentially expressed in response to virus infection and/or dsRNA injection (36). The antiviral roles of two of the genes that exhibited greater expression in response to virus and/or dsRNA, *dicer-like* and the gene encoded by GenBank: MF116383, herein renamed *bee antiviral protein-1* (*bap1*), were confirmed *in vivo*. Specifically, RNAi-mediated reduction of expression of either of these genes in honey bees resulted in greater levels of the model virus Sindbis virus (SINV) compared to virus levels in bees treated with a non-specific dsRNA control (36).

As described above, Dicer is an endoribonuclease that processes dsRNA into siRNAs that are bound by Ago2/RISC and serve as sequence-specific guides for the targeted cleavage of cognate RNAs, including viral genomes and transcripts. The precise antiviral function of *bee antiviral protein-1* (*bap1*) (GenBank: MF116383.1) is still unknown. It was originally described as putative cyclin-dependent serine/threonine kinase based on *in silico* analysis (36, 45). It was one of the most highly differentially expressed genes in response to virus-infection (36). Therefore, to further characterize the gene, the full-length cDNA was Sanger sequenced revealing that the transcript length was longer than predicted (i.e., 5,158 nucleotide) and shared 91% nucleotide identity with an *Apis cerana* probable cyclin-dependent serine/threonine-protein kinase transcript (XM_017051141.1) and therefore it was similarly described in a previous publication (36).

In this study, we replicated the finding that the gene encoded by GenBank: MF116383 restricts virus infection and further characterized this gene. Sequence analysis using HMMER and domain searches failed to identify a putative kinase domain, therefore we renamed this gene *bee antiviral protein-1* (*bap1*) to better reflect its function. Analysis of the phylogenetic distribution of *bap1* and its orthologs revealed that it is taxonomically restricted to Hymenopteran insects and the German cockroach (*Blatella germanica*). To evaluate the potential evolutionary selective pressures on this gene,

we calculated the site-wise dN/dS ratios and determined that the high substitution rate indicated by Bayesian phylogenetic inference is likely indicative of neutral selection across the majority of the gene, although some sites are under purifying and positive selection. To identify the pathways and proteins with which *bap1* may be associated, we used a weighted gene correlation network analysis (WGCNA) to identify highly coexpressed genes. This analysis revealed that *bap1* is coexpressed with several immune genes, most notably *ago2*, which we confirmed in an independent dataset. Together these data demonstrate that *bap1* is an important honey bee antiviral defense gene that is taxonomically restricted to primarily social insects (i.e., members of Apidae, Formicoidea, and *Blatella germanica*).

RESULTS AND DISCUSSION

Virus-Limiting Role of Honey Bee *bap1* (GenBank: MF116383) Confirmed

For experiments described herein and previous studies, we utilized Sindbis virus (SINV), which is a well-characterized model virus that has been extensively utilized to investigate antiviral defense mechanisms in a wide range of insects including fruit flies, mosquitos, and honey bees (22, 33, 36, 46, 47). Therefore, use of this virus in experiments facilitates comparison of immune responses in both natural mosquito hosts and non-native hosts (i.e., honey bee and fruit fly) that have not co-evolved with this virus. Additional advantages of using SINV for *in vivo* experiments include: lack of confounding infection with this virus, facile virus purification from cultured cells, quantification of virions *via* plaque assay, lack of viral encoded RNAi suppressor proteins, and the ability to generate virus from a cDNA cloned construct that includes a carboxy-terminal green fluorescent protein (GFP) tag that facilities virus tracking *via* microscopy and assessment of virus abundance at the protein level *via* Western blot analyses (33, 36). A disadvantage of SINV is that it does not naturally infect honey bees. However, the generation of infectious clones of honey bee viruses is relatively recent and they have not been utilized for *in vivo* studies in laboratories beyond those in which they were developed (48–50). While some studies utilize honey bee virus preparations obtained from pupae, and while there are advantages of these lines of investigation, these virus preparations may include co-purifying viruses, as well as other proteins, and they are not a standardized source of infectious material. It is important to investigate host-virus specific interactions using a panel of viruses, including both model viruses and naturally infecting honey bee viruses, however this was beyond the scope of this study.

In honey bees, dsRNA triggers both sequence specific and non-sequence specific virus limiting mechanisms that reduce virus abundance (30–36, 51). Therefore, the most relevant control for *in vivo* honey bee experiments aimed at investigating the impact of RNAi-mediated reduction of target gene expression on viral abundance approach, is a virus-infected and non-sequence specific dsRNA treated group, rather than comparison of virus and target gene expression levels in bees that did not

receive dsRNA. To confirm that *bap1* limits virus infection, we carried out honey bee virus infection studies in the context of either non-sequence specific control dsRNA (ns-dsRNA) or *bap1*-specific dsRNA and quantified the relative expression of *bap1* and virus abundance using quantitative PCR (qPCR). Bees were collected for analysis at 72 h post-infection because previous work reported that RNAi-mediated knock down efficiency of *bap1* and the effect on virus abundance was most apparent at this time point (36). In this study, similar to a previous report, honey bees infected with the model virus Sindbis-GFP (SINV) exhibited 1.65× greater expression of *bap1* compared to mock-infected bees (Dunnett's test, $p < 0.001$) (**Figure 1A**) (36). Co-administration of virus and ns-dsRNA, which shares no homology with either virus or host sequences and is a viral-associated molecular pattern (VAMP), also resulted in a 1.92× increase of *bap1* expression at 72 h post-infection (hpi) compared to levels in mock-infected honey bees (**Figure 1A**). The RNAi-mediated reduction of *bap1* expression was achieved by injecting *bap1*-seqence specific dsRNA, which reduced mean *bap1* expression to 0.77× and 0.23×, relative to mock-infected bees, in honey bees injected with *bap1* dsRNA or with *bap1* dsRNA and SINV, respectively (Dunnett's test, $p < 0.001$) (**Figure 1A**).

Virus abundance in virus-infected bees compared to virus-infected bees co-injected with either ns-dsRNA or *bap1*-dsRNA was examined by determining the viral RNA copies (which represent both viral genomes and transcripts during active infections) using qPCR. In this study, similar to previous studies, co-injection of ns-dsRNA reduced SINV levels by 40% relative to levels in virus-infected bees ($p = 0.045$), a result which confirmed the virus-limiting impact of administration of immune system stimulating ns-dsRNA (**Figure 1B**). Furthermore, honey bees with reduced levels of *bap1* expression harbored 193% more virus than control bees which received ns-dsRNA ($p = 0.016$; **Figure 1B**). Virus abundance was similar in SINV-infected honey bees and SINV-infected bees that were co-injected with *bap1*-dsRNA, as compared to virus-only infected bees, due to the virus limiting impact of dsRNA (36). Therefore, the role of *bap1 in vivo* is best assessed by comparing virus abundance in the SINV-infected and ns-dsRNA treated bees vs. virus abundance in virus-infected bees treated with *bap1*-specific dsRNA (**Figure 1B**) Together, these results indicate that antiviral defense was hindered in honey bees with reduced *bap1* expression and confirm the findings of Brutscher et al. that *bap1* has an antiviral function in honey bees (36).

Phylogenetic Analyses of *bee antiviral protein-1* Reveal It Is a Taxonomically Restricted, Divergent Gene That Is Evolving Under Neutral Selection

Bee antiviral protein-1 (GenBank: MF116383) is a recently described honey bee antiviral gene transcribed from LOC725387 to produce a 5,158 nucleotide (nt) long transcript that produces a 1,511 amino acid protein (**Supplementary Figure S1**) (36). Although, transcriptome and *in vivo* studies indicate it is important for honey bee antiviral defense, the mechanistic and
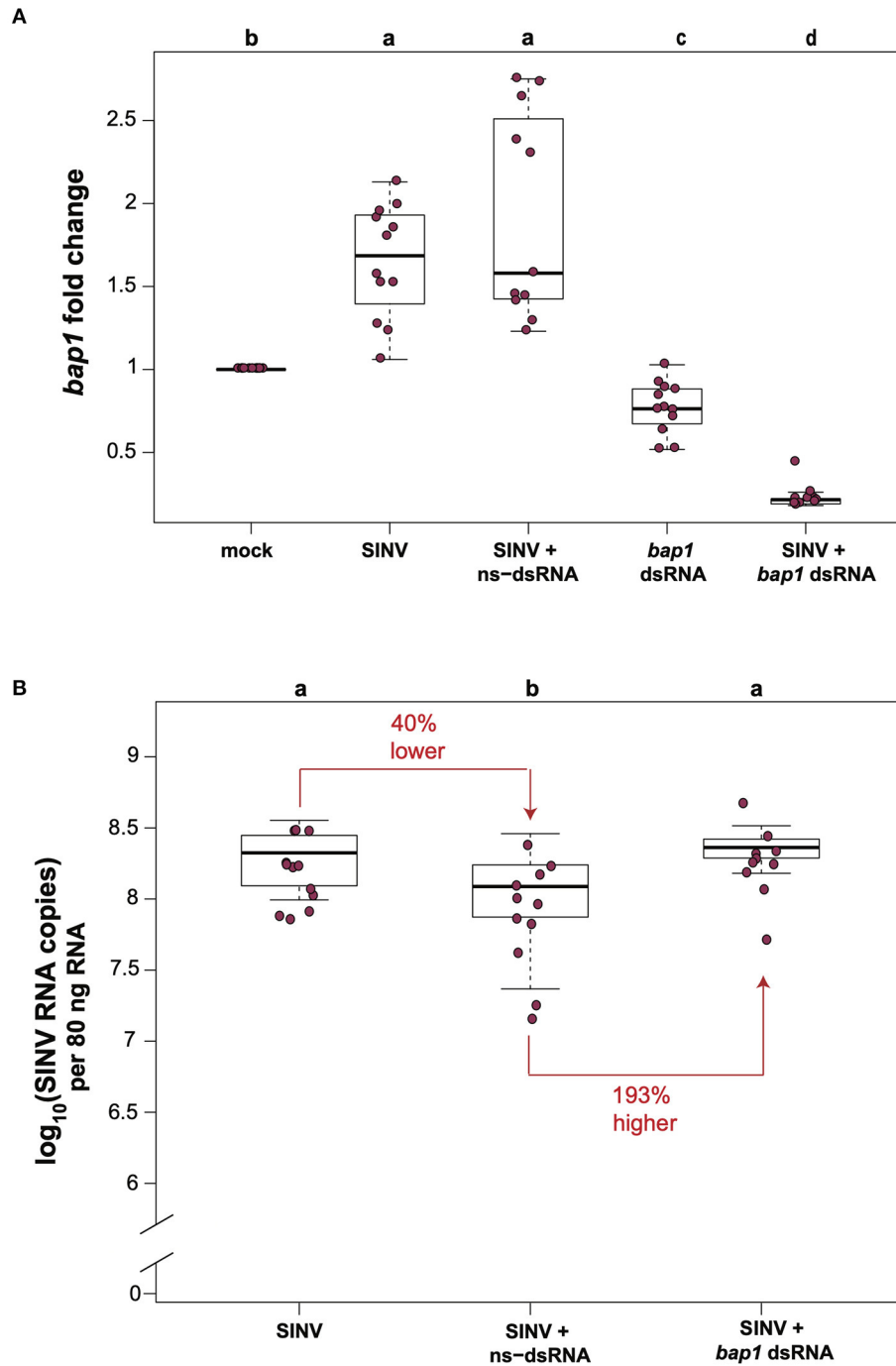
FIGURE 1 | *Bee antiviral protein-1 (bap1)* is an antiviral immune gene. **(A)** Relative expression of *bap1* was assessed by qPCR, normalized to *rpl8* and fold changes were calculated relative to buffer-injected bees (mock). Bees infected with Sindbis-GFP (SINV) exhibited 1.65× greater expression of *bap1* compared to mock-infected bees (Dunnett's test, $p < 0.001$). Co-injection of virus and ns-dsRNA also resulted in a 1.92× increase of *bap1* expression at 72 h post-infection (hpi) compared to levels in mock-infected honey bees. RNAi-mediated gene knock-down was achieved by injecting *bap1*-seqeunce specific dsRNA which resulted in a mean 0.77× reduction, and a 0.23 fold reduction in *bap1* expression in bees co-injected with *bap1* dsRNA and virus relative to mock ($p < 0.001$). **(B)** Virus abundance was measured by determining the viral RNA copies using qPCR. Bees coinjected with ns-dsRNA and SINV and 40% lower SINV levels compared to SINV-only injected ($p = 0.045$). Bees with reduced *bap1* expression had 193% higher SINV levels relative to bees coinjected with SINV and ns-dsRNA ($p = 0.016$). Different letters above the treatments groups indicates a statistically significant difference.

functional roles of *bap1* have not yet been elucidated. Therefore, we utilized HHpred with default parameters to identify any putative conserved amino acid domains to infer Bap1 function at the cellular level (52, 53). Hidden Markov Model searches against the protein data bank (PBD), Pfam-A_34, Swissprot, COG_KOG, or NCBI Conserved Domains databases all failed to identify any conserved domains. Therefore, in contrast to initial reports based on automated annotations, the gene encoded by GenBank: MF116383 does not likely have a kinase domain, and therefore *bap1* is a more accurate name for this gene and associated protein (ASQ15625).

