

## RESEARCH

# Transcriptomic responses to diet quality and viral infection in *Apis mellifera*

Lindsay Rutter<sup>1</sup>, Jimena Carrillo-Tripp<sup>2</sup>, Bryony C. Bonning<sup>3</sup>, Dianne Cook<sup>4</sup>, Amy L. Toth<sup>5,6</sup> and Adam G. Dolezal<sup>7\*</sup>

\*Correspondence:

adolezal@illinois.edu

<sup>7</sup>Department of Entomology,

University of Illinois at

Urbana-Champaign, Urbana, IL

61801, USA

Full list of author information is available at the end of the article

## Abstract

**Background:** Parts of Europe and the United States have witnessed dramatic losses in commercially managed honey bees over the past decade to what is considered an unsustainable extent. The large-scale loss of bees has considerable implications for the agricultural economy because bees are one of the leading pollinators of numerous crops. Bee declines have been associated with several interactive factors. Recent studies suggest nutritional and pathogen stress can interactively contribute to bee physiological declines, but the molecular mechanisms underlying interactive effects remain unknown. In this study, we provide insight into this question by using RNA-sequencing to examine how monofloral diets and Israeli acute paralysis virus inoculation influence gene expression patterns in bees.

**Results:** We found a considerable nutritional response, with almost 2,000 transcripts changing with diet quality. The majority of these genes were over-represented for nutrient signaling (insulin resistance) and immune response (Notch signaling and JaK-STAT pathways). In our experimental conditions, the transcriptomic response to viral infection was fairly limited. We only found 43 transcripts to be differentially expressed, some with known immune functions (argonaute-2), transcriptional regulation, and muscle contraction. We created contrasts to explore whether protective mechanisms of good diet were due to direct effects on immune function (resistance) or indirect effects on energy availability (tolerance). A similar number of resistance and tolerance candidate differentially expressed genes were found, suggesting both processes may play significant roles in dietary buffering from pathogen infection.

**Conclusions:** Through transcriptional contrasts and functional enrichment analysis, we contribute to our understanding of the mechanisms underlying feedbacks between nutrition and disease in bees. We also show that comparing results derived from combined analyses across multiple RNA-seq studies may allow researchers to identify transcriptomic patterns in bees that are concurrently less artificial and less noisy. This work underlines the merits of using data visualization techniques and multiple datasets to interpret RNA-sequencing studies.

**Keywords:** Honey bee; RNA-sequencing; Israeli acute paralysis virus; Monofloral pollen; Visualization

## 1 Background

2 Managed honey bees have undergone health declines in the United States and parts  
3 of Europe over the past decade [1, 2, 3], with annual mortality rates exceeding what  
4 beekeepers consider sustainable [4, 5]. More than 70 percent of major global food  
5 crops (including fruits, vegetables, and nuts) at least benefit from pollination, and  
6 yearly insect pollination services are valued worldwide at \$175 billion [6]. As honey  
7 bees are largely considered to be the leading pollinator of numerous crops, their  
8 marked loss has considerable implications for agricultural sustainability [7].

9 Honey bee declines have been associated with several factors, including pesticide  
10 use, parasites, pathogens, habitat loss, and poor nutrition [8, 9]. Researchers gen-  
11 erally agree that these stressors do not act in isolation; instead, they appear to  
12 influence the large-scale loss of honey bees in an interactive fashion as the environ-  
13 ment changes [10]. Nutrition and viral infection are two factors that pose heightened  
14 dangers to honey bee health in response to recent environmental changes. Interac-  
15 tions between nutrition and viral infection may create feedbacks that impact bee  
16 health through several mechanisms [11, 12].

17 Pollen is a main source of nutrition (including proteins, amino acids, lipids, sterols,  
18 starch, vitamins, and minerals) in honey bees [13, 14]. At the individual level, pollen  
19 supplies most of the nutrients necessary for physiological development [15] and is  
20 believed to have considerable impact on longevity [16]. At the colony level, pollen  
21 enables young workers to produce jelly, which then nourishes larvae, drones, older  
22 workers, and the queen [17, 18]. Various environmental changes (including urban-  
23 ization and monoculture crop production) have significantly altered the nutritional  
24 profile available to honey bees. In particular, honey bees are confronted with a  
25 less diverse selection of pollen, which is of concern because mixed-pollen (polyflo-  
26 ral) diets are generally considered healthier than single-pollen (monofloral) diets  
27 [19, 20, 21]. Indeed, reported colony mortality rates are higher in developed land

28 areas compared to undeveloped land areas [22], and beekeepers rank poor nutrition  
29 as one of the main reasons for colony losses [23]. Understanding how low diversity  
30 diets affect honey bee health will be crucial to resolve problems that may arise as  
31 agriculture continues to intensify throughout the world [24, 25].

32 Viral infection was a comparatively minor problem in honey bees until the last  
33 century when the ectoparasitic varroa mite (*Varroa destructor*) spread worldwide  
34 [26]. This mite feeds on honey bee hemolymph [27], transmits multiple viruses,  
35 and supports replication of some viruses [28, 29, 30, 31]. More than 20 honey bee  
36 viruses have been identified [32]. One of these viruses that has been linked to honey  
37 bee decline is Israeli acute paralysis virus (IAPV), a positive-sense RNA virus of  
38 the family Dicistroviridae [33]. IAPV infection causes shivering wings, decreased  
39 locomotion, muscle spasms, paralysis, and high premature death percentages in  
40 caged infected adult honey bees [34]. IAPV has demonstrated higher infectious  
41 capacities than other honey bee viruses under certain conditions [35] and is more  
42 prevalent in colonies that do not survive the winter [36].

43 Although there is growing interest in how viruses and diet quality affect the health  
44 and sustainability of honey bees, as well as a recognition that such factors might  
45 operate interactively, there are only a small number of experimental studies thus  
46 far directed toward elucidating the interactive effects of these two factors in honey  
47 bees [37, 38, 39, 40, 41]. We recently used laboratory cages and nucleus hive experi-  
48 ments to investigate the health effects of these two factors, and our results show the  
49 importance of the combined effects of both diet quality and virus infection. Specifi-  
50 cally, ingestion by honey bees of high quality pollen is able to mitigate virus-induced  
51 mortality to the level of diverse, polyfloral pollen [11].

52 Following up on these findings, we now aim to understand the corresponding  
53 underlying mechanisms by which high quality diets protect bees from virus-induced  
54 mortality. For example, it is not known whether the protective effect of good diet

is due to direct, specific effects on immune function (resistance), or if it is due to indirect effects of good nutrition on vigor (tolerance) [42]. Transcriptomics is one means to better understand the mechanistic underpinnings of dietary and viral effects on honey bee health. Transcriptomic analysis can help us identify 1) the genomic scale of transcriptomic response to diet and virus infection, 2) whether these factors interact in an additive or synergistic way on transcriptome function, and 3) the types of pathways affected by diet quality and viral infection. This information, heretofore lacking in the literature, can help us better understand how good nutrition may be able to serve as a “buffer” against other stressors [12].

There are only a small number of published experiments examining gene expression patterns related to diet effects [43] and virus infection effects [44] in honey bees, but there have been several such studies in model organisms. Model insect studies can inform studies of honey bee transcriptomic responses, using functional inference of as-of-yet uncharacterized honey bee genes based on orthology to *Drosophila* and other model organisms. Previous *Drosophila* studies that examined various diet effects have found gene expression changes related to immunity, metabolism, cell cycle activity, DNA binding, transcription, and insulin signaling [45, 46, 47, 43]. While similar transcriptomic studies have been limited in honey bees, one study found that pollen nutrition upregulates genes involved in macromolecule metabolism, longevity, and the insulin/TOR pathway required for physiological development [43]. Numerous studies on the transcriptomic effects of virus infection in model insect organisms have shown that RNA silencing, transcriptional pausing, Toll pathways, IMD pathways, JAK/STAT pathways, and Toll-7 autophagy pathways play substantial roles in virus-host systems [48, 49]. Studies of virus-bee systems have revealed some of the antiviral defense pathways known in model organisms are conserved and also related to bee antiviral immune responses [50].

81 To our knowledge, there are few to no studies investigating honey bee gene expres-  
82 sion patterns specifically related to monofloral diets, and few studies investigating  
83 honey bee gene expression patterns related to the combined effects of diet in any  
84 broad sense and viral inoculation in any broad sense [40]. In this study, we examine  
85 how monofloral diets and viral inoculation influence gene expression patterns in  
86 honey bees by focusing on four treatment groups (low quality diet without IAPV  
87 exposure, high quality diet without IAPV exposure, low quality diet with IAPV  
88 exposure, and high quality diet with IAPV exposure). For our diet factor, we exam-  
89 ined two monofloral pollen diets, rockrose (*Cistus* sp.) and chestnut (*Castanea* sp.).  
90 Rockrose pollen is generally considered less nutritious than chestnut pollen because  
91 it contains smaller amounts of protein, amino acids, antioxidants, calcium, and iron  
92 [11, 51]. We conduct RNA-sequencing analysis on a randomly selected subset of the  
93 honey bees we used in our previous study (as is further described in our methods  
94 section). We then examine pairwise combinations of treatment groups, the main  
95 effect of monofloral diet, the main effect of IAPV exposure, and the combined effect  
96 of the two factors on gene expression patterns.

97 Because RNA-seq data can be noisy and subject to high levels of inter-experiment  
98 variation, we further sought to validate our transcriptomic data via comparison to  
99 a previous RNA-seq study on honey bee responses to viral infection. To do this, we  
100 compare the main effect of IAPV exposure in our dataset to that obtained in a pre-  
101 vious study conducted by Galbraith and colleagues [44]. While our study examines  
102 honey bees derived from naturally-mated queens, the Galbraith study examined  
103 honey bees derived from single-drone inseminated queens. As a consequence, the  
104 honey bees in our study will be on average 25% genetically identical, whereas honey  
105 bees from the Galbraith study will be on average 75% genetically identical [52]. We  
106 note that the difference between these studies may be even greater than this as we  
107 used honey bees from 15 different colonies, i.e. from 15 different, naturally-mated

queens. We should therefore expect that the Galbraith study may generate data with higher signal:to:noise ratios than our data due to lower genetic variation between its replicates. At the same time, our honey bees will be more likely to display the health benefits gained from increased genotypic variance within colonies, including decreased parasitic load [53], increased tolerance to environmental changes [54], and increased colony performance [55, 56]. Given that honey bees are naturally very polyandrous [57], our naturally-mated honey bees may also reflect more realistic environmental and genetic conditions. Taken together, each study provides a different point of value: Our study likely presents less artificial data while the Galbraith data likely presents less messy data. To achieve this comparison, we use visualization techniques to assess the signal:to:noise ratio between these two datasets, and differential gene expression (DEG) analyses to determine any significantly overlapping genes of interest between these two datasets. As RNA-sequencing data can be biased [58, 59, 60], this comparison allowed us to characterize how repeatable and robust our RNA-sequencing results were in comparison to previous studies. It also allowed us to shine light on how experimental designs that control genetic variability to different extents might affect the resulting gene expression data in honey bees. We suggest that in-depth data visualization approaches can be useful for cross-study comparisons and validation of noisy RNA-sequencing data in the future.

## Results

### Mortality and virus titers

We reanalyzed our previously published dataset with a subset that focuses on diet quality and is more relevant to the current study. We show the data subset here to inform the RNA-sequencing comparison because we reduced the number of treatments from the original published data (from eight to four) [11] as a means to focus on diet quality effects.

135 As shown in Figure 1, mortality rates of honey bees 72 hours post-inoculation  
136 significantly differed among the treatment groups (mixed model ANOVA across all  
137 treatment groups,  $df = 3, 54$ ;  $F = 10.03$ ;  $p < 2.34e-05$ ). The effect of virus treatment  
138 (mixed model ANOVA,  $df = 1, 54$ ;  $F = 24.73$ ;  $p < 7.04e-06$ ) and diet treatment  
139 (mixed model ANOVA,  $df = 1, 54$ ;  $F = 5.32$ ;  $p < 2.49e-02$ ) were significant, but  
140 the interaction between the two factors (mixed model ANOVA,  $df = 1, 54$ ;  $F =$   
141  $4.72e-02$ ,  $p = 8.29e-01$ ) was not significant. We compared mortality levels based  
142 on pairwise comparisons: For a given diet, honey bees exposed to the virus showed  
143 significantly higher mortality rate than honey bees not exposed to the virus. Bees fed  
144 rockrose pollen had significantly elevated mortality with virus infection compared  
145 to uninfected controls (Benjamini-Hochberg,  $p < 1.53e-03$ ), and bees fed chestnut  
146 pollen similarly had significantly elevated mortality with virus infection compared  
147 to controls (Benjamini-Hochberg,  $p < 3.12e-03$ ) (Figure 1).

148 As shown in Figure 2, IAPV titers of honey bees 72 hours post-inoculation sig-  
149 nificantly differed among the treatment groups (mixed model ANOVA across all  
150 treatment groups,  $df = 3, 33$ ;  $F = 6.10$ ;  $p < 2.03e-03$ ). The effect of virus treatment  
151 (mixed model ANOVA,  $df = 1, 33$ ;  $F = 15.04$ ;  $p < 4.75e-04$ ) was significant, but the  
152 diet treatment (mixed model ANOVA,  $df = 1, 33$ ;  $F = 2.55$ ;  $p = 1.20e-01$ ) and the  
153 interaction between the two factors (mixed model ANOVA,  $df = 1, 33$ ;  $F = 7.02e-$   
154  $01$ ,  $p = 4.08e-01$ ) were not significant. We compared IAPV titers based on pairwise  
155 comparisons: Bees fed rockrose pollen had significantly elevated IAPV titers with  
156 virus infection compared to uninfected controls (Benjamini Hochberg,  $p < 7.56e-$   
157  $03$ ). However, bees fed chestnut pollen did not have significantly elevated IAPV  
158 titers with virus infection compared to uninfected controls (Benjamini Hochberg,  $p$   
159  $= 6.29e-02$ ). Overall, we interpreted these findings to mean that high-quality chest-  
160 nut pollen could partially “rescue” high virus titers resulting from the inoculation  
161 treatment, whereas low-quality rockrose pollen could not (Figure 2).

## 162 Transcriptomic responses to virus infection and diet

163 We observed a substantially larger number of differentially expressed genes (DEGs)  
164 in our diet main effect ( $n = 1,914$ ) than in our virus main effect ( $n = 43$ ) (Sup-  
165 plementary table 1 A and B, Additional file 1). In the diet factor, more DEGs  
166 were upregulated in the more-nutritious chestnut group ( $n = 1,033$ ) than in the  
167 less-nutritious rockrose group ( $n = 881$ ). In the virus factor, there were more virus-  
168 upregulated DEGs ( $n = 38$ ) than control-upregulated DEGs ( $n = 5$ ). While these  
169 reported DEG counts are from the DESeq2 package, we saw similar trends for the  
170 edgeR and limma package results (Supplementary table 1, Additional file 1 and  
171 Additional file 18).

172 GO analysis of the chestnut-upregulated DEGs revealed the following over-  
173 represented biological functions: Wnt signaling, hippo signaling, and dorso-ventral  
174 axis formation, as well as pathways related to circadian rhythm, mRNA surveillance,  
175 insulin resistance, inositol phosphate metabolism, FoxO signaling, ECM-receptor in-  
176 teraction, phototransduction, Notch signaling, JaK-STAT signaling, MAPK signal-  
177 ing, and carbon metabolism (Supplementary table 2, Additional file 1). GO analysis  
178 of the rockrose DEGs revealed pathways related to terpenoid backbone biosynthesis,  
179 homologous recombination, SNARE interactions in vesicular transport, aminoacyl-  
180 tRNA biosynthesis, Fanconi anemia, and pyrimidine metabolism (Supplementary  
181 table 3, Additional file 1).

182 With so few DEGs ( $n = 43$ ) in our virus main effect comparison, we focused on in-  
183 dividual genes and their known functionalities rather than GO over-representation  
184 (Table 1). Of the 43 virus-related DEGs, only 10 had GO assignments within the  
185 DAVID database. These genes had putative roles in the recognition of pathogen-  
186 related lipid products and the cleaving of transcripts from viruses, as well as in-  
187 volvement in ubiquitin and proteosome pathways, transcription pathways, apoptotic  
188 pathways, oxidoreductase processes, and several more functions (Table 1).



189 No interaction DEGs were observed between the diet and virus factors of the  
190 study, in any of the pipelines (DESeq2, edgeR, and limma).

191 The number of DEGs across the six treatment pairings between the diet and  
192 virus factor ranged from 0 to 955 (Supplementary table 8, Additional file 1). Again,  
193 diet level appeared to have greater influence on the number of DEGs than the virus  
194 level. Across every pair comparing the chestnut and rockrose levels, regardless of the  
195 virus level, the number of chestnut-upregulated DEGs was higher than the number  
196 of rockrose-upregulated DEGs (Supplementary table 8 C, D, E, F, Additional file 1).  
197 Virus-treated bees showed equal to or more upregulated genes relative to controls,  
198 under both diet treatments (Supplementary table 8 A and B, Additional file 1).  
199 These trends were observed for all three pipelines used (DESeq2, edgeR, and limma).

#### 200 Transcriptomic data visualization and comparison to a previous study

201 We wished to explore the signal:to:noise ratio between the Galbraith dataset and  
202 our dataset. Note that the Galbraith dataset contained three samples for each virus  
203 level, while our dataset contained twelve samples for each virus level. Basic PCA  
204 plots were constructed with the DESeq2 analysis pipeline and showed that the  
205 Galbraith dataset may separate the infected and uninfected honey bees better than  
206 our dataset (Additional file 2). Wanting to learn more about the data at the gene  
207 level, we continued with new visualization techniques that are available online [61].  
208 To understand how to interpret the visualizations used in this section, please read  
209 (<https://lrutter.github.io/bigPint/articles/plotIntro.html>).

210 We used parallel coordinate lines superimposed onto side-by-side boxplots to visu-  
211 alize the DEGs associated with virus infection in the two studies. The background  
212 side-by-side boxplot represents the distribution of *all* genes in the data, and each  
213 parallel coordinate line represents one DEG. In a parallel coordinate line, connec-  
214 tions between samples with positive correlations should be flat, while connections  
215 between samples with negative correlations should be crossed. We expect DEGs

216 to show more variability between treatments than between replicates. This means  
217 the parallel coordinate lines should be flat between replicates but crossed between  
218 treatments. However, overplotting problems would obscure our visualization if we  
219 were to plot all DEGs onto the same side-by-side boxplot. Therefore, we graphed  
220 clustered subsets of the DEGs (based on hierarchical clustering).

221 The 1,019 DEGs from the Galbraith dataset form relatively clean-looking visual  
222 displays, with consistent replicates and differences between treatments (Figure 3). In  
223 contrast, we see that the 43 virus-related DEGs from our dataset do not look as clean  
224 in their visual displays (Figure 4). The replicates appear somewhat inconsistent in  
225 their estimated expression levels and there is not always such a large (or even  
226 consistent) difference between treatment groups. We see a similar finding when we  
227 also examine a larger subset of 1,914 diet-related DEGs from our study (Additional  
228 file 3).

229 We next used repLIcate TREatment (“litre”) plots, which we recently developed  
230 for our bigPint software package. Litre plots allow users to visualize one DEG onto  
231 the Cartesian coordinates of one scatterplot matrix. In the litre plot, each gene  
232 in the data is plotted once for every combination of replicates between treatment  
233 groups. We use hexagon bins to summarize this massive information. Once the  
234 background of hexagons has been drawn to reveal the distribution of all between-  
235 treatment sample pair combinations for *all* genes, the user can superimpose all  
236 between-treatment sample pair combinations for one gene of interest.

237 Additional file 4 shows nine example litre plots for our dataset; each litre plot  
238 shows the 144 between-treatment sample pair combinations for one DEG of inter-  
239 est. Additional file 5 and 6 similarly each show nine example litre plots for the  
240 Galbraith dataset; each litre plot shows the nine between-treatment sample pair  
241 combinations for one DEG of interest. We see that indeed the virus DEGs from  
242 our data (Additional file 4) show less consistent replications and less differences be-

243 tween the treatment groups compared to the virus DEGs from the Galbraith data  
244 (Additional files 5 and 6). We also observe that, in the Galbraith dataset, the DEG  
245 points in the first cluster show less tight cluster patterns than the DEG points in  
246 the second cluster (Additional files 5 and 6), an observation we saw previously in  
247 the parallel coordinate plots (Figure 3).

248 Finally, we used scatterplot matrices from the bigPint software to further assess  
249 the DEGs. A scatterplot matrix is another effective multivariate visualization tool  
250 that plots read count distributions across all genes and samples. Specifically, it rep-  
251 resents every gene in the dataset as a black point in each scatterplot. DEGs can  
252 be superimposed as colored points to assess their patterns against the full dataset.  
253 We expect DEGs to mostly fall along the  $x=y$  line in replicate scatterplots (denot-  
254 ing replicate consistency) but deviate from the  $x=y$  line in treatment scatterplots  
255 (denoting significant treatment changes). The  $x=y$  line is shown in red in our plots.

256 We created standardized scatterplot matrices for each of the four clusters (from  
257 Figure 3) of the Galbraith data (Additional files 7, 8, 9, and 10). We also created  
258 standardized scatterplot matrices for our data. However, as our dataset contained  
259 24 samples, we would need to include 276 scatterplots in our matrix, which would  
260 be too numerous to allow for efficient visual assessment of the data. As a result,  
261 we created four scatterplot matrices of our data, each with subsets of 6 samples  
262 to be more comparable to the Galbraith data (Additional files 11, 12, 13, and 14).  
263 We can again confirm through these plots that the DEGs from the Galbraith data  
264 appeared more as expected: They deviated more from the  $x=y$  line in the treatment  
265 scatterplots while staying close to the  $x=y$  line in replicate scatterplots.

266 Despite the virus-related DEGs ( $n = 1,019$ ) from the Galbraith dataset displaying  
267 the expected patterns more than those from our dataset ( $n = 43$ ), there was signif-  
268 icant overlap ( $p\text{-value} < 2.2\text{e-}16$ ) in the DEGs between the two studies, with 26/38

(68%) of virus-upregulated DEGs from our study also showing virus-upregulated response in the Galbraith study (Figure 6).

#### Tolerance versus resistance

Using the contrasts specified in Table 2, we discovered 122 “tolerance” candidate DEGs and 125 “resistance” candidate DEGs. Within our 122 “tolerance” gene ontologies, we found functions related to metabolism (such as carbohydrate metabolism, fructose metabolism, and chitin metabolism). However, we also discovered gene ontologies related to RNA polymerase II transcription, immune response, and regulation of response to reactive oxygen species (Figure 5A). Within our 125 “resistance” gene ontologies, we found functions related to metabolism (such as carbohydrate metabolism, chitin metabolism, oligosaccharide biosynthesis, and general metabolism) (Figure 5B).