To gain insight on the phylogenetic distribution of *bap1*, which may also provide functional insights, we identified orthologous proteins in other insects and assessed their phylogenic relationship (**Figure 2A**). Honey bees belong to the order Hymenoptera, which is a large order of insects with over 150,000 species. Orthologs of *bap1* were identified in the major lineages of Hymenoptera, including Apocrita (wasps and Anthophila), and Eusymphyta (which include sawflies, horntails, and wood wasps). Whereas, orthologous genes were not identified in Diptera (including *Drosophila melanogaster,* as well as mosquitos in genera *Culex, Aedes,* and *Anopheles*) or most other insect sequences in NCBI databases with the exception of one protein in *Blatella germanica* (the German cockroach).

Orthologs of *bap1* were identified *via* the Position-Specific Iterated Basic Local Alignment (PSI-BLAST) tool on the website of the National Center for Biotechnology Information (NCBI) (**Supplementary Table S2**) (54–56). We identified 73 orthologs with percent amino acid identities ranging from 17 to 82% (**Supplementary Figure S1**; **Supplementary Table S3**). The protein sequences were aligned using MEGA X (MUSCLE with default parameters) (57) and analyzed by Bayesian inference phylogenetic analysis using MrBayes v3.2.7a (58, 59) with *Blatella germanica* as the out group since it is the only non-Hymenopteran taxon on the tree (**Figure 2A**). Unfortunately, Euymphyta and Apocrita (aside from Anthophila) remain very under sampled on this tree due to a lack of genome assemblies or low-quality genome assemblies. It is interesting that *bap1* orthologs were not identified in model insects such as *Drosophila melanogaster* (Diptera)*, Spodoptera frugiperda* (Lepidoptera), or *Tribolium castaneum* (Coleoptera). Phylogenomic analysis using 1,478 ortholog groups and published transcriptome data sets suggests the ancestors of Hymenoptera and Blattodea split roughly 400 million years ago, compared to the ancestors of Diptera, Lepidoptera, and Coleoptera, which are phylogenetically more similar to Hymenoptera, having diverged about 350 million years ago (60). Even though bees and cockroaches diverged long before bees and flies, a *bap1* ortholog was identified in *Blatella germanica*, the German cockroach. Interestingly, experiments aimed at investigating gene function in this species using RNAi-mediated gene knockdown, revealed that *dicer2* expression was increased in response to ns-dsRNA (61). Although complete transcriptome profiling of dsRNA-treated cockroaches has yet to be evaluated, we hypothesize that *bap1* orthologs were retained in several phylogenetic lineages that have a transcriptional response to dsRNA, and that this response may be antiviral– as it is in honey bees and bumble bees. Although, this hypothesis remains

to be tested. In line with our hypothesis, fruit flies and mosquitos, which do not have *bap1* orthologs, also do not exhibit virus-limiting responses to ns-dsRNA. However, there are organisms that limit virus infection in response to ns-dsRNA that lack a *bap1* ortholog (e.g., the sand fly, *Lutzomyia longipalpis*) (62). It not clear whether this phenotype in *L. longipalpis* is a retained (from an ancestor) or derived phenotype.

To confirm that some insects lack a *bap1* ortholog, PSI-BLAST searches directed to several genomes (i.e., *Drosophila melanogaster, Lutzomyia longipalpis,* and *Aedes albopictus*) were performed, and a Hidden Markov Model approach to identify deep homologs (JackHMMER) was utilized; all failed to identify additional orthologs (63). This indicates that *bap1* is a taxonomically restricted gene either due to lineage-specific losses or high divergence (i.e., high average amino acid substitution at each position) in other lineages thereby making orthologs unidentifiable.

The phylogenetic relationships based on the Bap1 protein sequence, grouped wasps and sawflies together, which was likely an artifact of long branch lengths, and not an accurate representation of lineage (**Figure 2A**). If the Bap1 protein tree reflected the true phylogenetic relationships within Hymenoptera, you would expect Eusymphyta to form the basal clade of Hymenoptera (64). Instead, in this tree, Anthophila appears to have split from the rest of Apocrita and Formicoidea before they diverged, when in fact Anthophila is the most derived clade within Hymenoptera (64).

The high rate of amino acid substitution in Bap1 orthologs (0.3 amino acid substitutions per position) is indicative of rapid evolution (**Figure 2A**). To contextualize the seemingly high amino acid substitution rate on the Bap1 tree (**Figure 2A**), a parallel analysis was carried out on RPB1, the largest subunit of eukaryotic DNA-directed RNA polymerase II, which showed a more than 10-fold lower substitution rate (0.02, **Supplementary Figure S2**; **Supplementary Table S4**) (65). Given the high substitution rate in Bap1 we hypothesized that Bap1 was under positive selection causing divergence between lineages. A gene is under positive selection when the ratio of non-synonymous mutations (a change in amino acid; dN) to synonymous mutations (no amino acid change, dS) as >1. This indicates some selective pressure is driving the fixation of advantageous mutations in lineages. A gene is under negative selection when the dN/dS ratio is <1, which indicates selection pressure is purging amino acid-changing mutations because they are deleterious. One might expect that an antiviral gene is under positive selection due to the arms-race between a pathogen an its host, which may favor non-synonymous mutations (26). Alternatively, genes that are important to numerous important biological processes, including some immune genes, could be under negative selection, in order to ensure structural and functional preservation. Furthermore, individual amino acids can experience unique selective pressure and, therefore, whole-gene dN/dS ratios are unlikely to be reflective of evolution at individual amino acids. A maximum likelihood analysis of site-wise dN/dS ratios using PAML revealed that *bap1* has codons evolving under negative (260 sites, dN/dS = 0.16) and positive selection (123 sites, dN/dS = 1.94) but most sites are evolving
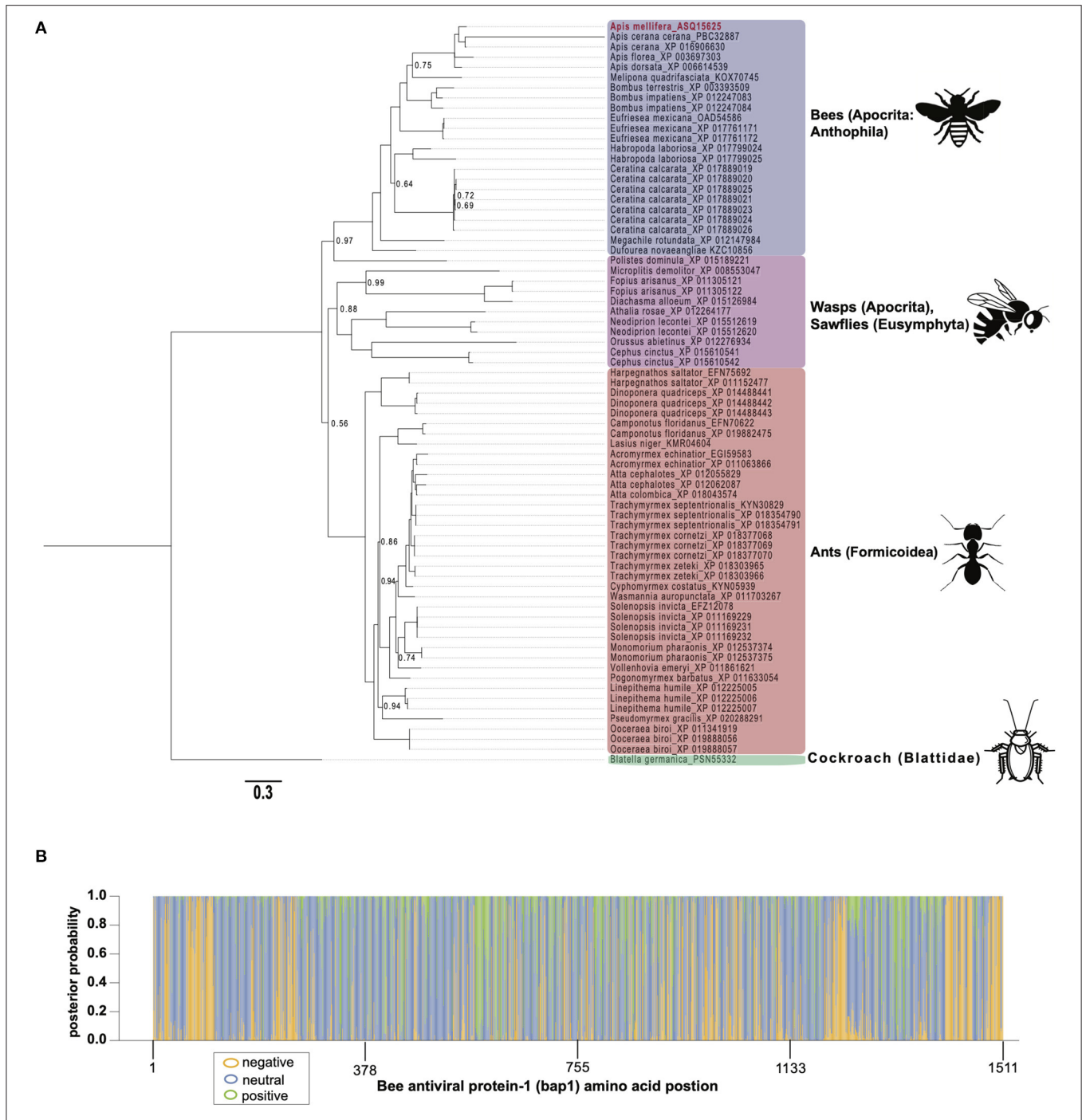
FIGURE 2 | Bee antiviral protein-1 phylogenic relationship to other Hymenopteran orthologs inferred from amino acid sequence. **(A)** Majority rule Bayesian consensus tree of Bap1 homologs derived from Bayesian analysis of amino acid alignment implemented in Mr. Bayes v3.2 using a Jones substitution model. Numbers on branches and nodes are posterior probabilities (0-1), though posterior probabilities values of 1 are not shown to improve clarity. The scale bar corresponds to proportion of amino acid change. Accession numbers are included on the branch tips and in **Supplementary Table S2**. **(B)** A corresponding majority Rule Bayesian consensus tree derived from Bayesian analysis of a codon alignment was used in a selection analysis in the CODEML package in Phylogenetic Analysis by Maximum Likelihood (PAML) 4.9. The Bayes Empirical Bayes method under model 2, which assumes 3 site classes (negative, neutral, and positive selection), was used to calculate the posterior probability that each amino acid position belonged to each site class. The posterior probabilities were then plotted as a stacked bar chart along the length of Bap1 amino acids. This method shows clear diffuse neutral selection with regions under strong positive or negative selection.

under neutral selection (797 sites, dN/dS = 1.00) (**Figure 2B**; **Table 1**; **Supplementary Figure S3**; **Supplementary Tables S5**, **S6**, **S10**) (66). This is in stark contrast to RPB1, the largest subunit of eukaryotic DNA-directed RNA polymerase II, which also has sites under neutral (191 sites, dN/dS = 1) and positive (98 sites, dN/dS= 14) selection, but most sites are under very strong negative selection (1,314 sites, dN/dS = 0.0006), which is characteristic of highly conserved proteins involved in fundamental biological processes, including transcription (**Supplementary Figures S2B**, **S4**; **Supplementary Tables S7**– **S10**). An Empirical Bayes approach was used to calculate the posterior probability that each amino acid was evolving under negative, neutral, or positive selection (**Figure 2B**; **Supplementary Figure S2B**). Overall, the pattern of selection and phylogenetic distribution indicate that the *bap1* is a taxonomically restricted ortholog group evolving primarily under neutral selection causing rapid divergence, as compared to a control gene (RPB1) (**Figure 2**; **Supplementary Figure S2**).

Bee antiviral protein-1 Is a Disordered Protein That Is Trafficked to the Endoplasmic Reticulum

Relaxed evolution is a hallmark of disordered proteins due to relaxed constraints on their structure (67–69). To ascertain whether the largely neutral selection on *bap1* orthologs (**Figure 2B**) was due to relaxed structural constraints, PrDOS (protein disorder prediction system) was used to calculate disorder prediction at each amino acid position with the default 5% false positive rate (**Figure 3A**) (70). This analysis indicated that the majority (i.e., 67.2%) of the Bap1 amino acid sequence is disordered or lacks structure (i.e., above 0.5 probability, **Figure 3A**). Intrinsic disorder in protein structure allows the function of a protein to be highly modular since it allows the binding of multiple partners, including proteins and nucleic acids (71). In fact the relaxed structural constraint, leading to decreased packing density, is associated with more rapid evolution (higher substitution rates) of disordered proteins, which is in-line with the bioinformatic analyses of *bap1,* described herein (**Figure 2A**) (67–69, 72). Due to their modularity, intrinsically disordered proteins (IDPs) often act as hubs in protein interaction networks (73–75). When bound to their target, many IDPs adopt a more structured conformation (76, 77). Future analyses, including the use of nuclear magnetic resonance (NMR) and X-ray scattering techniques, are required to definitively characterize the structure of the Bap1 protein in bound or unbound states.

To examine the putative subcellular localization of the protein encoded by *bap1*, we used SignalP-5.0 to predict the presence of a signal peptide on the N-terminus (78). Signal-P predicted the presence of a signal peptide with a probability of 0.983 with a cleavage site between amino acids 21 and 22 (**Figure 3B**). Signal peptides are N-terminal sequences on proteins that promotes targeting to the endoplasmic reticulum (ER), *via* recognition by the signal recognition particle (79, 80). Once translocated across the membrane into the ER lumen, the protein can either be released from the membrane *via* cleavage of the signal peptide by signal peptidases or it can remain associated with the membrane

(80–82). This process can happen co-translationally for larger proteins or post-translationally for smaller proteins (83). Once in the ER, however, the protein may be secreted, stay in the ER, or be sorted into intracellular compartments like the Golgi, endosomes or lysosomes (80, 84, 85). We were unable to identify any further localization signals that might indicate subcellular localization. Therefore, our data indicate that Bap1 likely resides in a membrane-bound compartment (or may be secreted) and may serve as a hub of molecular interactions (i.e., nucleic acid- protein or protein-protein) in the honey bee antiviral response. Further work to identify how Bap1 is limiting virus infection would be an exciting avenue of research.

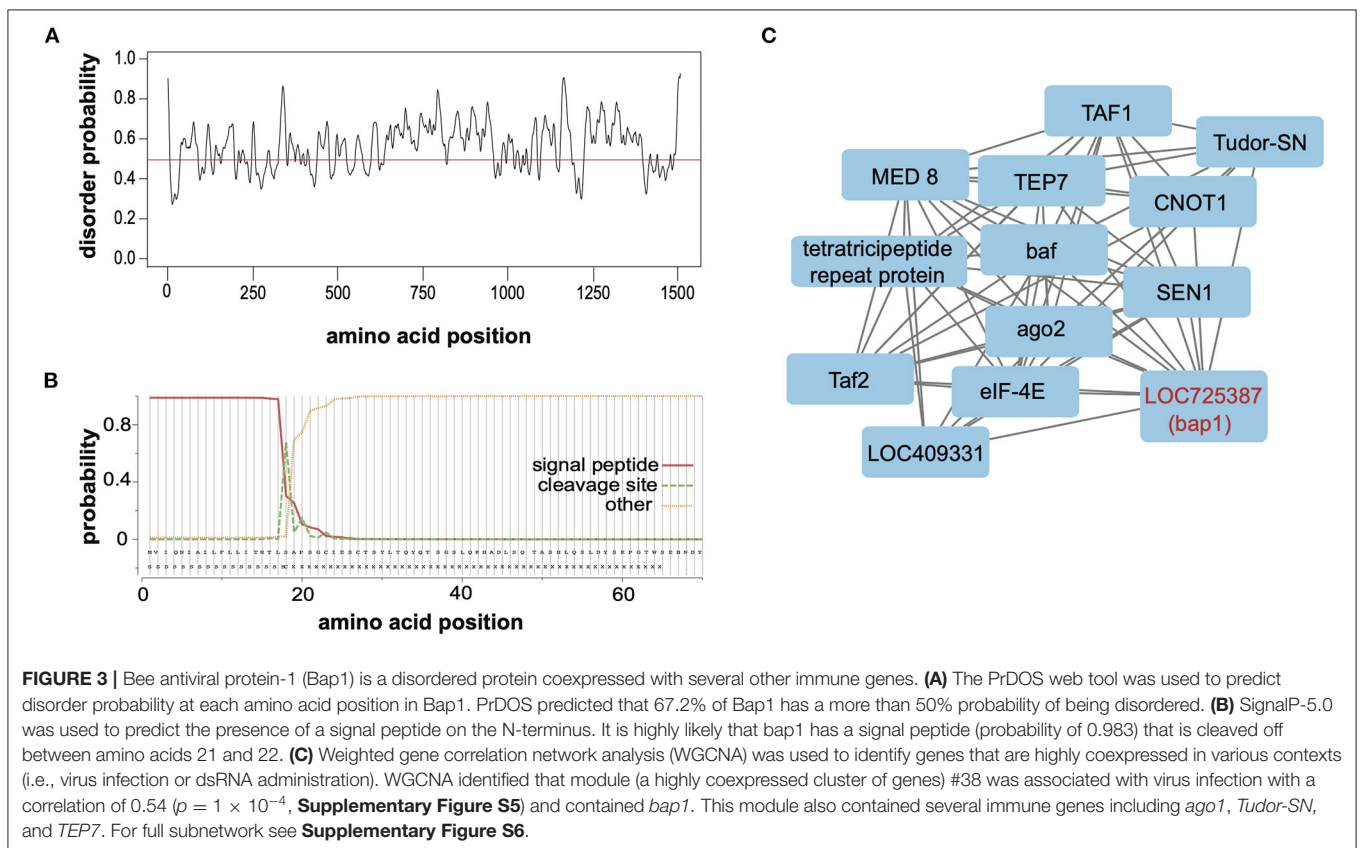## Bee antiviral protein-1 Is Co-expressed With Several Honey Bee Immune Genes

Genes with a coordinated function (e.g., act in antiviral defense) are often co-expressed. The honey bee antiviral gene *bap1* was first described and annotated in a transcriptome level study of honey bees infected with a model virus in the presence and absence of dsRNA species (36). This study assessed gene expression in individual bees over a time course (i.e., 6, 48, and 72 h post-injection) as compared to mock- infected (i.e., buffer-injected) control bees and presented fold- change and statistical significance data independently for each annotated gene at all time-points (36). To further evaluate the genes highly coexpressed with *bap1* we constructed a signed weighted gene correlation network from our previously published transcriptome dataset using the R package Weighted Gene Correlation Network Analysis (WGCNA) (36, 86, 87). After filtering, WGCNA identified 39 modules, which are clusters of highly co-expressed genes. Select modules are discussed below and pairwise correlations between genes in the identified modules are presented in **Supplementary Tables S13**– **S20**. There were three gene modules that most correlated with virus (SINV) infection (i.e., #8, #15, and #38), and one module that best correlated with dsRNA treatment (i.e., #20) (**Supplementary Figure S5**). These modules are described below, and the genes and gene-gene correlations of additional modules are included in **Supplementary Tables S17**–**S20**.

Gene module #8 and SINV infection had a correlation coefficient of 0.56 ($p = 4 \times 10^{-5}$) (**Supplementary Figure S5**). This is a small module with only 38 genes, therefore gene ontology analysis of this group was not possible, and most of the genes in this module have no known or predicted function. However, interestingly, several predicted non-coding RNAs (ncRNAs) belong to this module (e.g., LOC100578712, LOC102654940, LOC102655797, and mir3764) (**Supplementary Table S13**). Non-coding RNAs perform a wide range of functions, but most notably pre- and post-transcriptional regulation of gene expression through methylation or RNAi, respectively (88–91). Non-coding RNAs may have either a pro- or anti-viral role [reviewed in (92)]. One gene in this module was IGFn3-6 (also known as DSCAM2-like), a gene in the immunoglobulin-like superfamily. DSCAM2 is a gene that undergoes extensive mutually exclusive alternative splicing to create a wide range of somatic diversity. It is involved

TABLE 1 | Bee antiviral protein-1 (*bap1*) orthologs are primarily under neutral selection.

	M0	M1	M2
Description	Assumes one dN/dS ratio for whole gene	Assumes near-neutral selection	Assumes 3 classes with varying selection (negative, neutral or positive)
-lnL	−160,366.55	−154,730.907	−154,513.0653
Test Statistic (null is M0)		11,271.286	M2-M0: 11,706.97
			M2-M1: 434
p-value		$<1 \times 10^{-15}$	M2-M0: $<1 \times 10^{-15}$
			M2-M1: $p < 0.001$
dN/dS	0.628	negative = 0.146, neutral = 1.000	negative = 0.16
			neutral = 1.00
			positive = 1.94

Measures of selection (dN/dS) were calculated for bap1 (GenBank: MF116383) orthologs under three models. M0 assumes one dN/dS ratio for the whole protein. M1 assumes the protein is evolving under near-neutral selection and calculates dN/dS for two site classes (dN/dS = 1 and dN/dS <1). Lastly, M2 assumes there are three codon site classes in the protein (dN/dS < 1, dN/dS = 1 and dN/dS >1). The tests show M2 is the best model and that bap1 orthologs have sites under negative, neutral, and positive selection.