To visually explore gene expression patterns related to tolerance and resistance, we used hierarchical clustering to separate candidate DEGs into common patterns, and then visualized these clusters using parallel coordinate lines superimposed onto side-by-side boxplots. To reduce overplotting of parallel coordinate lines, we again used hierarchical clustering techniques to separate DEGs into common patterns. Perhaps unsurprisingly, we still see a substantial amount of noise (inconsistency between replicates) in our resulting candidate DEGs (Additional files 15 and 16). However, the broad patterns we expect to see still emerge: For example, based on the contrasts we created to obtain the ‘tolerance’ candidate DEGs, we expect them to display larger count values in the “NC” group compared to the “NR” group and larger count values in the “VC” group compared to the “VR” group. Indeed, we see this pattern in the associated parallel coordinate plots (Additional file 15). Likewise, based on the contrasts we created to obtain the ‘resistance’ candidate DEGs, we still expect them to display larger count values in the “VC” group compared to the “VR” group, but we no longer expect to see a difference between the “NC”

and “NR” groups. We do generally see these expected patterns in the associated parallel coordinate plots: While there are large outliers in the “NC” group, the “NR” replicates are no longer typically below a standardized count of zero (Additional file 16). The genes in Cluster 3 may follow the expected pattern the most distinctively (Additional file 16).

### Post hoc analysis

To better understand sources of transcriptomic noise, we explored whether pathogen response measurements (virus titers and mortality), which varied widely across samples, were correlated with observed patterns in gene expression.

The R-squared values between gene read counts and pathogen response measurements were generally low ( $R\text{-squared} < 0.1$ ) across our dataset (Supplementary table 9, Additional file 1). We further explored whether clusters of DEGs showed higher correlations with pathogen response measurements than non-DEGs (the latter serving as a control, where we do not expect a correlation). A Kruskal–Wallis test was used to determine if R-squared distributions of DEG clusters significantly differed from those in the rest of the data. The p-values and Bonferroni correction values for each of the 36 tests (as described in the methods section) is provided in Supplementary table 9, Additional file 1. An overall trend emerges to suggest that DEGs may have significantly larger correlation with the pathogen response measurements compared to non-DEGs.

## Discussion

Challenges to honey bee health are a growing concern, in particular the combined, interactive effects of nutritional stress and pathogens [12]. In this study, we used RNA-sequencing to probe mechanisms underlying honey bee responses to two effects, diet quality and infection with the prominent virus of concern, IAPV. In general, we found a major nutritional transcriptomic response, with nearly 2,000 transcripts changing in response to diet quality (rockrose/poor diet versus chest-

nut/good diet). The majority of these genes were upregulated in response to high quality diet, and these genes were over-represented for functions such as nutrient signaling metabolism (insulin resistance), immune response (Notch signaling and JaK-STAT pathways), and carbon metabolism (Supplementary table 2, Additional file 1). These data suggest high quality nutrition may allow bees to alter their metabolism, favoring investment of energy into innate immune responses.

One of the few studies that has investigated transcriptomic response to nutrition in honey bees similarly found that pollen upregulates genes related to macromolecule metabolism, insulin pathways, and TOR pathways [43]. Diet effects on transcriptomics have been more extensively studied in the insect model *Drosophila*. One recent transcriptomic study in *Drosophila melanogaster* reported an overexpression of genes related to immunity, metabolism, and hemocyanin in a high-fat diet and overexpression of genes related to cell cycle activity, DNA binding and transcription, and CHK kinase-like protein activity in a high-sugar diet [45]. This same study also discovered an upregulation of genes related to peptide and carbohydrate processing in both high-fat and high-sugar diets, a finding the authors attributed to a general increase in caloric intake. Another recent study investigated the transcriptomic effects of diets high in protein relative to sugar, diets high in sugar relative to protein, and diets with equal amounts of protein and sugar [46]. *Drosophila mojavensis* and *Drosophila arizonae* showed substantial differential expression between the dietary conditions: genes involved in carbohydrate and lipid metabolism were upregulated in response to high sugar low protein diets and genes involved in juvenile hormone (JH) and ecdysone were upregulated in response to low sugar high protein diets. Interestingly, prior studies have suggested that JH regulates body size by controlling ecdysone production, which modifies insulin signaling [47]. As we saw in our study, these studies generally suggest that diet differences may relate to gene expression changes in metabolism and immune responses in honey bees.

350 While some insect systems have shown relatively low transcriptional responses  
351 to dicistrovirus infection [62, 63], previous work on honey bees has revealed many  
352 hundreds of DEGs [44]. Discrepancies between datasets may be due to noise and  
353 complexity of the honey bee microbiome. The transcriptomic response to virus infec-  
354 tion in our experiment was fairly limited. We found only 43 differentially expressed  
355 transcripts, some with known immune functions such as a gene with similarity to  
356 MD-2 lipid recognition protein and argonaute-2, a protein that plays a central role  
357 in RNA silencing (Table 1). We also found genes related to transcriptional regu-  
358 lation and muscle contraction. The small number of DEGs in this study may be  
359 partly explained by the large amount of noise in the data (Figure 4 and Additional  
360 files 2B, 4, 11, 12, 13, and 14).

361 There have been numerous studies on the transcriptomic effects of virus infection  
362 in model organisms like fruit flies and mosquitoes that can provide a useful frame-  
363 work for interpreting virus responses in honey bees. These studies have showed that  
364 RNA silencing is a major antiviral strategy, along with transcriptional pausing, Toll  
365 pathways, IMD pathways, JAK/STAT pathways, and Toll-7-autophagy pathways  
366 [48, 49]. Recent transcriptomic studies in honey bees have shown similar hallmarks  
367 of these same antiviral defense mechanisms, including RNA silencing, Toll path-  
368 ways, IMD pathways, JAK/STAT pathways, autophagy, and endocytosis [50]. It is  
369 important to note that general immune responses to viral infection in insects might  
370 be an indirect result of cellular damage [49]. In fact, every virus-host interaction has  
371 its own particularities derived from the diverse methods of replication and infection  
372 cycle evolved by different viruses. An intricate set of pro- and anti-virus host factors  
373 such as ribosomal proteins and autophagy pathways are involved, but the response  
374 depends on the virus species, as has been elucidated in *Drosophila* [48, 49]. In ad-  
375 dition, a non-sequence-specific antiviral response mediated by unspecific dsRNA  
376 pathway was discovered in honey bees [64, 65]. In the case of dicistroviruses, few

works have studied the impact of IAPV infection at transcriptional level. Chen et al. 2014 analyzed responses to IAPV infection in larvae and workers using microarrays [66]. Many of the DEGs found were involved in immune response and energy-related metabolism, particularly in adults but not in brood. The authors propose this observed difference could be connected to latent infections in larvae (where host immunity is not perturbed) versus acute infections in adulthood (induced by stressors faced during development) [66]. IAPV acute infection also alters the DNA methylation pattern of numerous genes that do not overlap the genes that are up- or down-regulated at the transcriptional level [44]. These works reiterate the conclusion that viruses trigger particular antiviral mechanisms by different means and depending on several factors. The honey bee antiviral pathways induced by specific viruses were recently reviewed [50]; it is noteworthy that many honey bee factors discovered by transcriptomics need further characterization to uncover their role in controlling (or promoting) viral infection in honey bees.

Given the noisy nature of our data, and our desire to home in on genes with real expression differences, we compared our data to the Galbraith study [44], which also examined bees response to IAPV infection. In contrast to our study, Galbraith et al. identified a large number of virus responsive transcripts, and generally had less noise in their data (Figure 3 and Additional files 2A, 5, 6, 7, 8, 9, and 10). To identify the most consistent virus-responsive genes from our study, we looked for overlap in the DEGs associated with virus infection on both experiments. We found a large, statistically significant ( $p\text{-value} < 2.2\text{e-}16$ ) overlap, with 26/38 (68%) of virus-responsive DEGs from our study also showing response to virus infection in Galbraith et al. (Figure 6). This result gives us confidence that, although noisy, we were able to uncover reliable, replicable gene expression responses to virus infection with our data.



403 Data visualization is a useful method to identify noise and robustness in RNA-  
404 sequencing data [67]. In this study, we used extensive data visualization to improve  
405 the interpretation of our RNA-sequencing results. For example, the DESeq2 pack-  
406 age comes with certain visualization options that are popular in RNA-sequencing  
407 analysis. One of these visualization is the principal component analysis (PCA) plot,  
408 which allows users to visualize the similarity between samples within a dataset. We  
409 could determine from this plot that indeed the Galbraith data may show more simi-  
410 larity between its replicates and differences between its treatments compared to our  
411 data (Additional file 2). However, the PCA plot only shows us information at the  
412 sample level. We wanted to investigate how these differences in the signal:to:noise  
413 ratios of the datasets would affect the structure of any resulting DEGs. As a result,  
414 we also used three plotting techniques from the bigPint package: We investigated  
415 the 1,019 virus-related DEGs from the Galbraith dataset and the 43 virus-related  
416 DEGs from our dataset using parallel coordinate lines, scatterplot matrices, and  
417 litre plots. To prevent overplotting issues in our graphics, we used a hierarchical  
418 clustering technique for the parallel coordinate lines to separate the set of DEGs  
419 into smaller groups. We also needed to examine four subsets of samples from our  
420 dataset to make effective use of the scatterplot matrices. After these tailorizations,  
421 we determined that the same patterns we saw in the PCA plots regarding the entire  
422 dataset extended down the pipeline analysis into the DEG calls: Even the DEGs  
423 from the Galbraith dataset showed more similarity between their replicates and dif-  
424 ferences between their treatments compared to those from our data. However, the  
425 365 DEGs from the Galbraith data in Cluster 1 of Figure 3 showed an inconsistent  
426 first replicate in the treatment group (“V.1”), which was something we observed  
427 in the PCA plot. This indicates that this feature also extended down the analysis  
428 pipeline into DEG calls. Despite the differences in signal between these two datasets,  
429 there was substantial overlap in the resulting DEGs. We believe these visualization

430 applications can be useful for future researchers analyzing RNA-sequencing data to  
431 quickly and effectively ensure that the DEG calls look reliable or at least overlap  
432 with DEG calls from similar studies that look reliable. We also expect this type of  
433 visualization exploration can be especially crucial when studying wild populations  
434 with high levels of genetic and environmental variation between replicates and/or  
435 when using experiments that may lack rigid design control.

436 One of the goals of this study was to use our RNA-sequencing data to assess  
437 whether transcriptomic responses to diet quality and virus infection provide insight  
438 into whether high quality diet can buffer bees from pathogen stress via mechanisms  
439 of “resistance” or “tolerance”. Recent evidence has suggested that overall immu-  
440 nity is determined by more than just “resistance” (the reduction of pathogen fitness  
441 within the host by mechanisms of avoidance and control) [68]. Instead, overall im-  
442 munity is related to “resistance” in conjunction with “tolerance” (the reduction  
443 of adverse effects and disease resulting from pathogens by mechanisms of heal-  
444 ing) [42, 68]. Immune-mediated resistance and diet-driven tolerance mechanisms  
445 are costly and may compete with each other [42, 69]. Data and models have sug-  
446 gested that selection can favor an optimum combination of both resistance and  
447 tolerance [70, 71, 72, 73]. We attempted to address this topic through specific gene  
448 expression contrasts (Table 2), accompanied by GO analysis of the associated gene  
449 lists. We found an approximately equal number of resistance ( $n = 125$ ) and toler-  
450 ance ( $n = 122$ ) related candidate DEGs, suggesting both processes may be playing  
451 significant roles in dietary buffering from pathogen induced mortality. Resistance  
452 candidate DEGs had functions related to several forms of metabolism (chitin and  
453 carbohydrate), regulation of transcription, and cell adhesion (Figure 5B). Toler-  
454 ance candidate DEGs had functions related to carbohydrate metabolism and chitin  
455 metabolism; however, they also showed functions related to immune response, in-  
456 cluding RNA polymerase II transcription (Figure 5A). Previous studies have shown

457 that transcriptional pausing of RNA polymerase II may be an innate immune re-  
458 sponse in *D. melanogaster* that allows for a more rapid response by increasing  
459 the accessibility of promoter regions of virally induced genes [74]. These possible  
460 immunological defense mechanisms within our “tolerance” candidate DEGs and  
461 metabolic processes within our “resistance” candidate DEGs may provide addi-  
462 tional evidence of feedbacks between diet and disease in honey bees [12].