FIGURE 3 | Bee antiviral protein-1 (Bap1) is a disordered protein coexpressed with several other immune genes. **(A)** The PrDOS web tool was used to predict disorder probability at each amino acid position in Bap1. PrDOS predicted that 67.2% of Bap1 has a more than 50% probability of being disordered. **(B)** SignalP-5.0 was used to predict the presence of a signal peptide on the N-terminus. It is highly likely that bap1 has a signal peptide (probability of 0.983) that is cleaved off between amino acids 21 and 22. **(C)** Weighted gene correlation network analysis (WGCNA) was used to identify genes that are highly coexpressed in various contexts (i.e., virus infection or dsRNA administration). WGCNA identified that module (a highly coexpressed cluster of genes) #38 was associated with virus infection with a correlation of 0.54 ($p = 1 \times 10^{-4}$, **Supplementary Figure S5**) and contained *bap1.* This module also contained several immune genes including *ago1, Tudor-SN,* and *TEP7*. For full subnetwork see **Supplementary Figure S6**.

in neuronal development in *Drosophila melanogaster* and DSCAMs have also been implicated in immunity, including potential roles in antiviral immunity (93–96). In honey bees (*A. mellifera syriaca*) DSCAM2 SNPs were also associated with *Varroa destructor* mite resistance phenotypes, and DSCAM exhibited higher expression in virus-infected bees than in control bees (33, 97).

Gene module #15 and SINV infection had a correlation coefficient of 0.54 ($p = 8 \times 10^{-5}$) (**Supplementary Figure S5**).

This module contains 122 genes, but since many of them lacked functional assignments, no significant gene ontology terms were detected (**Supplementary Table S14**). Like gene module #8, gene module #15 also included several non-coding RNAs (LOC100577428, LOC102655361, and LOC102654929), as well as genes involved in pathogen recognition (*pgrp-s2*), transcriptional repression (*snail*), cell morphogenesis (*trbl*), cell migration regulation (*PVF1*), and a gene involved in the anti-parasitoid response (*Flo2*). *Pgrp-s2* also exhibited increased

expression in this data set in response to SINV infection. Pgrp-s2 is a member of a class of pathogen recognition receptors that activate the Toll or Imd pathway in response to pathogens, and it has also been implicated in antiviral immunity in *Bombyx mori* where it activates the Imd pathway (98). The Toll and Imd pathways, which are mediated by NF-kB-like transcription factors are both important for antiviral defense against certain viruses in *Drosophila melanogaster* and other insects (99–101) [reviewed in (5)]. The mechanism of activation of these pathways by virus infection is unclear and their precise role is not fully characterized, though Imd may contribute through regulation of apoptosis (101).

Gene module #38 and SINV infection had a correlation coefficient of 0.54 ($p = 1 \times 10^{-4}$, **Supplementary Figure S5**). This module contains 481 genes, similar to other modules, gene ontology could not be assessed since numerous genes in this cluster lacked functional annotation. However, this module included genes involved in metabolism, transcription, translation, and immunity (**Supplementary Table S15**). The immune genes in this module included *bap1* (LOC725387), which was co-expressed with additional immune genes *tep7*, *tudor-SN*, and *ago-2* (15, 16, 102–104) [reviewed in (5)]. Thioester containing protein 7 (TEP7) is a member of the thioester containing protein family which includes vertebrate complement proteins [reviewed in (105)]. In fruit flies and mosquitoes, *TEP* expression is likely regulated by JAK in response to bacterial infection and they function in bacterial recognition, opsonization and phagocytosis [reviewed in (105, 106)]. However, in *Aedes aegypti* TEP1 and TEP2 limit West Nile virus infection and TEP is indirectly involved in combatting Dengue virus (107, 108). Lastly, Tudor-SN is a nuclease and a core component of RISC of Drosophila that is also involved in cleavage of hyper-ADAR-edited dsRNAs (109–111). In summary, *bap1* is co-expressed with several genes that are implicated in antiviral immunity in several species (**Figure 3C**).

Gene module #20 and dsRNA-treatment were moderately associated with a correlation coefficient of 0.35 ($p = 0.02$) (**Supplementary Figure S5**; **Supplementary Table S16**). This module contained 806 genes and had significant enrichment for genes involved in nucleic acid binding ($p = 0.016$), metal ion binding ($p = 0.016$), and genes with a zinc finger C2H2-like domain ($p = 0.002$). These groups comprise proteins with a wide range of functions including protein, DNA, ssRNA, and dsRNA binding functions [reviewed in (112)], sometimes interacting with both nucleic acids and proteins. The zinc finger C2H2 domain is by far the most common domain in metazoan transcription factors [reviewed in (113)], but their wide range of binding activities suggests functional diversity. Indeed, there are antiviral mammalian zinc finger nucleases which can either directly bind to viral nucleic acids and promote their degradation, or otherwise regulate the immune response [reviewed in (114)].

Confirmation of *bap1* and *ago2* Coexpression in Honey Bees

To confirm the co-expression of *bap1* and *ago2*, a *post-hoc* analysis was performed on a dataset, composed of previously published data and data generated in this study (**Figure 1B**)

(47). The previously published dataset was from a study that determined that the heat shock response in honey bees is antiviral (47). Specifically, expression data from individual bees that were (a) SINV-infected, (b) exposed to heat shock (42°C for 4 h) only, or (c) both virus-infected and heat shocked (47). This combined dataset was utilized for *post-hoc* analysis because bees were subjected to various treatments that impact *bap1* expression, including heat shock which increases *bap1* expression (47). The association between *bap1* and *ago2* expression was analyzed using the following linear model:
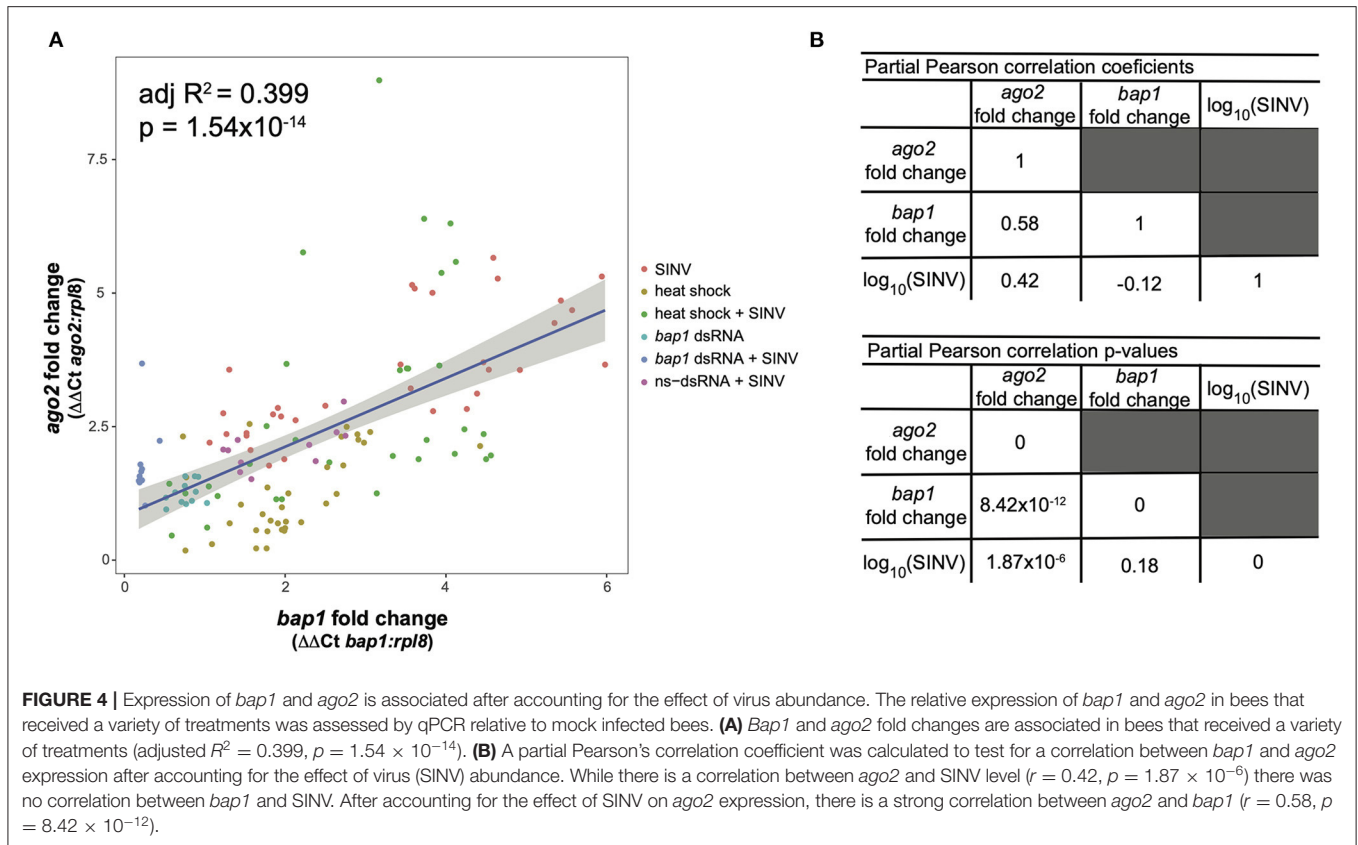
$$\text{ago2 fold change} \sim \beta_0 + \beta_1 * \text{bap1 fold change}$$
$$* \beta_2 * \text{SINV RNA copies}$$

This model structure was selected because SINV and *ago2* expression are also weakly associated (adjusted $R^2 = 0.06$, $p = 0.003$, log likelihood $= -233.4$), and therefore including the interaction term significantly improved the model. The model indicates that *bap1* expression (fold change relative to control buffer-injected bees) is a good predictor of *ago2* expression (Adjusted $R^2 = 0.399$, $p = 1.55 \times 10^{-14}$, log likelihood $= -203.3$, **Figure 4A**). This model also indicates that SINV remains a weak predictor of *ago2* expression, which was expected since *ago2* expression is induced by SINV infection (36). However, the chosen model structure explains only 40% of the variance in the dataset, indicating there are additional variables, which are not accounted for in the model, that effect *ago2* expression. See **Table 2** for full model output.

Because there is multicollinearity in the model, partial Pearson's correlation coefficients were calculated to isolate the association between *bap1* and *ago2* expression from the effect of SINV RNA copies on either variable (**Figure 4B**). After accounting for the partial effect of $\log_{10}$ (SINV RNA copies) on *ago2* expression ($r = 0.42$, $p = 1.87 \times 10^{-6}$) *bap1* expression has a high partial correlation with *ago2* expression ($r = 0.58$, $p = 8.42 \times 10^{-12}$, **Figure 4B**). Therefore, *bap1* and *ago2* are likely co-regulated. It is intriguing to speculate that this co-expression could be due to functions in a common pathway. Alternatively, it could simply be case of co-expression for a similar function (i.e., antiviral).