463 There were several limitations in this study that could be improved upon in fu-  
464 ture studies. For instance, our comparison between the Galbraith data (single-drone  
465 colonies) and our data (naturally-mated colonies) was limited by numerous extra-  
466 neous variables between these studies. In addition to different molecular pipelines  
467 and bioinformatic preprocessing pipelines used between these studies, the Galbraith  
468 study focused on worker honey bees that were fed sugar and artificial pollen di-  
469 ets, whereas our study focused on worker honey bees that were fed bee-collected  
470 monofloral diets. Furthermore, the Galbraith data used eviscerated abdomens with  
471 attached fat bodies and only considered symptomatic honey bees for their infected  
472 treatment group, whereas we used whole bodies and considered both asymptomatic  
473 and symptomatic honey bees for our infected treatment group. There are also dif-  
474 ferences in the hours post inoculation and possible differences in the inoculation  
475 amount between the studies. Further differences between the studies can be found  
476 in their corresponding published methods sections [11, 44]. The different factors  
477 between these two studies may be critical because particular antiviral factors in  
478 honey bees are linked to specific viruses, specific developmental stages, the ana-  
479 lyzed tissue, the route of inoculation, and the time (post-inoculation) during which  
480 the study was performed. This was clearly demonstrated when comparing honey bee  
481 responses to two related iflaviruses with very different infection dynamics, sacbrood  
482 bee virus (SBV) vs. deformed wing virus (DWV) [75]. Authors observed differences

483 in induction of defensin and hymenoptaecin immune-related genes, and suggested  
484 the results reflect adaptations to the different routes of transmission [75].

485 Moreover, our comparative visualization assessment between these two datasets  
486 was also somewhat limited because the virus effect in the Galbraith study used  
487 three replicates for each level, whereas the virus effect in our study used twelve  
488 replicates for each level that were actually further subdivided into six replicates for  
489 each diet level. Hence the apparent reduction in noise observed in the Galbraith  
490 data compared to our data in the PCA plots, parallel coordinate plots, scatterplot  
491 matrices, and litre plots may be an inadvertent product of the smaller number of  
492 replicates used and the lack of a secondary treatment group rather than solely the  
493 reduction in genetic variability through the single-drone colony design itself. With  
494 this in mind, while our current efforts may be a starting point, future studies can  
495 shed more light on signal:to:noise and differential expression differences between  
496 naturally-mated colony designs and single-drone colony designs by controlling for  
497 extraneous factors more strictly than what we were able to do in the current line  
498 of work.

499 In addition, this study used a whole body RNA-sequencing approach. In future  
500 related studies, it may be informative to use tissue-specific methods. Previous work  
501 has shown that even though IAPV replication occurs in all honey bee tissues, it  
502 localizes more in gut and nerve tissues and in the hypopharyngeal glands. Likewise,  
503 the highest IAPV titers have been observed in gut tissues [36]. Recent evidence has  
504 suggested that RNA-sequencing approaches toward composite structures in honey  
505 bees leads to false negatives, implying that genes strongly differentially expressed  
506 in particular structures may not reach significance within the composite structure  
507 [76]. These studies have also found that within a composite extraction, structures  
508 therein may contain opposite patterns of differential expression. We can provide  
509 more detailed answers to our original transcriptomic questions if we were to repeat

510 this same experimental design only now at a more refined tissue level. Another  
511 future direction related to this work would be to integrate multiple omics datasets  
512 to investigate monofloral diet quality and IAPV infection in honey bees. Indeed,  
513 previous studies in honey bees have found that multiple omics datasets do not  
514 always align in a clear-cut manner, and hence may broaden our understanding of  
515 the molecular mechanisms being explored [44].

## 516 **Conclusions**

517 To the best of our knowledge, there are few to no studies investigating honey bee  
518 gene expression specifically related to monofloral diets, and few to no studies ex-  
519 amining honey bee gene expression related to the combined effects of diet in any  
520 general sense and viral inoculation in any general sense. It also remains unknown  
521 whether the protective effects of good diet in honey bees is due to direct effects on  
522 immune function (resistance) or indirect effects of energy availability on vigor and  
523 health (tolerance). We attempted to address these unresolved areas by conducting  
524 a two-factor RNA-sequencing study that examined how monofloral diets and IAPV  
525 inoculation influence gene expression patterns in honey bees. Overall, our data sug-  
526 gest complex transcriptomic responses to multiple stressors in honey bees. Diet has  
527 the capacity for large and profound effects on gene expression and may set up the  
528 potential for both resistance and tolerance to viral infection, adding to previous  
529 evidence of possible feedbacks between diet and disease in honey bees [12].

530 Moreover, this study also demonstrated the benefits of using data visualizations  
531 and multiple datasets to address inherently messy biological data. For instance, by  
532 verifying the substantial overlap in our DEG lists to those obtained in another study  
533 that addressed a similar question using specimens with less genetic variability, we  
534 were able to place much higher confidence in the differential gene expression results  
535 from our otherwise noisy data. We also suggested that comparing results derived  
536 from multiple studies varying in level of genetic and environmental variability may

allow researchers to identify transcriptomic patterns that are concurrently more realistic and less noisy. Altogether, we hope our results underline the merits of using data visualization techniques and multiple datasets to understand and interpret RNA-sequencing datasets.

## Methods

### Mortality and virus titers

Details of the procedures we used to prepare virus inoculum, infect and feed caged honey bees, and quantify IAPV can be reviewed in our previous work [11, 35]. A linear mixed effects model was used to relate the mortality rates and IAPV titers to the main and interaction effects of the diet and virus factors. The model was fitted to the data by restricted maximum likelihood (REML) using the “lme” function in the R package “nlme”. A random (intercept) effect for experimental setup was included in the model. Post-hoc pairwise comparisons of the four (diet and virus combination) treatment groups were performed and Benjamini-Hochberg adjusted p-values were calculated to limit familywise Type I error rates [77].

### Design of two-factor experiment

For our nutrition factor, we examined two monofloral pollen diets, rockrose (*Cistus* sp.) and chestnut (*Castanea* sp.). Rockrose pollen is generally considered less nutritious than chestnut pollen due to its lower levels of protein, amino acids, antioxidants, calcium, and iron [11, 51]. For our virus factor, one level contained bees that were infected with IAPV and another level contained bees that were not infected with IAPV. This experimental design resulted in four treatment groups (rockrose pollen without IAPV exposure, chestnut pollen without IAPV exposure, rockrose pollen with IAPV exposure, and chestnut pollen with IAPV exposure) that allowed us to assess main effects and interactive effects between diet quality and IAPV infection in honey bees.

There are several reasons why our design focused only on diet quality (monofloral diets) as opposed to diet diversity (monofloral diets versus polyfloral diets). First, when assessing diet diversity, a sugar diet is often used as a control. However, such an experimental design does not reflect real-world conditions for honey bees as they rarely face a total lack of pollen [51]. Second, in studies that compared honey bee health using monofloral and polyfloral diets at the same time, if the polyfloral diet and one of the high-quality monofloral diets both exhibited similarly beneficial effects, then it was difficult for the authors to assess if the polyfloral diet was better than most of the monofloral diets because of its diversity or because it contained as a subset the high-quality monofloral diet [51]. Third, as was previously mentioned, honey bees are now confronted with less diverse sources of pollen. As a result, there is a need to better understand how monofloral diets affect honey bee health.

#### RNA extraction

Fifteen cages per treatment were originally produced for monitoring of mortality. From these, six live honey bees were randomly selected from each cage 36 hours post inoculation and placed into tubes [35]. Tubes were kept on dry ice and then transferred into a -80C freezer until processing. From the fifteen possible cages, eight were randomly selected for RNA-sequencing. From these eight cages, two of the honey bees per cage were randomly selected from the original six live honey bees per cage. These two bees were combined to form a pooled sample representing the cage. Whole body RNA from each pool was extracted using Qiagen RNeasy MiniKit followed by Qiagen DNase treatment. Samples were suspended in water to 200-400 ng/ $\mu$ l. All samples were then tested on a Bioanalyzer at the Iowa State University DNA Facility to ensure quality (RIN > 8).

#### Gene expression

Samples were sequenced starting on January 14, 2016 at the Iowa State University DNA Facility (Platform: Illumina HiSeq Sequencing 2500 in rapid run mode; Cat-

egory: Single End 100 cycle sequencing). A standard Illumina mRNA library was prepared by the DNA facility. Reads were aligned to the BeeBase Version 3.2 genome [78] from the Hymenoptera Genome Database [79] using the programs GMAP and GSNAP [80]. There were four lanes of sequencing with 24 samples per lane. Each sample was run twice. Approximately 75-90% of reads were mapped to the honey bee genome. Each lane produced around 13 million single-end 100 basepair reads.

We tested all six pairwise combinations of treatments for DEGs (pairwise DEGs). We also tested the diet main effect (diet DEGs), virus main effect (virus DEGs), and interaction term for DEGs (interaction DEGs). We then also tested for virus main effect DEGs (virus DEGs) in public data derived from a previous study exploring the gene expression of IAPV virus infection in honey bees [44]. We tested each DEG analysis using recommended parameters with DESeq2 [81], edgeR [67], and LimmaVoom [82]. In all cases, we used a false discovery rate (FDR) threshold of 0.05 [83]. Fisher's exact test was used to determine significant overlaps between DEG sets (whether from the same dataset but across different analysis pipelines or from different datasets across the same analysis pipelines). The eulerr shiny application was used to construct Venn diagram overlap images [84]. In the end, we focused on the DEG results from DESeq2 [81] as this pipeline was also used in the Galbraith study [44]. We used the independent filtering process built into the DESeq2 software that mitigates multiple comparison corrections on genes with no power rather than defining one filtering threshold.

#### Comparison to prior studies on transcriptomic response to viral infection

We compare the main effect of IAPV exposure in our dataset to that obtained in a previous study conducted by Galbraith and colleagues [44] who also addressed honey bee transcriptomic responses to virus infection. We applied the same downstream bioinformatics analyses between our count table and the count table provided in the Galbraith study. When we applied our bioinformatics pipeline to the Galbraith



count table, we obtained different differential expression counts compared to the results published in the Galbraith study. However, there was substantial overlap and we considered this justification to use the differential expression list we obtained in order to keep the downstream bioinformatics analyses as similar as possible between the two datasets (Additional file 17).

We used honey bees from naturally-mated colonies, whereas Galbraith et al. [44] used honey bees from single-drone colonies. In light of this, we should expect the Galbraith et al. dataset to contain lower genetic variation between its replicates and higher signal:to:noise ratios than our dataset. We use visualization techniques to assess the signal:to:noise ratio between these two datasets, and differential gene expression (DEG) analyses to determine any significantly overlapping genes of interest between these two datasets.

## Visualization

We used an array of visualization tools as part of our analysis. We used the PCA plot [85] from the DESeq2 package, a well-known and established tool. Along with that, we used lesser-known multivariate visualization tools from our work-in-progress R package called bigPint. Specifically, we used parallel coordinate plots [86], scatter-plot matrices [87], and litre plots (which we recently developed based on “replicate line plots” [88]) to assess the variability between the replicates and the treatments in our data. We also used these plotting techniques to assess for normalization problems and other common problems in RNA-sequencing analysis pipelines [88].

Furthermore, we used statistical graphics to better understand patterns in our DEGs. However, in cases of large DEG lists, these visualization tools had overplotting problems (where multiple objects are drawn on top of one another, making it impossible to detect individual values). To remedy this problem, we first standardized each DEG to have a mean of zero and standard deviation of unity [89, 90]. Then, we performed hierarchical clustering on the standardized DEGs using Ward’s

linkage. This process divided large DEG lists into smaller clusters of similar patterns, which allowed us to more efficiently visualize the different types of patterns within large DEG lists (see Figures 3 and 4 for examples).

#### Gene ontology

DEGs were uploaded as a background list to DAVID Bioinformatics Resources 6.7 [91, 92]. The overrepresented gene ontology (GO) terms of DEGs were determined using the BEEBASE\_ID identifier option (honey bee gene model) in the DAVID software. To fine-tune the GO term list, only terms correlating to Biological Processes were considered. The refined GO term list was then imported into REVIGO [93], which uses semantic similarity measures to cluster long lists of GO terms.

#### Probing tolerance versus resistance

To investigate whether the protective effect of good diet is due to direct, specific effects on immune function (resistance), or if it is due to indirect effects of good nutrition on energy availability and vigor (tolerance), we created contrasts of interest (Table 2). In particular, we assigned “resistance candidate DEGs” to be the ones that were upregulated in the chestnut group within the virus infected bees but not upregulated in the chestnut group within the non-infected bees. Our interpretation of these genes is that they represent those that are only activated in infected bees that are fed a high quality diet. We also assigned “tolerance candidate DEGs” to be the ones that were upregulated in the chestnut group for both the virus infected bees and non-infected bees. Our interpretation of these genes is that they represent those that are constitutively activated in bees fed a high quality diet, regardless of whether they are experiencing infection or not. We then determined how many genes fell into these two categories and analyzed their GO terminologies.

## 668 Post hoc analysis

669 We found considerable noisiness in our data and saw, through gene-level visual-  
670 izations, that our DEGs contained outliers and inconsistent replicates. Hence, we  
671 wanted to explore whether our DEG read counts correlated with pathogen response  
672 metrics, including IAPV titers, sacbrood bee virus (SBV) titers, and mortality rates.  
673 For this process, we considered virus main effect DEGs (Figure 4), “tolerance can-  
674 didate” DEGs (Additional file 15), and “resistance candidate” DEGs (Additional  
675 file 16). For each DEG in each cluster, we calculated a coefficient of determination  
676 (R-squared) value to estimate the correlation between its raw read counts and the  
677 pathogen response metrics across its 24 samples. We then used the Kruskal–Wallis  
678 test to determine if the distribution of the R-squared values in any of the DEG clus-  
679 ters significantly differed from those in the non-DEG genes (the rest of the data).  
680 As there were four clusters for each of the nine combinations of DEG lists (“tol-  
681 erance” candidate DEGs, “resistance” candidate DEGs, and virus-related DEGs)  
682 and pathogen response measurements (IAPV titer, SBV titer, and mortality rate),  
683 this process resulted in 36 statistical tests.

## 684 Declarations

### 685 Ethics approval and consent to participate

686 All honey bees used in this work were sampled in the United States, and no ethical use approval is required for this  
687 species in this country.

### 688 Consent for publication

689 Not applicable.

### 690 Availability of data and materials

691 The data discussed in this publication have been deposited in NCBI’s Gene Expression Omnibus [94] and are  
692 accessible through GEO Series accession number GSE121885  
693 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121885>). Scripts to reproduce visualizations in this  
694 paper are available online (<https://github.com/lrutter/HoneyBeePaper>). Information about bigPint visualizations,  
695 including how to interpret them are available online (<https://lrutter.github.io/bigPint>). Lists of BeeBase IDs for  
696 contrast DEGs from this study are found in Additional File 20.

### 697 Competing interests

698 The authors declare that they have no competing interests.

### 699 Funding

700 This work was supported by the United States Department of Agriculture, Agriculture and Food Research Initiative  
701 (USDA-AFRI) 2011-04894.

### 702 Author’s contributions

703 LR performed the bioinformatic and statistical analyses and produced the figures and tables. LR and ALT drafted  
704 the manuscript. ALT, AGD, JCT, BCB contributed to experimental design. AGD and JCT carried out laboratory  
705 experiments. AGD processed samples for virus and RNA-seq. DC advised on statistical analyses and visualization.  
706 All authors revised and approved the manuscript.

## 707 Acknowledgements

708 We would like to thank Giselle Narvaez for assisting with cage experiments.

## 709 Author details

710 <sup>1</sup>Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA 50011, USA. <sup>2</sup>Department of  
711 Microbiology, Center for Scientific Research and Higher Education of Ensenada, Ensenada, Baja California 22860,  
712 Mexico. <sup>3</sup>Department of Entomology and Nematology, University of Florida, Gainesville, FL 32611, USA.  
713 <sup>4</sup>Econometrics and Business Statistics, Monash University, Clayton, VIC 3800, Australia. <sup>5</sup>Department of  
714 Entomology, Iowa State University, Ames, IA 50011, USA. <sup>6</sup>Department of Ecology, Evolution, and Organismal  
715 Biology, Iowa State University, Ames, IA 50011, USA. <sup>7</sup>Department of Entomology, University of Illinois at  
716 Urbana-Champaign, Urbana, IL 61801, USA.

## 717 References

- 718 1. van Engelsdorp, D., Evans, J.D., Saegerman, C., Mullin, C., Haubruge, E., Nguyen, B.K., Frazier, M., Frazier,  
719 J., Cox-Foster, D., Chen, Y., Underwood, R., Tarpay, D.R., Pettis, J.S.: Colony collapse disorder: A descriptive  
720 study. *PLoS ONE* **4**, 6481 (2009)
- 721 2. Kulhanek, K., Steinhauer, N., Rennich, K., Caron, D.M., Sagili, R.R., Pettis, J.S., Ellis, J.D., Wilson, M.E.,  
722 Wilkes, J.T., Tarpay, D.R., Rose, R., Lee, K., Rangel, J., vanEngelsdorp, D.: A national survey of managed  
723 honey bee 2014–2015 annual colony losses in the USA. *Journal of Apicultural Research* **56**, 328–340 (2017)
- 724 3. Laurent, M., Hendrikx, P., Ribiere-Chabert, M., Chauzat, M.-P.: A pan-European epidemiological study on  
725 honeybee colony losses 2012–2014. *Epilobee* **2013**, 44 (2016)
- 726 4. Caron, D., Sagili, R.: Honey bee colony mortality in the Pacific Northwest: Winter 2009/2010. *Am Bee J* **151**,  
727 73–76 (2011)
- 728 5. Bond, J., Plattner, K., Hunt, K.: Fruit and Tree Nuts Outlook: Economic Insight U.S. Pollination- Services  
729 Market. Economic Research Service Situation and Outlook FTS-357SA, USDA (2014)
- 730 6. Gallai, N., Salles, J.-M., Settele, J., Vaissière, B.B.: Economic valuation of the vulnerability of world agriculture  
731 confronted with pollinator decline. *Ecol. Econ.* **68**, 810–821 (2009)
- 732 7. Klein, A.-M., Vaissière, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T.:  
733 Importance of pollinators in changing landscapes for world crops. *Proc Biol Sci* **274**, 303–313 (2007)
- 734 8. Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., Kunin, W.E.: . Global pollinator  
735 declines: trends, impacts and drivers **25**, 345–353 (2010)
- 736 9. Spivak, M., Mader, E., Vaughan, M., Euliss, N.H.: The Plight of the Bees. *Environ Sci Technol* **45**, 34–38  
737 (2011)
- 738 10. Goulson, D., Nicholls, E., Botías, C., Rotheray, E.L.: Bee declines driven by combined stress from parasites,  
739 pesticides, and lack of flowers. *Science* **347**, 1255957 (2015)
- 740 11. Dolezal, A.G., Carrillo-Tripp, J., Judd, T., Miller, A., Bonning, B., Toth, A.: Interacting stressors matter: Diet  
741 quality and virus infection in honey bee health. In prep (2018)
- 742 12. Dolezal, A.G., Toth, A.L.: Feedbacks between nutrition and disease in honey bee health. *Current Opinion in*  
743 *Insect Science* **26**, 114–119 (2018)
- 744 13. Roulston, T.H., Buchmann, S.L.: A phylogenetic reconsideration of the pollen starch-pollination correlation.  
745 *Evol Ecol Res* **2**, 627–643 (2000)
- 746 14. Stanley, R.G., Linsens, H.F.: Pollen: Biology, Biochemistry, Management
- 747 15. Brodschneider, R., Crailsheim, K.: Nutrition and health in honey bees. *Apidologie* **41**, 278–294 (2010)
- 748 16. Haydak, M.H.: Honey bee nutrition. *Annu Rev Entomol* **15**, 143–156 (1970)
- 749 17. Crailsheim, K., Schneider, L.H.W., Hrassnigg, N., Bühlmann, G., Brosch, U., Gmeinbauer, R., Schöffmann, B.:  
750 Pollen consumption and utilization in worker honeybees (*Apis mellifera carnica*): dependence on individual age  
751 and function. *J Insect Physiol* **38**, 409–419 (1992)
- 752 18. Crailsheim, K.: The flow of jelly within a honeybee colony. *J Comp Physiol B* **162**, 681–689 (1992)
- 753 19. Schmidt, J.O.: Feeding preference of *Apis mellifera* L. (Hymenoptera: Apidae): Individual versus mixed pollen  
754 species. *J. Kans. Entomol. Soc.* **57**, 323–327 (1984)
- 755 20. Schmidt, J.O., Thoenes, S.C., Levin, M.D.: Survival of honey bees, *Apis mellifera* (Hymenoptera: Apidae), fed  
756 various pollen sources. *J. Econ. Entomol.* **80**, 176–183 (1987)
- 757 21. Alaux, C., Ducloz, F., Conte, D.C.Y.L.: Diet effects on honeybee immunocompetence. *Biol. Lett.* **6**, 562–565  
758 (2010)
- 759 22. Naug, D.: Nutritional stress due to habitat loss may explain recent honeybee colony collapses. *Biol Conserv*  
760 **142**, 2369–2372 (2009)
- 761 23. Engelsdorp, D.V., Hayes, J.J., Underwood, R.M., Pettis, J.: A survey of honey bee colony losses in the U.S., fall  
762 2007 to spring 2008. *PLoS ONE* **3**, 4071 (2008)
- 763 24. Neumann, P., Carreck, N.L.: Honey bee colony losses. *J Apicult Res* **49**, 1–6 (2010)
- 764 25. Engelsdorp, D.V., Meixner, M.D.: A historical review of managed honey bee populations in Europe and the  
765 United States and the factors that may affect them. *J Invertebr Pathol* **103**, 80–95 (2010)
- 766 26. Rosenkranz, P., Aumeier, P., Ziegelmann, B.: Biology and control of *Varroa destructor*. *J Invertebr Pathol* **103**,  
767 96–119 (2010)
- 768 27. Weinberg, K.P., Madel, G.: The influence of the mite *Varroa Jacobsoni* Oud. on the protein concentration and  
769 the haemolymph volume of the brood of worker bees and drones of the honey bee *Apis Mellifera* L. *Apidologie*  
770 **16**, 421–436 (1985)
- 771 28. Shen, M.Q., Cui, L.W., Ostiguy, N., Cox-Foster, D.: Intricate transmission routes and interactions between  
772 picorna-like viruses (Kashmir bee virus and sacbrood virus) with the honeybee host and the parasitic varroa  
773 mite. *J Gen Virol* **86**, 2281–2289 (2005)
- 774 29. Yang, X., Cox-Foster, D.: Effects of parasitization by *Varroa destructor* on survivorship and physiological traits  
775 of *Apis mellifera* in correlation with viral incidence and microbial challenge. *Parasitology* **134**, 405–412 (2007)