CONCLUSIONS

Honey bee antiviral defense mechanisms include the Toll and Imd (NFkB) signaling pathways and dsRNA-triggered mechanisms, including sequence-specific RNAi and a non-sequence specific triggered pathway that is less well-characterized (33, 36, 41, 44, 51, 115–119) [reviewed in (6, 120)]. Transcriptome level analysis and *in vivo* experiments determined that *bee antiviral protein-1* (*bap1*, GenBank: MF116383) plays an important role in honey bee antiviral defense, in the context of SINV-infection. Herein, we further characterized *bap1* using bioinformatic and phylogenetic approaches that revealed that *bap1* is taxonomically restricted to Hymenoptera and *B. germanica* and that it is evolving primarily under neutral selection, although some sites were under positive or negative selection. In line with those analyses, structural analysis predicted that Bap1 is highly disordered and neutral evolution

FIGURE 4 | Expression of *bap1* and *ago2* is associated after accounting for the effect of virus abundance. The relative expression of *bap1* and *ago2* in bees that received a variety of treatments was assessed by qPCR relative to mock infected bees. **(A)** *Bap1* and *ago2* fold changes are associated in bees that received a variety of treatments (adjusted $R^2 = 0.399$, $p = 1.54 \times 10^{-14}$). **(B)** A partial Pearson's correlation coefficient was calculated to test for a correlation between *bap1* and *ago2* expression after accounting for the effect of virus (SINV) abundance. While there is a correlation between *ago2* and SINV level ($r = 0.42$, $p = 1.87 \times 10^{-6}$) there was no correlation between *bap1* and SINV. After accounting for the effect of SINV on *ago2* expression, there is a strong correlation between *ago2* and *bap1* ($r = 0.58$, $p = 8.42 \times 10^{-12}$).

is common for disordered proteins. We hypothesize that, like other disordered proteins, Bap1 may have the capability to bind many different protein or nucleic acid targets and thus serve as a molecular interaction hub in the honey bee antiviral defense. Future work will determine whether regions of positive selection are particularly important for interfacing with binding partners (i.e., nucleic acids or protein), and identifying those binding partners. To begin to identify the pathways *bap1* may interact with, a weighted gene correlation network analysis was performed and determined *bap1* expression is highly correlated with several immune genes, including *ago2* which was confirmed in an independent dataset. Whether coexpression of *bap1* and *ago2* is due to coregulation by a shared pathway or simply due to their shared antiviral function remains to be elucidated. Overall, these data indicated that *bap1* is an exciting avenue of research as a novel antiviral gene. Since host responses to specific viruses may vary, future studies that investigate the potential antiviral function of *bap1* in response to other viruses are required. The identification of new genes and pathways involved in combatting virus infections in honey bees may present opportunities to develop strategies that mitigate future honey bee colony deaths due to virus infection.

MATERIALS AND METHODS

Age-Matched Live Honey Bees

Honey bee (*Apis mellifera*) colonies were established from packages (∼1.5 kg of worker bees and a naturally-mated queen)

TABLE 2 | Linear model output for the association between *bap1* and *ago2* expression.

Term	Estimate	Standard error	*t*-value	*p*-value
Intercept	0.444	0.25	1.77	0.079
bap1 fold change	0.733	0.092	7.99	7.26×10^{-13}
SINV RNA copies	3.92×10^{-9}	1.39×10^{-9}	2.82	0.0056
*bap1*SINV	-7.23×10^{-10}	3.28×10^{-10}	-2.213	0.0287

Ago2 expression was modeled as a function of bap1 expression coupled with SINV RNA copy number. The model indicates that, after accounting for the interaction between SINV RNA copies and bap1 expression, the expression of bap1 and ago2 are highly associated, and corroborated WGCNA analysis.
Multiple $R^2 = 0.413$, adjusted $R^2 = 0.399$, $F = 29.56_{df1=3, df2=126}$, $p = 1.54 \times 10^{-14}$.

of primarily *Apis mellifera carnica* stock purchased from a commercial producer in Montana in April 2018. Honey bees were kept in Langstroth hives located on Montana State University's Horticulture Farm in Bozeman, MT, US. Colonies were maintained using standard apicultural practices, including bi-monthly evaluation of *Varroa destructor* mite infestation levels using the powdered sugar roll method (121). Colonies were treated with formic acid polysaccharide gel strips (Mite Away Quick Strips®, Nature's Own Design Apiary Products) when mite infestation was >3% [3 mites per 100 bees (121)].

Honey bees for laboratory-based experiments were obtained from brood containing frames with newly emerging bees, which were collected 1 day prior to each experiment and maintained

at 32°C and ∼ 20% relative humidity in a laboratory incubator overnight. Young, age-matched (∼24 h post-emergence), female adult bees were utilized for all experiments. For the duration of the experiment, honey bees were housed in modified deli-containers at 32°C and fed bee candy (powdered sugar mixed with corn syrup until pliable) and water *ad libitum* (33, 36, 47, 122).

Double Stranded RNA Preparation

Double stranded RNA was generated by *in vitro* transcription with T7 RNA polymerase (22, 36). T7 promoter-containing PCR-products were amplified using primers listed in **Supplementary Table S1**, with the following thermocycler program: pre-incubation of 95°C (5 min), 35 cycles of 95°C (30 s), 60°C (30 s), and 72°C (1 min) followed by a final incubation at 72°C (5 min). PCR products were used as template for T7 polymerase transcription (100 μl reactions: NTPs (each 7.5 mM final), RNase OUT (40 units) (Invitrogen), buffer (400 mM HEPES pH 7.5, 120 mM MgCl$_2$, 10 mM spermidine, 200 mM DTT); reactions were carried out at 37°C overnight (8–10 h). DNA was removed by adding 1 unit of RQ1 DNAse (Promega) and incubating for 15 min at 37°C. dsRNA products were ethanol precipitated with 1:10 volume 3M sodium acetate (pH 5.5), suspended in 200 μL RNase-free water, the RNA secondary structure was denatured *via* incubation at 100°C for 5 min, and then complementary RNA strands were annealed by slow cooling to room temperature. The dsRNA products were purified by phenol:chloroform extraction and subsequent ethanol precipitation with a 1:10 volume of 5M ammonium acetate. The precipitated dsRNAs were dissolved in 60–100 μL 10 mM Tris HCl (pH 7.5). The dsRNA quality was assessed by agarose gel electrophoresis and spectrophotometry. The dsRNA was quantified based on the band intensity relative to a standard with ImageJ version 1.50i (123).

Virus Infection and Heat Shock Protocol

Honey bee virus infections were established *via* intra-thoracic injections using glass needles made by pulling borosilicate glass capillary tubes (100 mm long, 1 mL capacity, Kimble-Chase) with a coil temperature of 61°C using a PC-10 Dual-Stage Glass Micropipette Puller (Narishige). Prior to injection, age-matched honey bees (∼24-h post-emergence) were cold anesthetized for 10 min at 4°C. Honey bees were infected with recombinant Sindbis virus expressing green fluorescent protein (SINV-GFP; 3,750 plaque forming units (pfu) in 2 μl 10 mM Tris HCl buffer pH 7.5) (33, 36, 124) *via* intra-thoracic injection using a Harbo syringe (Honey bee Insemination Service) and microcapillary glass needles; mock-infected bees were injected with 2 μL buffer (10 mM Tris HCl, pH 7.5).

Experimental treatment groups that were subjected to temperature stress were intrathoracically injected with either buffer or virus, allowed to recover for 6 h at 32°C, exposed to heat shock (i.e., 42°C for 4 h), and then transferred back to 32°C for the remainder of the study. The heat shock experiments were carried out using three independent honey bee cohorts obtained from three different colonies on distinct dates (i.e., replicate 1 in June 2018, replicate 2 in August 2018, and replicate 3 in July 2019). The data from these sampled are analyzed *post-hoc* from a previously published manuscript (47).

RNA Extraction

Honey bee samples were dissected into head, thorax, and abdomen. The abdomen was chosen for further analysis as it is the primary site of immune cell generating fat bodies and it is distal from the site of injection, and thus virus infection naturally spread to that tissue. Honey bee abdomens were homogenized in 300 μL of deionized water with a sterile steel ball (5 mm) using a Tissue Lyser II (Qiagen) at 30 Hz for 2 min. Then 300 μL of TRIzol reagent (Invitrogen) was added to the homogenate, vortexed for 15 s and incubated at room temperature for 5 min. Next, 60 μL of chloroform was added, samples were shaken by hand for 15 s and incubated on the benchtop for another 2 min. Samples were then centrifuged at 12,000 × g at 4°C for 15 min and the aqueous phase was transferred to a clean centrifuge tube. One volume of isopropanol was added to the aqueous phase, mixed by inversion, and nucleic acid was precipitated by incubation at room temperature for 10 min. The precipitate was pelleted by centrifugation at 12,000 × g at 4°C for 10 min. Pellets were then washed with 500 μL of 75% ethanol and centrifuged at 7,500 × g at 4°C for 5 min, then air dried for 10 min at room temperature and dissolved in 30 μL of deionized water. RNA concentrations and quality were assessed on a Nanodrop 2000 spectrophotometer (Thermo Fisher). When quality was low, RNA was precipitated a second time by addition of four volumes of cold ethanol and 1:10 of a volume 3M sodium acetate (pH 5.5) and incubation at −20°C overnight. Nucleic acids were pelleted by centrifugation at 12,000 × g at 4°C for 10 min and pellets were washed one time with 500 μL 70% ethanol and centrifuged at 12,000 × g at 4°C for 5 min. Pellets were air dried and suspended in 30 μL dH$_2$O. Samples were stored at −80°C until analysis.

Reverse Transcription/cDNA Synthesis

Reverse transcription reactions were performed by incubating 2,000 ng total RNA, 200 units M-MLV reverse-transcriptase (Promega) and 500 ng random hexamer primers (IDT) for 1 h at 37°C, according to the manufacturer's instructions. cDNA was diluted 1:2 and 2 μL was used for PCR or qPCR analysis.

Quantitative Polymerase Chain Reaction

Quantitative PCR (qPCR) was used to analyze the abundance of virus (i.e., SINV-GFP) and the relative abundances of honey bee immune gene and heat shock protein transcripts. All qPCR reactions were performed in triplicate with 2 μL of cDNA template. Each 20 μL reaction contained 1× ChoiceTaq Mastermix (Denville), 0.4 μM each forward and reverse primer, 1× SYBR Green (Life Technologies), and 3 mM MgCl$_2$. A CFX Connect Real Time instrument (BioRad) was used for the following thermo-profile: pre-incubation 95°C for 1 min followed by 40 cycles of 95°C for 10 s, 58°C for 20 s, and 72°C for 15 s, with a final melt curve analysis at 65°C for 5 s to 95°C.

To quantify viral RNA copy numbers in the samples, SINV-GFP plasmid standards were used as templates, with concentrations ranging from 10^3 to 10^9 copies per reaction to create a linear standard curve. The detection limit was

10^3 copies of the SINV cDNA using primers qSindbisFW4495 and qSindbisREV4635. The host gene *Am rpl8* was used as a reference gene to assess the relative expression of two immune genes (*Am bap1* and *Am ago2*). Amplicons were amplified in triplicate per sample for comparison, using primers in **Supplementary Table S1**. Reactions without template were carried out as negative controls. The qPCR specificity was verified through melt point analysis and *via* gel electrophoresis, and all products were previously verified by sequencing. The linear equation for the plasmid standard for SINV was: $Ct = -3.348x + 40.25$ ($R^2 = 0.996$, efficiency = 98.9%) where "x" is the log (SINV genome equivalents). The relative expression of host genes was determined by a ranked $\Delta\Delta Ct$ method in which the ΔCt was calculated by subtracting the *rpl8* Ct value from the Ct of the gene of interest. Then the ΔCt values were ranked to control for natural inter-individual variation in gene expression and the matching mock-infected ΔCt was subtracted from the treatment group ΔCt to obtain the $\Delta\Delta Ct$. The fold-change in cDNA abundance was calculated by the equation $2^{-\Delta\Delta Ct}$.

Phylogenetic Analysis

Bee antiviral protein-1 (*bap1*) (GenBank: MF116383) encodes an antiviral protein. The precise function of Bap1 has yet to be elucidated. Phylogenetic analyses of *bap1* were performed to gain functional insight. To perform a phylogenetic analysis, amino acid and nucleic acid sequences were acquired from NCBI non-redundant protein sequence database by searching using NCBI PSI BLAST function and exporting any sequence with an e-value below 0.001. An amino acid alignment was generated in MEGAX (57) using MUSCLE with default parameters.