30. Yang, X.L., Cox-Foster, D.L.: Impact of an ectoparasite on the immunity and pathology of an invertebrate: Evidence for host immunosuppression and viral amplification. *P Natl Acad Sci USA* **102**, 7470–7475 (2005)
31. Emsen, B., Hamiduzzaman, M.M., Goodwin, P.H., Guzman-Novoa, E.: Lower virus infections in Varroa destructor-infested and uninfested brood and adult honey bees (*Apis mellifera*) of a low mite population growth colony compared to a high mite population growth colony. *PLoS ONE* **10**, 0118885 (2015)
32. Chen, Y.P., Siede, R.: Honey bee viruses. *Adv Virus Res* **70**, 33–80 (2007)
33. Bonning, B.C., Miller, W.A.: Dicistroviruses. *Annu Rev Entomol* **55**, 129–150 (2010)
34. Maori, E., Paldi, N., Shafir, S., Kalev, H., Tsur, E., Glick, E., Sela, I.: IAPV, a bee-affecting virus associated with Colony Collapse Disorder can be silenced by dsRNA ingestion. *Insect Mol Biol* **18**, 55–60 (2009)
35. Carrillo-Tripp, J., Dolezal, A.G., Goblirsch, M.J., Miller, W.A., Toth, A.L., Bonning, B.C.: In vivo and in vitro infection dynamics of honey bee viruses. *Sci Rep* **6**, 22265 (2016)
36. Chen, Y.P., Pettis, J.S., Corona, M., Chen, W.P., Li, C.J., Spivak, M., Visscher, P.K., DeGrandi-Hoffman, G., Boncristiani, H., Zhao, Y., van Engelsdorp, D., Delaplane, K., Solter, L., Drummond, F., Kramer, M., Lipkin, W.I., Palacios, G., Hamilton, M.C., Smith, B., Huang, S.K., Zheng, H.Q., Li, J.L., Zhang, X., Zhou, X.F., Wu, L.Y., Zhou, J.Z., Lee, M.-L., Teixeira, E.W., Li, Z.G., Evans, J.D.: Israeli acute paralysis virus: Epidemiology, pathogenesis and implications for honey bee health. *PLoS Pathog* **10**, 1004261 (2014)
37. DeGrandi-Hoffman, G., Chen, Y.: Nutrition, immunity and viral infections in honey bees. *Current Opinion in Insect Science* **10**, 170–176 (2015)
38. DeGrandi-Hoffman, G., Chen, Y., Huang, E., Huang, M.H.: The effect of diet on protein concentration, hypopharyngeal gland development and virus load in worker honey bees (*Apis mellifera* L.). *J Insect Physiol* **56**, 1184–1191 (2010)
39. Le Conte, Y., BRUNET, J.-L., McDonnell, C., Dussaubat, C., Alaux, C.: Interactions Between Risk Factors in Honey Bees
40. Annoscia, D., Zanni, V., Galbraith, D., Quirici, A., Grozinger, C., Bortolomeazzi, R., Nazzi, F.: Elucidating the mechanisms underlying the beneficial health effects of dietary pollen on honey bees (*Apis mellifera*) infested by Varroa mite ectoparasites. *Scientific Reports* **7**, 6258 (2017)
41. Nazzi, F., Pennacchio, F.: Honey bee antiviral immune barriers as affected by multiple stress factors: A novel paradigm to interpret colony health decline and collapse. *Viruses* **10**, 159 (2018)
42. Miller, C.V.L., Cotter, S.C.: Resistance and tolerance: The role of nutrients on pathogen dynamics and infection outcomes in an insect host. *Journal of Animal Ecology* **87**, 500–510 (2017)
43. Alaux, C., Dantec, C., Parrinello, H., Conte, Y.L.: Nutrigenomics in honey bees: digital gene expression analysis of pollen's nutritive effects on healthy and varroa-parasitized bees. *BMC Genomics* **12**, 496 (2011)
44. Galbraith, D.A., Yang, X., Niño, E.L., Yi, S., Grozinger, C.: Parallel epigenomic and transcriptomic responses to viral infection in honey bees (*Apis mellifera*). *PLoS Pathogens* **11**, 1004713 (2015)
45. Hemphill, W., Rivera, O., Talbert, M.: RNA-Sequencing of *Drosophila melanogaster* head tissue on high-sugar and high-fat diets. *G3: Genes, Genomes, Genetics* **8**, 279–290 (2018)
46. Nazario-Yepiz, N.O., Loustalot-Laclette, M.R., Carpintero-Ponce, J., Abreu-Goodger, C., Markow, T.A.: Transcriptional responses of ecologically diverse *Drosophila* species to larval diets differing in relative sugar and protein ratios. *PLoS ONE* **12**, 0183007 (2017)
47. Mirth, C.K., Tang, H.Y., Makohon-Moore, S.C., Salhadar, S., Gokhale, R.H., Warner, R.D., Koyama, T., Riddiford, L.M., Shingleton, A.W.: Juvenile hormone regulates body size and perturbs insulin signaling in *Drosophila*. *Proceedings of the National Academy of Sciences* **25**, 201313058 (2014)
48. Xu, J., Cherry, S.: Viruses and antiviral immunity in *Drosophila*. *Dev Comp Immunol* **42**, 67–84 (2014)
49. Swevers, L., Liu, J., Smagghe, G.: Defense Mechanisms against Viral Infection in *Drosophila*: RNAi and Non-RNAi. *Viruses* **10**, 230 (2018)
50. McMenamin, A.J., Daughenbaugh, K.F., Parekh, F., Pizzorno, M.C., Flenniken, M.L.: Honey Bee and Bumble Bee Antiviral Defense. *Viruses* **10**, 395 (2018)
51. Pasquale, G.D., Salignon, M., Conte, Y.L., Belzunces, L.P., Decourtye, A., Kretzschmar, A., Suchail, S., Brunet, J.-L., Alaux, C.: Influence of pollen nutrition on honey bee health: Do pollen quality and diversity matter? *PLoS ONE* **8**, 72016 (2013)
52. Page, R.E., Laidlaw, H.H.: Full sisters and supersisters: A terminological paradigm. *Anim. Behav.* **36**, 944–945 (1988)
53. Sherman, P.W., Seeley, T.D., Reeve, H.K.: Parasites, pathogens, and polyandry in social Hymenoptera. *Am. Nat* **131**, 602–610 (1988)
54. Crozier, R.H., Page, R.E.: On being the right size: Male contributions and multiple mating in social Hymenoptera. *Behav. Ecol. Sociobiol.* **18**, 105–115 (1985)
55. Mattila, H.R., Seeley, T.D.: Genetic diversity in honey bee colonies enhances productivity and fitness. *Science* **317**, 362–364 (2007)
56. Tarpay, D.R.: Genetic diversity within honeybee colonies prevents severe infections and promotes colony growth. *Proc. R. Soc. Lond. B* **270**, 99–103 (2003)
57. Brodschneider, R., Arnold, G., Hrasnigg, N., Crailsheim, K.: Does patriline composition change over a honey bee queen's lifetime? *Insects* **3**, 857–869 (2012)
58. Hansen, K.D., Brenner, S.E., Dudoit, S.: Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research* **38**, 131 (2010)
59. Oshlack, A., Robinson, M.D., Young, M.D.: From RNA-seq reads to differential expression results. *Genome Biology* **11**, 220 (2010)
60. McIntyre, L.M., Lopiano, K.K., Morse, A.M., Amin, V., Oberg, A.L., Young, L.J., Nuzhdin, S.V.: RNAseq: Technical variability and sampling. *BMC Genomics* **12**, 293 (2011)
61. Rutter, L., Cook, D.: bigPint: Make big data pint-sized. <https://lutter.github.io/bigPint>. Accessed 22 November 2018 (2018)
62. Merklings, S.H., Overheul, G.J., van Mierlo, J.T., Arends, D., Gilissen, C., van Rij, R.P.: The heat shock response restricts virus infection in *Drosophila*. *Scientific Reports* **5**, 12758 (2015)

63. Dostert, C., Jouanguy, E., Irving, P., Troxler, L., Galiana, D., Hetru, C., Hoffmann, J.A., Imler, J.-L.: The JAK-STAT signaling pathway is required but not sufficient for the antiviral response of *Drosophila*. *Nature Immunology* **6**, 946–953 (2005)
64. Flenniken, M.L., Andino, R.: Non-specific dsRNA-mediated antiviral response in the honey bee. *PLoS ONE* **8**, 77263 (2013)
65. Brutscher, L.M., Daughenbaugh, K.F., Flenniken, M.L.: Virus and dsRNA-triggered transcriptional responses reveal key components of honey bee antiviral defense. *Scientific Reports* **7**, 6448 (2017)
66. Chen, Y.P., Pettis, J.S., Corona, M., Chen, W.P., Li, C.J., Spivak, M., Visscher, P.K., DeGrandi-Hoffman, G., Boncristiani, H., Zhao, Y., vanEngelsdorp, D., Delaplane, K., Solter, L., Drummond, F., Kramer, M., Lipkin, W.I., Palacios, G., Hamilton, M.C., Smith, B., Huang, S.K., Zheng, H.Q., Li, J.L., Zhang, X., Zhou, A.F., Wu, L.Y., Zhou, J.Z., Lee, M.-L., Teixeira, E.W., Li, Z.G., Evans, J.D.: Israeli Acute Paralysis Virus: Epidemiology, pathogenesis and implications for honey bee health. *PLoS Pathogens* **10**, 1004261 (2014)
67. Robinson, M.D., McCarthy, D.J., Smyth, G.K.: edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010)
68. Carval, D., Ferriere, R.: A unified model for the coevolution of resistance, tolerance, and virulence. *Evolution* **64**, 2988–3009 (2010)
69. Moret, Y.: Trans-generational immune priming: Specific enhancement of the antimicrobial immune response in the mealworm beetle, *Tenebrio molitor*. *Proceedings of the Royal Society B: Biological Sciences* **273**, 1399–1405 (2006)
70. Mauricio, R., Rausher, M.D., Burdick, D.S.: Variation in the defense strategies of plants: are resistance and tolerance mutually exclusive? *Ecology* **78**, 1301–1310 (1997)
71. Fornoni, J., Nunez-Farfan, J., Valverde, P.L., Rausher, M.D.: Evolution of mixed plant defense allocation against natural enemies. *Evolution* **58**, 1685–1695 (2004)
72. Restif, O., Koella, J.C.: Shared control of epidemiological traits in a coevolutionary model of host-parasite interactions. *The American Naturalist* **161**, 827–836 (2003)
73. Chambers, M.C., Schneider, D.S.: Balancing resistance and infection tolerance through metabolic means. *PNAS* **109**, 13886–13887 (2012)
74. Xu, J., Grant, G., Sabin, L.R., Gordesky-Gold, B., Yasunaga, A., Tudor, M., Cherry, S.: Transcriptional pausing controls a rapid antiviral innate immune response in *Drosophila*. *Cell Host Microbe* **12**, 531–543 (2012)
75. Ryabov, E.V., Fannon, J.M., Moore, J.D., Wood, G.R., Evans, D.J.: 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**, 1591 (2016)
76. Johnson, B.R., Atallah, J., Plachetzki, D.C.: The importance of tissue specificity for RNA-seq: highlighting the errors of composite structure extractions. *BMC Genomics* **14**, 586 (2013)
77. Thissen, D., Steinberg, L., Kuang, D.: Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *J Educ Behav Stat* **27**, 77–83 (2002)
78. Consortium, H.B.G.S.: Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* **15**, 86 (2014)
79. Elsik, C.G., Tayal, A., Diesh, C.M., Unni, D.R., Emery, M.L., Nguyen, H.N., Hagen, D.E.: Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine. *Nucleic Acids Research* **4**, 793–800 (2016)
80. Wu, T.D., Reeder, J., Lawrence, M., Becker, G., Brauer, M.J.: GMAP and GSNAP for genomic sequence alignment: Enhancements to speed, accuracy, and functionality. *Methods Mol Biol* **1418**, 283–334 (2016)
81. Love, M.I., Huber, W., Anders, S.: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014)
82. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K.: limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Research* **43**(7), 47 (2015)
83. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995)
84. Larsson, J.: eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. (2018). R package version 4.0.0. <https://cran.r-project.org/package=eulerr>
85. I.T. Jolliffe: Principal Component Analysis. Springer, Berlin, Heidelberg (2002)
86. Inselberg, A.: The plane with parallel coordinates. *The Visual Computer* **1**, 69–91 (1985)
87. W.S. Cleveland: Visualizing Data. Hobart Press, Summit, New Jersey (1993)
88. Cook, D., Hofmann, H., Lee, E., Yang, H., Nikolau, B., Wurtele, E.: Exploring gene expression data, using plots. *Journal of Data Science* **5**, 151–182 (2007)
89. Chandrasekhar, T., Thangavel, K., Elayaraja, E.: Effective Clustering Algorithms for Gene Expression Data. *International Journal of Computer Applications* **32**, 4 (2011)
90. de Souto D. de Araujo, M., Costa, I., Soares, R., Luderger, T., Schliep, A.: Comparative Study on Normalization Procedures for Cluster Analysis of Gene Expression Datasets. *International Joint Conference on Neural Networks*, 2793–2799 (2008)
91. Huang, D.W., Sherman, B.T., Lempicki, R.: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009)
92. Huang, D.W., Sherman, B.T., Lempicki, R.A.: Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1–13 (2009)
93. Supek, F., Bošnjak, M., Škunca, N., Šmuc, T.: REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS ONE* **6**, 21800 (2011)
94. Edgar, R., Domrachev, M., Lash, A.E.: Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207–210 (2002)
95. Schlicker, A., Domingues, F.S., Rahnenfuhrer, J., Lengauer, T.: A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics* **7**, 302 (2006)