A majority rule Bayesian consensus tree of Bap1 orthologs was derived from Bayesian analysis of amino acid alignment implemented in Mr. Bayes v3.2 using a Jones substitution model. Metropolis-coupled Markov Chain Monte Carlo (MCMC) permutation of parameters were initiated with a random tree and involved two runs each with 16 chains set at default temperatures (58, 59). Markov chains were run for 10,000,000 generations and sampled every 10,000 generations. MrBayes identified this sampling rate and a 25% burn-in as sufficient because of the non-autocorrelation of adjacently sampled trees and the complete convergence of the two separate MCMC runs at likelihood stationarity. Numbers on nodes are posterior probabilities (0-1), though posterior probabilities of 1 were not shown to improve figure clarity. The scale bar corresponds to proportion of amino acid change per site.

To perform PAML analysis, nucleic acid accession numbers were retrieved by creating a local BLAST database containing all of the taxa on the amino acid tree and performing BLAST analyses of all the amino acid sequences against the BLASTn database (54, 55). Nucleic acid sequences were retrieved by submitting the accession numbers to NCBI's batch ENTREZ (125). Sequences were again aligned by codon in MEGAX using MUSCLE with default parameters. Majority rule Bayesian consensus trees of *bap1* and *RPB1* orthologs were derived from a Bayesian analysis of a nucleic acid codon alignments implemented in Mr. Bayes v3.2 using the protein nucleic acid model. In this model, MrBayes first translates codons into protein and then infers relationships. Metropolis-coupled Markov Chain Monte Carlo (MCMC) permutation of parameters were initiated with a random tree and involved two runs each with 16 chains set at default temperatures (58, 59). Markov chains were run for 5,000,000 generations and sampled every 10,000 generations. MrBayes identified this sampling rate and a 25% burn-in as sufficient because of the non-autocorrelation of adjacently sampled trees and the complete convergence of the two separate MCMC runs at likelihood stationarity. Numbers on nodes are posterior probabilities (0-1), though posterior probabilities of 1 were left off to improve clarity. Generated trees were then partially edited in FigTree v 1.4.3 (126) and Adobe Illustrator to further improve clarity. The scale bar corresponds to proportional changes per codon site. Accession numbers are included on the branch tips and in **Supplementary Tables S4**, **S5**.

The codon alignments and the majority rule Bayesian consensus trees generated with this method were then used for selection analysis using the CODEML package in Phylogenetic Analysis by Maximum Likelihood (PAML) 4.9 (66). Relative rates of non-synonymous substitution (parameter ω) were estimated using fixed branch lengths for models M0, M1, and M2a. All models were run twice to check convergence. The log likelihood ratio test (LRT) statistic was calculated as twice the log likelihood difference between a test and null models ($2\Delta lnL$) which was then compared against the χ^2 distribution with critical values of 3.841 (M1-M0), 5.99 (M2-M1), and 7.814 (M2a-M0) at an α level of 0.05. PAML's Bayes Empirical Bayes (BEB) method was used on M2a to calculate the posterior probability that each amino site along Bap1 belonged to each site class (ω < 1, ω = 1, ω > 1). The dN/dS ratios calculated by the BEB method at each site containing an amino acid were then plotted as a stacked bar chart (**Figure 2B**; **Supplementary Figure S2B**). Site-specific dN/dS ratios were plotted and those with posterior probability of belonging to a given site class greater than an arbitrary cutoff of 0.8 were color coded as evolving under positive (red), neutral (blue) or negative (green) selection (**Supplementary Figures S3**, **S4**). Raw BEB probabilities and dN/dS ratios listed in **Supplementary Tables S5**, **S8**.

Honey Bee RNAseq Data and Analysis

Honey bee transcriptome sequence data were generated and described in Brutscher et al. (36). Sequence data were deposited into NCBI Sequence Read Archive under accession number SRP101337 and is linked with NCBI BioProject #PRJNA377749. In brief, RNA samples obtained from individual honey bees (*n* = 47) from bees that were mock-infected, SINV-infected, dsRNA-treated, SINV + non-specific dsRNA, treated and SINV+ specific dsRNA treated. Samples were processed and the normalized number of Fragments Per Kilobase of transcript per Million mapped reads (FPKM) was determined using CuffDiff as described in Brutscher et al. (36, 127, 128) (**Supplementary Table S10**). FPKM values were $\log_2$ transformed and used for Weighted Gene Correlation Network Analysis (WGCNA).

Weighted Gene Correlation Network Analysis

To identify genes are that co-expressed with *bap1* a weighted gene correlation network was constructed using the package WGCNA (86, 87) in R 4.0.2 from data analyzed *post-hoc* (36). WGCNA uses an unsupervised hierarchical clustering algorithm that groups genes into modules based on similarity of expression. In this case WGCNA analysis was performed on $\log_2$ transformed transcript FPKM values from three individuals in several treatment groups at several time points as described above. Recommended settings were used for the analysis (86, 87). However, briefly, minimum module size was set to 20 and deep split was set to 2. A signed correlation matrix was generated using default parameters, and a signed gene correlation network was constructed using a soft power threshold of 10. Modules were merged based on a branch cut height of 0.25. Trait data were used to identify module eigengenes differentially expressed in response to various traits (e.g., SINV RNA copies or whether or not the individual was injected with dsRNA) and plotted as a heat map (**Supplementary Figure S5**; **Supplementary Tables S13–S20**). The module that MF116383 belonged to was manually identified and pared down to include only the strongest connections with correlations above 0.5 or below −0.5. Then this network was imported into Cytoscape v 3.8.2 (129) for visualization and the *bap1* subnetwork was isolated manually (**Supplementary Figure S6**). For clarity, genes with no predicted function were removed from the network (**Figure 3C**). All remaining connections in this subnetwork are positive correlations above 0.5.

Statistical Analysis

All statistical analyses were carried out in R 4.0.2, except the likelihood-ratio test statistics to compare models of selection which were calculated manually by subtracting the null model from the test model and multiplying by 2 ($2\Delta\ln L$). ANOVAs and linear models were constructed using the base R package (130, 131). Partial Pearson's Correlation coefficients were calculated using the *ppcor* package in R (132). Because our dataset contains zero values in our SINV measurement (untreated bees), SINV levels were $\log_{10}$ transformed prior to calculating correlation coefficients to improve homoscedasticity and improve normality.

DATA AVAILABILITY STATEMENT

Sequence data were deposited into NCBI Sequence Read Archive under Accession Number SRP101337 and is linked with NCBI BioProject PRJNA377749.

AUTHOR CONTRIBUTIONS

AJM and MLF: conceptualization, formal analysis, writing—original draft preparation, writing—review and editing, and visualization. AJM, KFD, and MLF: methodology. AJM, LMB, KFD, and MLF: investigation. MLF: resources, supervision, project administration, and funding acquisition. AJM, LMB, and MLF: data curation. All authors have read and agreed to the published version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/finsc.2021.749781/full#supplementary-material

REFERENCES

1. Müller WEG, Müller IM. Origin of the metazoan immune system: identification of the molecules and their functions in sponges. *Integr Comp Biol.* (2003) 43:281–92. doi: 10.1093/icb/43.2.281
2. Cooper MD, Alder MN. The evolution of adaptive immune systems. *Cell.* (2006) 124:815–22. doi: 10.1016/j.cell.2006.02.001
3. Ronald PC, Beutler B. Plant and animal sensors of conserved microbial signatures. *Science.* (2010) 330:1061–4. doi: 10.1126/science.1189468
4. Kingsolver MB, Hardy RW. Making connections in insect innate immunity. *PNAS.* (2012) 109:18639–40. doi: 10.1073/pnas.1216736109
5. Kingsolver MB, Huang Z, Hardy RW. Insect antiviral innate immunity: pathways, effectors, and connections. *J Mol Biol.* (2013) 425:4921–36. doi: 10.1016/j.jmb.2013.10.006
6. McMenamin AJ, Daughenbaugh KF, Parekh F, Pizzorno MC, Flenniken ML. Honey bee and bumble bee antiviral defense. *Viruses.* (2018) 10:395. doi: 10.3390/v10080395
7. Klein A-M, Vaissière BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, et al. (2007). Importance of pollinators in changing landscapes for world crops. *Proc Biol Sci.* 274:303–13. doi: 10.1098/rspb.2006.3721
8. Gallai N, Salles J-M, Settele J, Vaissière BE. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecoll Econ.* (2009) 68:810–21. doi: 10.1016/j.ecolecon.2008.06.014
9. Eilers EJ, Kremen C, Smith Greenleaf S, Garber AK, Klein A-M. Contribution of pollinator-mediated crops to nutrients in the human food supply. *PLoS ONE.* (2011) 6:e21363. doi: 10.1371/journal.pone.0021363
10. Ferrandon D, Imler JL, Hetru C, Hoffmann JA. The *Drosophila* systemic immune response: sensing and signalling during bacterial and fungal infections. *Nat Rev Immunol.* (2007) 7:862–74. doi: 10.1038/nri2194

11. Hoffmann JA. The immune response of *Drosophila*. *Nature*. (2003) 426:33–8. doi: 10.1038/nature02021

12. Buchon N, Silverman N, Cherry S. Immunity in *Drosophila melanogaster*-from microbial recognition to whole-organism physiology. *Nat Rev Immunol*. (2014) 14:796–810. doi: 10.1038/nri3763

13. Brutscher LM, Flenniken ML. (2015). RNAi and antiviral defense in the honey bee. *J Immunol Res*. (2015) 2015:941897. doi: 10.1155/2015/941897

14. Galiana-Arnoux D, Dostert C, Schneemann A, Hoffmann JA, Imler J-L. Essential function *in vivo* for Dicer-2 in host defense against RNA viruses in *Drosophila*. *Nat Immunol*. (2006) 7:590–7. doi: 10.1038/ni1335

15. van Rij RP, Saleh M-C, Berry B, Foo C, Houk A, Antoniewski C, et al. The RNA silencing endonuclease argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. *Genes Dev*. (2006) 20:2985–95. doi: 10.1101/gad.1482006

16. Wang X-H, Aliyari R, Li W-X, Li H-W, Kim K, Carthew R, et al. RNA Interference directs innate immunity against viruses in adult *Drosophila*. *Science*. (2006) 312:452–4. doi: 10.1126/science.1125694

17. Zambon RA, Vakharia VN, Wu LP. RNAi is an antiviral immune response against a dsRNA virus in *Drosophila melanogaster*. *Cell Microbiol*. (2006) 8:880–9. doi: 10.1111/j.1462-5822.2006.00688.x

18. Deddouche S, Matt N, Budd A, Mueller S, Kemp C, Galiana-Arnoux D, et al. The DExD/H-box helicase Dicer-2 mediates the induction of antiviral activity in *Drosophila*. *Nat Immunol*. (2008) 9:1425–32. doi: 10.1038/ni.1664

19. Myles KM, Wiley MR, Morazzani EM, Adelman ZN. Alphavirus-derived small RNAs modulate pathogenesis in disease vector mosquitoes. *PNAS*. (2008) 105:19938. doi: 10.1073/pnas.0803408105

20. Bronkhorst AW, van Rij RP. The long and short of antiviral defense: small RNA-based immunity in insects. *Curr Opin Virol*. (2014) 7:19–28. doi: 10.1016/j.coviro.2014.03.010