918 **Figures**

**Figure 1 Mortality rates for the four treatment groups, two virus groups, and two diet groups.** Left to right: Mortality rates for the four treatment groups, two virus groups, and two diet groups. “N” represents non-inoculation, “V” represents viral inoculation, “C” represents chestnut pollen, and “R” represents rockrose pollen. The mortality rate data included 59 samples with 15 replicates per treatment group, except for the “NC” group having 14 replicates. ANOVA values and p-values for the statistical tests are listed in the text of the paper. The letters above the bars represent significant differences with a confidence level of 95%.

**Figure 2 IAPV titers for the four treatment groups, two virus groups, and two diet groups.** Left to right: IAPV titers for the four treatment groups, two virus groups, and two diet groups. “N” represents non-inoculation, “V” represents viral inoculation, “C” represents chestnut pollen, and “R” represents rockrose pollen. The IAPV titer data included 38 samples with 10 replicates per treatment group, except for the “NR” group having 8 replicates. ANOVA values and p-values for the statistical tests are listed in the text of the paper. The letters above the bars represent significant differences with a confidence level of 95%.

**Figure 3 Parallel coordinate plots of the 1,019 virus-related DEGs of the Galbraith data [44].** Parallel coordinate plots of the 1,019 DEGs after hierarchical clustering of size four between the virus-infected and control groups of the Galbraith study. “N” represents non-inoculation, “V” represents viral inoculation. Clusters 1, 2, and 4 seem to represent DEGs that were overexpressed in the virus inoculated group, and Cluster 3 seems to represent DEGs that were overexpressed in the non-inoculated control group. In general, the DEGs appeared as expected, but there is rather noticeable deviation of the first replicate from the virus-treated sample (“V.1”) from the other virus-treated replicates in Cluster 1.

**Figure 4 Parallel coordinate plots of the 43 virus-related DEGs of our data.** Parallel coordinate plots of the 43 DEGs after hierarchical clustering of size four between the virus-infected and control groups of our study. “N” represents non-infected control group, and “V” represents treatment of virus. The vertical red line indicates the distinction between treatment groups. We see from this plot that the DEG designations for this dataset do not appear as clean compared to what we saw in the Galbraith dataset in Figure 3.

**Figure 5 Gene ontology results for the 122 “tolerance” and 125 “resistance” DEG candidates in our data.** GO analysis results for the 122 DEGs related to our “tolerance” hypothesis (A) and for the 125 DEGs related to our “resistance” hypothesis (B). The color and size of the circles both represent the number of genes in that ontology. The x-axis and y-axis are organized by SimRel, a semantic similarity metric [95].

**Figure 6 Venn diagrams comparing the virus-related DEG overlaps between our dataset and the Galbraith dataset.** Venn diagrams comparing the virus-related DEG overlaps between the Galbraith study (labeled as “G”) and our study (labeled as “R”). From left to right: Total virus-related DEGs (subplot A), virus-upregulated DEGs (subplot B), control-upregulated DEGs (subplot C). Both the total virus-related and virus-upregulated DEGs showed significant overlap between the studies ( $p\text{-value} < 2.2\text{e-}16$ ) as per Fisher’s Exact Test for Count Data. There was one gene that was virus-upregulated in the Galbraith study but control-upregulated in our study.

919 **Tables**

BeeBase ID	Gene Name	Known functions	Us	Galbraith
GB41545	MD-2-related lipid-recognition protein-like	Implicated in lipid recognition, particularly in the recognition of pathogen related products	N	-
GB50955	Protein argonaute-2	Interacts with small interfering RNAs to form RNA-induced silencing complexes which target and cleave transcripts that are mostly from viruses and transposons	V	V
GB48755	UBA-like domain-containing protein 2	Found in diverse proteins involved in ubiquitin/proteasome pathways	V	V
GB47407	Histone H4	Capable of affecting transcription, DNA repair, and DNA replication when post-transcriptionally modified	V	V
GB42313	Leishmanolysin-like peptidase	Encodes a protein involved in cell migration and invasion; implicated in mitotic progression in <i>D. melanogaster</i>	V	V
GB50813	Rho guanine nucleotide exchange factor 11	Implicated in regulation of apoptotic processes, cell growth, signal transduction, and transcription	V	V
GB54503	Thioredoxin domain-containing protein	Serves as a general protein disulphide oxidoreductase	N	-
GB53500	Transcriptional regulator Myc-B	Regulator gene that codes for a transcription factor	V	V
GB51305	Tropomyosin-like	Related to protein involved in muscle contraction	N	N
GB50178	Cilia and flagella-associated protein 61-like	Induces components required for wild-type motility and stable assembly of motile cilia	V	V

**Table 1** Known functions of the mapped subset of 43 DEGs in the virus main effect of our study. Whether the gene was overrepresented in the virus or non-virus group is also indicated for both our study and the Galbraith study. Functionalities were extracted from Flybase, National Center for Biotechnology Information and The European Bioinformatics Institute databases.



Contrast	DEGs	Interpretation	Results
V (all) vs N (all)	43	Genes that change expression due to virus effect regardless of diet status in bees	Table 1
NC vs NR	941	Genes that change expression due to diet effect in uninfected bees	Supplementary tables 4 and 5, Additional file 1
VC vs VR	376	Genes that change expression due to diet effect in infected bees	Supplementary tables 6 and 7, Additional file 1
VC upregulated in VC vs VR, and NC upregulated in NC vs NR	122	“Tolerance” genes that turn on by good diet regardless of virus infection status in bees	Figure 5A
VC upregulated in VC vs VR, but NC not upregulated in NC vs NR	125	“Resistance” genes that turn on by good diet only in infected bees	Figure 5B

**Table 2** Contrasts in our study for assessing GO and pathways analysis.

**Additional Files**

Additional file 1 — Supplementary tables.

**Table 1:** Number of DEGs across three analysis pipelines for (A) the diet main effect in our study, (B) the virus main effect in our study, and (C) the virus main effect in the Galbraith study. For the diet effects, “C” represents chestnut diet and “R” represents rockrose diet. For the virus effects, “N” represents control non-inoculated and “V” represents virus-inoculated. **Table 2:** Pathways related to the 1,033 DEGs that were upregulated in the chestnut treatment from the diet main effect. **Table 3:** Pathways related to the 881 DEGs that were upregulated in the rockrose treatment from the diet main effect. **Table 4:** GO analysis results for the 601 DEGs that were upregulated in the NC treatment from the NC versus NR treatment pair analysis. These DEGs represent genes that are upregulated when non-infected honey bees are given high quality chestnut pollen compared to being given low quality rockrose pollen. **Table 5:** GO analysis results for the 340 DEGs that were upregulated in the NR treatment from the NC versus NR treatment pair analysis. These DEGs represent genes that are upregulated when non-infected honey bees are given low quality rockrose pollen compared to being given high quality chestnut pollen. **Table 6:** GO analysis results for the 247 DEGs that were upregulated in the VC treatment from the VC versus VR treatment pair analysis. These DEGs represent genes that are upregulated when infected honey bees are given high quality chestnut pollen compared to being given low quality rockrose pollen. **Table 7:** GO analysis results for the 129 DEGs that were upregulated in the VR treatment from the VC versus VR treatment pair analysis. These DEGs represent genes that are upregulated when infected honey bees are given low quality rockrose pollen compared to being given high quality chestnut pollen. **Table 8:** Number of DEGs across three analysis pipelines for all six treatment pair combinations between the diet and virus factor. “C” represents chestnut diet, “R” represents rockrose diet, “V” represents virus-inoculated, and “N” represents control non-inoculated. **Table 9:** Kruskal-Wallis p-value and Bonferroni corrections for the 36 combinations of DEG lists, pathogen response metrics, and cluster number. (XLS).

Additional file 2 — PCA plots for the Galbraith dataset and for our dataset.

PCA plots for the Galbraith dataset (A) and for our dataset (B). “V” represents virus-inoculated, and “N” represents control non-inoculated. The x-axis represents the principal component with the most variation and the y-axis represents the principal component with the second-most variation (PNG).

Additional file 3 — Parallel coordinate lines of the diet-related DEGs of our dataset.

Parallel coordinate plots of the 1,914 DEGs after hierarchical clustering of size six between the chestnut and rockrose groups of our study. Here “C” represents chestnut samples, and “R” represents rockrose samples. The vertical red line indicates the distinction between treatment groups. We see from this plot that the DEG designations for this dataset do not appear as clean compared to what we saw in the Galbraith dataset in Figure 3 (PNG).

Additional file 4 — Example litre plots from the virus-related DEGs of our dataset.

Example litre plots of the nine DEGs with the lowest FDR values from the 43 virus-related DEGs of our dataset. “N” represents non-infected control samples and “V” represents virus-treated samples. Most of the magenta points (representing the 144 combinations of samples between treatment groups for a given DEG) do not reflect the expected pattern as clearly compared to what we saw in the litre plots of the Galbraith data. They are not as clustered together (representing replicate inconsistency) and they sometimes cross the x=y line (representing lack of difference between treatment groups). This finding reflects what we saw in the messy looking parallel coordinate lines of Figure 4 (PNG).

959 Additional file 5 — Example litre plots of DEGs from Cluster 1 of the Galbraith dataset.

960 Example litre plots of the nine DEGs with the lowest FDR values from the 365 DEGs in Cluster 1 (originally shown  
961 in Figure 3) of the Galbraith dataset. "N" represents non-infected control samples and "V" represents virus-treated  
962 samples. Most of the light orange points (representing the nine combinations of samples between treatment groups  
963 for a given DEG) deviate from the  $x=y$  line in a tight bundle as expected (PNG).

964 Additional file 6 — Example litre plots of DEGs from Cluster 2 of the Galbraith dataset.

965 Example litre plots of the nine DEGs with the lowest FDR values from the 327 DEGs in Cluster 2 (originally shown  
966 in Figure 3) of the Galbraith dataset. "N" represents non-infected control samples and "V" represents virus-treated  
967 samples. Most of the dark orange points (representing the nine combinations of samples between treatment groups  
968 for a given DEG) deviate from the  $x=y$  line in a compact clump as expected. However, they are not as tightly  
969 bunched together compared to what we saw in the example litre plots of Cluster 1 (shown in Additional file 5). As a  
970 result, what we see in these litre plots reflects what we saw in the parallel coordinate lines of Figure 3: The replicate  
971 consistency in the Cluster 1 DEGs is not as clean as that in the Cluster 2 DEGs, but is still relatively clean (PNG).

972 Additional file 7 — Scatterplot matrix of DEGs from Cluster 1 of the Galbraith dataset.

973 The 365 DEGs from the first cluster of the Galbraith dataset (originally shown in Figure 3) superimposed as light  
974 orange dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. "N"  
975 represents non-infected control samples and "V" represents virus-treated samples. We confirm that the DEGs  
976 mostly follow the expected structure, with their placement deviating from the  $x=y$  line in the treatment  
977 scatterplots, but adhering to the  $x=y$  line in the replicate scatterplots. However, we do see that sample "V.1" may  
978 be somewhat inconsistent in these DEGs, as its presence in the replicate scatterplots shows DEGs deviating from  
979 the  $x=y$  line more than expected and its presence in the treatment scatterplots shows DEGs adhering to the  $x=y$   
980 line more than expected. This inconsistent sample was something we observed in Figure 3 (PNG).

981 Additional file 8 — Scatterplot matrix of DEGs from Cluster 2 of the Galbraith dataset.

982 The 327 DEGs from the second cluster of the Galbraith dataset (originally shown in Figure 3) superimposed as dark  
983 orange dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. "N"  
984 represents non-infected control samples and "V" represents virus-treated samples. We confirm that the DEGs  
985 mostly follow the expected structure, with their placement deviating from the  $x=y$  line in the treatment  
986 scatterplots, but adhering to the  $x=y$  line in the replicate scatterplots (PNG).

987 Additional file 9 — Scatterplot matrix of DEGs from Cluster 3 of the Galbraith dataset.

988 The 224 DEGs from the third cluster of the Galbraith dataset (originally shown in Figure 3) superimposed as  
989 turquoise dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. "N"  
990 represents non-infected control samples and "V" represents virus-treated samples. We confirm that the DEGs  
991 mostly follow the expected structure, with their placement deviating from the  $x=y$  line in the treatment  
992 scatterplots, but adhering to the  $x=y$  line in the replicate scatterplots (PNG).

993 Additional file 10 — Scatterplot matrix of DEGs from Cluster 4 of the Galbraith dataset.

994 The 103 DEGs from the fourth cluster of the Galbraith dataset (originally shown in Figure 3) superimposed as pink  
995 dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. "N"  
996 represents non-infected control samples and "V" represents virus-treated samples. We confirm that the DEGs  
997 mostly follow the expected structure, with their placement deviating from the  $x=y$  line in the treatment  
998 scatterplots, but adhering to the  $x=y$  line in the replicate scatterplots. We also see that the second replicate from  
999 the virus-treated sample ("V.2") may be somewhat inconsistent in these DEGs, as its presence in the replicate  
1000 scatterplots results in the DEGs unexpectedly deviating from the  $x=y$  line and its presence in the treatment  
1001 scatterplots results in the DEGs unexpectedly adhering to the  $x=y$  line (PNG).

1002 Additional file 11 — Scatterplot matrix of virus-related DEGs from our dataset, showing only replicates 1, 2, and 3.

1003 The 43 virus-related DEGs from our dataset superimposed as magenta dots onto all genes in the form of a  
1004 scatterplot matrix. Only replicates 1, 2, and 3 are shown from both treatment groups. The data has been  
1005 standardized. "N" represents non-infected control samples and "V" represents virus-treated samples. We see that,  
1006 compared to the scatterplot matrices from certain clusters of the Galbraith data, the 43 DEGs from this subset of  
1007 six samples from our data do not paint as clear of a picture, sometimes unexpectedly deviating from the  $x=y$  line in  
1008 the replicate plots and sometimes unexpectedly adhering to the  $x=y$  line in the treatment plots (PNG).

1009 Additional file 12 — Scatterplot matrix of virus-related DEGs from our dataset, showing only replicates 4, 5, and 6.

1010 The 43 virus-related DEGs from our dataset superimposed as magenta dots onto all genes in the form of a  
1011 scatterplot matrix. Only replicates 4, 5, and 6 are shown from both treatment groups. The data has been  
1012 standardized. "N" represents non-infected control samples and "V" represents virus-treated samples. We see that,  
1013 compared to the scatterplot matrices from certain clusters of the Galbraith data, the 43 DEGs from this subset of  
1014 six samples from our data do not paint as clear of a picture, and most of them unexpectedly adhere to the  $x=y$  line  
1015 in the treatment plots (PNG).

1016 Additional file 13 — Scatterplot matrix of virus-related DEGs from our dataset, showing only replicates 7, 8, and 9.  
 1017 The 43 virus-related DEGs from our dataset superimposed as magenta dots onto all genes in the form of a  
 1018 scatterplot matrix. Only replicates 7, 8, and 9 are shown from both treatment groups. The data has been  
 1019 standardized. "N" represents non-infected control samples and "V" represents virus-treated samples. We see that,  
 1020 compared to the scatterplot matrices from certain clusters of the Galbraith data, the 43 DEGs from this subset of  
 1021 six samples from our data do not paint as clear of a picture, sometimes unexpectedly deviating from the  $x=y$  line in  
 1022 the replicate plots and sometimes unexpectedly adhering to the  $x=y$  line in the treatment plots (PNG).

1023 Additional file 14 — Scatterplot matrix of virus-related DEGs from our dataset, showing only replicates 10, 11, and  
 1024 12.  
 1025 The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only  
 1026 replicates 10, 11, and 12 are shown from both treatment groups. The data has been standardized. "N" represents  
 1027 non-infected control samples and "V" represents virus-treated samples. We see that, compared to the scatterplot  
 1028 matrices from certain clusters of the Galbraith data, the 43 DEGs from this subset of six samples from our data do  
 1029 not paint as clear of a picture, and most of them unexpectedly deviate from the  $x=y$  line in the virus-related  
 1030 replicate plots (PNG).

1031 Additional file 15 — Parallel coordinate plots of the "tolerance" candidate DEGs.  
 1032 Parallel coordinate plots of the 122 DEGs after hierarchical clustering of size four between the "tolerance" candidate  
 1033 DEGs. Here "N" represents non-infected control group, "V" represents treatment of virus, "C" represents  
 1034 high-quality chestnut diet, and "R" represents low-quality rockrose diet. The vertical red line indicates the  
 1035 distinction between treatment groups. We see there is considerable noise in the data (non-consistent replicate  
 1036 values), but that the general patterns of the DEGs follow what we expect based on our "tolerance" contrast (PNG).

1037 Additional file 16 — Parallel coordinate plots of the "resistance" candidate DEGs.  
 1038 Parallel coordinate plots of the 125 DEGs after hierarchical clustering of size four between the "resistance"  
 1039 candidate DEGs. Here "N" represents non-infected control group, "V" represents treatment of virus, "C" represents  
 1040 high-quality chestnut diet, and "R" represents low-quality rockrose diet. The vertical red line indicates the distinction  
 1041 between treatment groups. We see there is considerable noise in the data (non-consistent replicate values), but that  
 1042 the general patterns of the DEGs follow what we expect based on our "resistance" contrasts (PNG).

1043 Additional file 17 — Venn diagrams comparing the virus-related DEG overlaps in the Galbraith data using our  
 1044 pipeline and the pipeline used by Galbraith *et al.*  
 1045 Venn diagrams comparing the virus-related DEG overlaps of the Galbraith data from the DESeq2 bioinformatics  
 1046 pipelines used in the Galbraith study (labeled as "G.O.") and the DESeq2 bioinformatics pipelines used in our study  
 1047 (labeled as "G.R."). While we were not able to fully replicate the DEG list published in the Galbraith study, our DEG  
 1048 list maintained significant overlaps with their DEG list. From left to right: Total virus-related DEGs (subplot A),  
 1049 virus-upregulated DEGs (subplot B), control-upregulated DEGs (subplot C) (PNG).

1050 Additional file 18 — Venn diagrams of main effect DEG overlaps across DESeq2, edgeR, and limma  
 1051 Venn diagrams comparing DEG overlaps across DESeq2, edgeR, and limma for our diet main effect (top row), our  
 1052 virus main effect (middle row), and the Galbraith virus main effect (bottom row). Within a given subplot, "D"  
 1053 represents DESeq2, "E" represents edgeR, and "L" represents limma. From left to right on top row: Total  
 1054 diet-related DEGs (subplot A), chestnut-upregulated DEGs (subplot B), rockrose-upregulated DEGs (subplot C).  
 1055 From left to right on middle row: Total virus-related DEGs (subplot D), virus-upregulated DEGs (subplot E),  
 1056 control-upregulated DEGs in our data (subplot F). From left to right on bottom row: Total virus-related DEGs  
 1057 (subplot G), virus-upregulated DEGs (subplot H), control-upregulated DEGs in the Galbraith data (subplot I)  
 1058 (PNG). With the exception of the limma pipeline resulting in zero DEGs in our virus main effect analysis, we found  
 1059 significant overlaps between DEG lists across the different pipelines (DESeq2, edgeR, and limma). In general,  
 1060 DESeq2 resulted in the largest number of DEGs and limma resulted in the least number of DEGs (PNG).

1061 Additional file 19 — Analysis of correlation between DEG read counts and pathogen response metrics  
 1062 Distribution of R-squared values for DEG cluster read counts and pathogen response metrics. Columns left to right:  
 1063 SBV titers, mortality rates, and IAPV titers. Rows top to bottom: Tolerance candidate DEGs, resistance candidate  
 1064 DEGs, and virus-related DEGs. Each subplot includes five boxplots which represent the R-squared value distributions  
 1065 for four DEG clusters and all remaining non-DEGs in the data. The top number above each boxplot represents the  
 1066 number of genes included. The first four boxplots also include a bottom number, which represents the  
 1067 Kruskal-Wallis p-value of the comparison of the R-squared distribution of the cluster and the R-squared distribution  
 1068 of the non-DEG data (PNG).

1069 Additional file 20 — Tables listing DEGs for contrasts.  
 1070 **Table 1:** IDs of 1,914 DEGs in our diet main effect. **Table 2:** IDs of 43 DEGs in our virus main effect. **Table 3:** IDs  
 1071 of 178 DEGs in our NR versus VR contrast. **Table 4:** IDs of 376 DEGs in our VC versus VR contrast. **Table 5:** IDs  
 1072 of 774 DEGs in our NC versus VR contrast. **Table 6:** IDs of 955 DEGs in our VC versus NR contrast. **Table 7:** IDs  
 1073 of 941 DEGs in our NC versus NR contrast. **Table 8:** IDs of 125 resistance candidate genes. **Table 9:** IDs of 122  
 1074 tolerance candidate genes. (XLS).