21. Brennan CA, Anderson KV. *Drosophila*: the genetics of innate immune recognition and response. *Annu Rev Immunoly*. (2004) 22:457–83. doi: 10.1146/annurev.immunol.22.012703.104626

22. Saleh MC, Tassetto M, van Rij RP, Goic B, Gausson V, Berry B, et al. Antiviral immunity in *Drosophila* requires systemic RNA interference spread. *Nature*. (2009) 458:346–50. doi: 10.1038/nature07712

23. Tassetto M, Kunitomi M, Andino R. Circulating immune cells mediate a systemic RNAi-based adaptive antiviral response in *Drosophila*. *Cell*. (2017) 169:314–25.e313. doi: 10.1016/j.cell.2017.03.033

24. Poirier EZ, Goic B, Tomé-Poderti L, Frangeul L, Boussier J, Gausson V, et al. Dicer-2-dependent generation of viral DNA from defective genomes of RNA viruses modulates antiviral immunity in insects. *Cell Host Microbe*. (2018) 23:353–65.e358. doi: 10.1016/j.chom.2018.02.001

25. Suzuki Y, Baidaliuk A, Miesen P, Frangeul L, Crist AB, Merkling SH, et al. Non-retroviral endogenous viral element limits cognate virus replication in *Aedes aegypti* ovaries. *Curr Biol*. (2020) 30:3495–06.e3496. doi: 10.1016/j.cub.2020.06.057

26. Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson NR, et al. Antagonistic coevolution accelerates molecular evolution. *Nature*. (2010) 464:275–8. doi: 10.1038/nature08798

27. Daugherty MD, Malik HS. Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet*. (2012) 46:677–700. doi: 10.1146/annurev-genet-110711-155522

28. Obbard DJ, Jiggins FM, Halligan DL, Little TJ. Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr Biol*. (2006) 16:580–5. doi: 10.1016/j.cub.2006.01.065

29. Palmer WH, Hadfield JD, Obbard DJ. RNA-interference pathways display high rates of adaptive protein evolution in multiple invertebrates. *Genetics*. (2018) 208:1585–99. doi: 10.1534/genetics.117.300567

30. Maori E, Paldi N, Shafir S, Kalev H, Tsur E, Glick E, et al. IAPV, a bee-affecting virus associated with colony collapse disorder can be silenced by dsRNA ingestion. *Insect Mol Biol*. (2009) 18:55–60. doi: 10.1111/j.1365-2583.2009.00847.x

31. Liu X, Zhang Y, Yan X, Han R. Prevention of chinese sacbrood virus infection in apis cerana using rna interference. *Curr Microbiol*. (2010) 61:422–8. doi: 10.1007/s00284-010-9633-2

32. Desai SD, Eu YJ, Whyard S, Currie RW. Reduction in deformed wing virus infection in larval and adult honey bees (*Apis mellifera*

L.) by double-stranded RNA ingestion. *Insect Mol Biol*. (2012) 21:446–55. doi: 10.1111/j.1365-2583.2012.01150.x

33. Flenniken ML, Andino R. Non-specific dsRNA-mediated antiviral response in the honey bee. *PLoS ONE*. (2013) 8:e77263. doi: 10.1371/journal.pone.0077263

34. Chejanovsky N, Ophir R, Schwager MS, Slabezki Y, Grossman S, Cox-Foster D. Characterization of viral siRNA populations in honey bee colony collapse disorder. *Virology*. (2014) 454-455:176–83. doi: 10.1016/j.virol.2014.02.012

35. Niu J, Meeus I, Cappelle K, Piot N, Smagghe G. The immune response of the small interfering RNA pathway in the defense against bee viruses. *Curr Opin Insect Sci*. (2014) 6:22–7. doi: 10.1016/j.cois.2014.09.014

36. Brutscher LM, Daughenbaugh KF, Flenniken ML. Virus and dsRNA-triggered transcriptional responses reveal key components of honey bee antiviral defense. *Sci Rep*. (2017) 7:6448. doi: 10.1038/s41598-017-06623-z

37. Leonard SP, Powell JE, Perutka J, Geng P, Heckmann LC, Horak RD, et al. Engineered symbionts activate honey bee immunity and limit pathogens. *Science*. (2020) 367:573. doi: 10.1126/science.aax9039

38. Piot N, Snoeck S, Vanlede M, Smagghe G, Meeus I. The effect of oral administration of dsRNA on viral replication and mortality in *Bombus terrestris*. *Viruses*. (2015) 7:3172–85. doi: 10.3390/v7062765

39. Cappelle K, Smagghe G, Dhaenens M, Meeus I. Israeli acute paralysis virus infection leads to an enhanced RNA interference response and not its suppression in the bumblebee *Bombus terrestris*. *Viruses*. (2016) 8:334. doi: 10.3390/v8120334

40. Niu J, Smagghe G, De Coninck DI, Van Nieuwerburgh F, Deforce D, Meeus I. *In vivo* study of Dicer-2-mediated immune response of the small interfering RNA pathway upon systemic infections of virulent and avirulent viruses in *Bombus terrestris*. *Insect Biochem Mol Biol*. (2016) 70:127–37. doi: 10.1016/j.ibmb.2015.12.006

41. Niu J, Meeus I, Smagghe G. Differential expression pattern of Vago in bumblebee (*Bombus terrestris*), induced by virulent and avirulent virus infections. *Sci Rep*. (2016) 6:34200. doi: 10.1038/srep34200

42. Wang H, Meeus I, Smagghe G. Israeli acute paralysis virus associated paralysis symptoms, viral tissue distribution and Dicer-2 induction in bumblebee workers (*Bombus terrestris*). *J Gen Virol*. (2016) 97:1981–9. doi: 10.1099/jgv.0.000516

43. Niu J, Meeus I, Coninck DIMD, Deforce D, Etebari K, Asgari S, et al. Infections of virulent and avirulent viruses differentially influenced the microRNAs in *Bombus terrestris*. *Sci Rep*. (2017) 7:1–11. doi: 10.1038/srep45620

44. Wang H, Smagghe G, Meeus I. The role of a single gene encoding the Single von Willebrand factor C-domain protein (SVC) in bumblebee immunity extends beyond antiviral defense. *Insect Biochem Mol Biol*. (2017) 91:10–20. doi: 10.1016/j.ibmb.2017.10.002

45. Elsik CG, Worley KC, Bennett AK, Beye M, Camara F, Childers CP, et al. Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genom*. (2014) 15:86. doi: 10.1186/1471-2164-15-86

46. Venticinque L, Meruelo D. Sindbis viral vector induced apoptosis requires translational inhibition and signaling through Mcl-1 and Bak. *Mol Cancer*. (2010) 9:1–16. doi: 10.1186/1476-4598-9-37

47. McMenamin AJ, Daughenbaugh KF, Flenniken ML. The heat shock response in the Western honey bee (*Apis mellifera*) is antiviral. *Viruses*. (2020) 12:245. doi: 10.3390/v12020245

48. Lamp B, Url A, Seitz K, Eichhorn J, Riedel C, Sinn LJ, et al. Construction and rescue of a molecular clone of deformed wing virus (DWV). *PLoS ONE*. (2016) 11:e0164639. doi: 10.1371/journal.pone.0164639

49. Seitz K, Buczolich K, Dikunová A, Plevka P, Power K, Rümenapf T, et al. A molecular clone of chronic bee paralysis virus (CBPV) causes mortality in honey bee pupae (*Apis mellifera*). *Sci Rep*. (2019) 9:16274. doi: 10.1038/s41598-019-52822-1

50. Ryabov EV, Christmon K, Heerman MC, Posada-Florez F, Harrison RL, Chen Y, et al. Development of a honey bee RNA virus vector based on the genome of a deformed wing virus. *Viruses*. (2020) 12:374. doi: 10.3390/v12040374

51. Ryabov EV, Wood GR, Fannon JM, Moore JD, Bull JC, Chandler D, et al. A virulent strain of deformed wing virus (DWV) of honeybees (*Apis mellifera*) prevails after varroa destructor-mediated, or *in vitro*, transmission. *PLoS Path*. (2014) 10:e1004230. doi: 10.1371/journal.ppat.1004230

52. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol.* (2018) 430:2237–43. doi: 10.1016/j.jmb.2017.12.007

53. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, et al. Protein sequence analysis using the MPI bioinformatics toolkit. *Curr Protoc Bioinform.* (2020) 72:e108. doi: 10.1002/cpbi.108

54. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* (1990) 215:403–10. doi: 10.1016/S0022-2836(05)80360-2

55. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* (1997) 25:3389–402. doi: 10.1093/nar/25.17.3389

56. Gertz E. *BLAST Scoring Parameters.* Austin, TX: Oden Institute in Austin (2005).

57. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* (2018) 35:1547–9. doi: 10.1093/molbev/msy096

58. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinform.* (2001) 17:754–5. doi: 10.1093/bioinformatics/17.8.754

59. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinform.* (2003) 19:1572–4. doi: 10.1093/bioinformatics/btg180

60. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science.* (2014) 346:763. doi: 10.1126/science.1257570

61. Lozano J, Gomez-Orte E, Lee H-J, Belles X. Super-induction of Dicer-2 expression by alien double-stranded RNAs: an evolutionary ancient response to viral infection? *Dev Genes Evol.* (2012) 222:229–35. doi: 10.1007/s00427-012-0404-x

62. Pitaluga AN, Mason PW, Traub-Cseko YM. Non-specific antiviral response detected in RNA-treated cultured cells of the sandfly, Lutzomyia longipalpis. *Dev Comp Immunol.* (2008) 32:191–7. doi: 10.1016/j.dci.2007.06.008

63. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinform.* (2010) 11:431. doi: 10.1186/1471-2105-11-431

64. Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, et al. Evolutionary history of the hymenoptera. *Curr Biol.* (2017) 27:1013–8. doi: 10.1016/j.cub.2017.01.027

65. Koleske AJ, Young RA. The RNA polymerase II holoenzyme and its implications for gene regulation. *Trends Biochem Sci.* (1995) 20:113–6. doi: 10.1016/S0968-0004(00)88977-X

66. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* (2007) 24:1586–91. doi: 10.1093/molbev/msm088

67. Lin YS, Hsu WL, Hwang JK, Li WH. Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Mol Biol Evol.* (2007) 24:1005–11. doi: 10.1093/molbev/msm019

68. Brown CJ, Johnson AK, Daughdrill GW. Comparing models of evolution for ordered and disordered proteins. *Mol Biol Evol.* (2010) 27:609–21. doi: 10.1093/molbev/msp277

69. Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and disorder. *Curr Opin Struct Biol.* (2011) 21:441–6. doi: 10.1016/j.sbi.2011.02.005

70. Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* (2007) 35:W460–4. doi: 10.1093/nar/gkm363

71. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol.* (2014) 16:18. doi: 10.1038/nrm3920

72. Xia Y, Franzosa EA, Gerstein MB. Integrated assessment of genomic correlates of protein evolutionary rate. *PLoS Comput Biol.* (2009) 5:e1000413. doi: 10.1371/journal.pcbi.1000413

73. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* (2005) 272:5129–48. doi: 10.1111/j.1742-4658.2005.04948.x

74. Higurashi M, Ishida T, Kinoshita K. Identification of transient hub proteins and the possible structural basis for their multiple interactions. *Prot Sci.* (2008) 17:72–8. doi: 10.1110/ps.073196308

75. Kim PM, Sboner A, Xia Y, Gerstein M. The role of disorder in interaction networks: a structural analysis. *Mol Syst Biol.* (2008) 4:179. doi: 10.1038/msb.2008.16

76. Spolar RS, Record MT Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science.* (1994) 263:777–84. doi: 10.1126/science.8303294

77. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol.* (2002) 12:54–60. doi: 10.1016/S0959-440X(02)00289-0

78. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* (2019) 37:420–3. doi: 10.1038/s41587-019-0036-z

79. Nothwehr SF, Gordon JI. Targeting of proteins into the eukaryotic secretory pathway: signal peptide structure/function relationships. *Bioessays.* (1990) 12:479–84. doi: 10.1002/bies.950121005

80. von Heijne G. The signal peptide. *J Membr Biol.* (1990) 115:195–201. doi: 10.1007/BF01868635

81. Coleman J, Inukai M, Inouye M. Dual functions of the signal peptide in protein transfer across the membrane. *Cell.* (1985) 43:351–60. doi: 10.1016/0092-8674(85)90040-6

82. Kapp K, Schrempf S, Lemberg M, Bernhard D. Post-targeting functions of signal peptides. In: Zimmermann R, editor. *Protein Transport into the Endoplasmic Reticulum.* Austin, TX: Madame Curie Bioscience Database: Landes Bioscience (2009).

83. Meyer DI. Protein translocation into the endoplasmic reticulum: a light at the end of the tunnel. *Trends Cell Biol.* (1991) 1:154–9. doi: 10.1016/0962-8924(91)90016-3

84. Bonifacino JS, Traub LM. Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu Rev Biochem.* (2003) 72:395–447. doi: 10.1146/annurev.biochem.72.121801.161800

85. Tokarev AA, Alfonso A, Segev N. Intracellular compartments and trafficking pathways. In: Segev N, editor. *Trafficking Inside Cells: Pathways, Mechanisms and Regulation.* Austin, TX: Madame Curie Bioscience Database (2009).

86. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* (2008) 9:559. doi: 10.1186/1471-2105-9-559

87. Langfelder P, Horvath S. Fast R functions for robust correlations and hierarchical clustering. *J Stat Softw.* (2012) 46:17. doi: 10.18637/jss.v046.i11

88. Mattick JS. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* (2001) 2:986–91. doi: 10.1093/embo-reports/kve230

89. Jackson RJ, Standart N. How do microRNAs regulate gene expression? *Science.* (2007) 367:re1. doi: 10.1126/stke.3672007re1

90. Fernandes JCR, Acuña SM, Aoki JI, Floeter-Winter LM, Muxel SM. Long non-coding RNAs in the regulation of gene expression: physiology and disease. *Noncoding RNA.* (2019) 5:17. doi: 10.3390/ncrna5010017

91. Gil N, Ulitsky I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat Rev Genet.* (2020) 21:102–17. doi: 10.1038/s41576-019-0184-5

92. Bruscella P, Bottini S, Baudesson C, Pawlotsky JM, Feray C, Trabucchi M. Viruses and miRNAs: more friends than foes. *Front Microbiol.* (2017) 8:824. doi: 10.3389/fmicb.2017.00824

93. Millard SS, Flanagan JJ, Pappu KS, Wu W, Zipursky SL. Dscam2 mediates axonal tiling in the *Drosophila* visual system. *Nature.* (2007) 447:720–4. doi: 10.1038/nature05855

94. Lah GJ-E, Li JSS, Millard SS. Cell-specific alternative splicing of *Drosophila* Dscam2 is crucial for proper neuronal wiring. *Neuron.* (2014) 83:1376–88. doi: 10.1016/j.neuron.2014.08.002

95. Armitage SA, Peuss R, Kurtz J. Dscam and pancrustacean immune memory - a review of the evidence. *Dev Comp Immunol.* (2015) 48:315–23. doi: 10.1016/j.dci.2014.03.004

96. Armitage SAO, Kurtz J, Brites D, Dong Y, Du Pasquier L, Wang HC. Dscam1 in pancrustacean immunity: current status and a look to the future. *Front Immunol.* (2017) 8:662. doi: 10.3389/fimmu.2017.00662

97. Haddad N, Mahmud Batainh A, Suleiman Migdadi O, Saini D, Krishnamurthy V, Parameswaran S, et al. Next generation sequencing of

Apis mellifera syriaca identifies genes for Varroa resistance and beneficial bee keeping traits. *Insect Sci.* (2016) 23:579–90. doi: 10.1111/1744-7917.12205

98. Zhao P, Xia F, Jiang L, Guo H, Xu G, Sun Q, et al. Enhanced antiviral immunity against Bombyx mori cytoplasmic polyhedrosis virus *via* overexpression of peptidoglycan recognition protein S2 in transgenic silkworms. *Dev Comp Immunol.* (2018) 87:84–9. doi: 10.1016/j.dci.2018.05.021

99. Zambon RA, Nandakumar M, Vakharia VN, Wu LP. The Toll pathway is important for an antiviral response in *Drosophila*. *PNAS.* (2005) 102:7257–62. doi: 10.1073/pnas.0409181102

100. Avadhanula V, Weasner BP, Hardy GG, Kumar JP, Hardy RW. A novel system for the launch of alphavirus RNA synthesis reveals a role for the Imd pathway in arthropod antiviral response. *PLoS Path.* (2009) 5:e1000582. doi: 10.1371/journal.ppat.1000582

101. Costa A, Jan E, Sarnow P, Schneider D. The Imd pathway is involved in antiviral immune responses in *Drosophila*. *PLoS ONE.* (2009) 4. doi: 10.1371/journal.pone.0007436

102. Rand TA, Ginalski K, Grishin NV, Wang X. Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. *Proc Natl Acad Sci USA.* (2004) 101:14385–9. doi: 10.1073/pnas.0405913101

103. Tomari Y, Matranga C, Haley B, Martinez N, Zamore PD. A protein sensor for siRNA asymmetry. *Science.* (2004) 306:1377. doi: 10.1126/science.1102755

104. Kim K, Lee YS, Carthew RW. Conversion of pre-RISC to holo-RISC by Ago2 during assembly of RNAi complexes. *RNA.* (2007) 13:22–9. doi: 10.1261/rna.283207

105. Blandin S, Levashina EA. Thioester-containing proteins and insect immunity. *Mol Immunol.* (2004) 40:903–8. doi: 10.1016/j.molimm.2003.10.010

106. Shokal U, Eleftherianos I. Evolution and function of thioester-containing proteins and the complement system in the innate immune response. *Front Immunol.* (2017) 8:759. doi: 10.3389/fimmu.2017.00759

107. Cheng G, Liu L, Wang P, Zhang Y, Zhao YO, Colpitts TM, et al. An *in vivo* transfection approach elucidates a role for aedes aegypti thioester-containing proteins in flaviviral infection. *PLoS ONE.* (2011) 6:e22786. doi: 10.1371/journal.pone.0022786

108. Xiao X, Liu Y, Zhang X, Wang J, Li Z, Pang X, et al. Complement-related proteins control the flavivirus infection of *Aedes aegypti* by inducing antimicrobial peptides. *PLoS Path.* (2014) 10:e1004027. doi: 10.1371/journal.ppat.1004027

109. Caudy AA, Ketting RF, Hammond SM, Denli AM, Bathoorn AMP, Tops BBJ, et al. A micrococcal nuclease homologue in RNAi effector complexes. *Nature.* (2003) 425:411–4. doi: 10.1038/nature01956

110. Schwarz DS, Tomari Y, Zamore PD. The RNA-induced silencing complex is a Mg2+-dependent endonuclease. *Curr Biol.* (2004) 14:787–91. doi: 10.1016/j.cub.2004.03.008

111. Scadden ADJ. The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nature Struct Mol Biol.* (2005) 12:489–96. doi: 10.1038/nsmb936

112. Iuchi S. Three classes of C2H2 zinc finger proteins. *Cell Mol Life Sci.* (2001) 58:625–35. doi: 10.1007/PL00000885

113. Stubbs L, Sun Y, Caetano-Anolles D. Function and evolution of C2H2 zinc finger arrays. In: Harris RJ, editor. *A Handbook of Transcription Factors.* Dordrecht: Springer (2011).

114. Wang G, Zheng C. Zinc finger proteins in the host-virus interplay: multifaceted functions based on their nucleic acid-binding property. *FEMS Microbiol Rev.* (2021) 45:fuaa059. doi: 10.1093/femsre/fuaa059

115. Nazzi F, Brown SP, Annoscia D, Del Piccolo F, Di Prisco G, Varricchio P, et al. Synergistic parasite-pathogen interactions mediated by host immunity can drive the collapse of honeybee colonies. *PLoS Path.* (2012) 8:e1002735. doi: 10.1371/journal.ppat.1002735

116. Di Prisco G, Cavaliere V, Annoscia D, Varricchio P, Caprio E, Nazzi F, et al. Neonicotinoid clothianidin adversely affects insect immunity and promotes replication of a viral pathogen in honey bees. *PNAS.* (2013) 110:18466–71. doi: 10.1073/pnas.1314923110

117. Chen YP, Pettis JS, Corona M, Chen WP, Li CJ, Spivak M, et al. Israeli acute paralysis virus: epidemiology, pathogenesis and implications for honey bee health. *PLoS Path.* (2014) 10:e1004261. doi: 10.1371/journal.ppat.1004261

118. Galbraith DA, Yang X, Nino EL, Yi S, Grozinger C. Parallel epigenomic and transcriptomic responses to viral infection in honey bees (*Apis mellifera*). *PLoS Pathog.* (2015) 11:e1004713. doi: 10.1371/journal.ppat.1004713

119. Ryabov EV, Fannon JM, Moore JD, Wood GR, Evans DJ. The iflaviruses sacbrood virus and deformed wing virus evoke different transcriptional responses in the honeybee which may facilitate their horizontal or vertical transmission. *PeerJ.* (2016) 4:e1591. doi: 10.7717/peerj.1591

120. Brutscher LM, Daughenbaugh KF, Flenniken ML. Antiviral defense mechanisms in honey bees. *Curr Opin Insect Sci.* (2015) 2:1–12. doi: 10.1016/j.cois.2015.04.016

121. Milbrath M. *Varroa Mite Monitoring: Using a Sugar Roll to Identify Populations of Varroa Destructor in Honey Bee Colonies.* (2018). pollinators.msu.edu. Available onlin at: https://pollinators.msu.edu/resources/beekeepers/varroa-mite-monitoring1/ (accessed January 2, 2020).

122. Evans JD, Chen YP, Prisco Gd, Pettis J, Williams V. Bee cups: single-use cages for honey bee experiments. *J Apic Res.* (2009) 48:300–2. doi: 10.1080/00218839.2009.11101548

123. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Meth.* (2012) 9:671–5. doi: 10.1038/nmeth.2089

124. Hahn CS, Hahn YS, Braciale TJ, Rice CM. Infectious Sindbis virus transient expression vectors for studying antigen processing and presentation. *PNAS.* (1992) 89:2679. doi: 10.1073/pnas.89.7.2679

125. Batch Entrez. In: Ganten D, Ruckpaul K, editors. *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine.* Berlin, Heidelberg: Springer Berlin Heidelberg (2006). p. 131.

126. Rambaut A. *FigTree-Version 1.4.3, a Graphical Viewer of Phylogenetic Trees.* Edinburgh: GitHub (2017).

127. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* (2012) 7:562–78. doi: 10.1038/nprot.2012.016

128. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol.* (2013) 31:46–53. doi: 10.1038/nbt.2450

129. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498–504. doi: 10.1101/gr.1239303

130. Team R. *RStudio: Integrated Development Environment for R.* 1.1.456 ed. Vienna, Austria: RStudio, Inc. (2016).

131. Team RC. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing (2019).

132. Kim S. ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *CSAM.* (2015) 22:665–74. doi: 10.5351/CSAM.2015.22.6.665

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.