

RESEARCH

Gene expression responses to diet quality and viral infection in *Apis mellifera*

Lindsay Rutter¹, Dianne Cook², Amy L. Toth^{3,4} and Adam Dolezal^{5*}

*Correspondence:

adolezal@illinois.edu

⁵Department 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 honey bees has considerable implications for the agricultural economy because honey bees are one of the leading pollinators of numerous crops. Honey bee declines have been associated with several interactive factors. Poor nutrition and viral infection are two environmental stressors that pose heightened dangers to honey bee health. In this study, we used RNA-sequencing to examine how monofloral diets and Israeli Acute Paralysis Virus inoculation influence gene expression patterns in honey 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). Somewhat unexpectedly, 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 determine if any 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 DEGs were found, suggesting both processes may play significant roles in dietary buffering from pathogen infection. We also compared the virus main effect in our study (polyandrous colonies) to that obtained in a previous study (single-drone colonies) and verified significant overlap in differential expression despite visualization methods showing differences in the noisiness levels between these two datasets.

Conclusions: Through transcriptional contrasts and functional enrichment analysis, we add to evidence of feedbacks between diet and disease in honey bees. We also show that comparing results derived from polyandrous colonies (which are typically more natural) and single-drone colonies (which usually yield more signal) may allow researchers to identify transcriptomic patterns in honey bees that are concurrently less artificial and less noisy. Altogether, we hope this work underlines possible merits of using data visualization techniques and multiple datasets when interpreting RNA-sequencing studies.

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

1 Background

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



10 Honey bee declines have been associated with several factors, including pesti-
11 cide use, parasites, pathogens, habitat loss, and poor nutrition [8, 9]. Researchers
12 generally agree that these stressors do not act in isolation; instead, they appear
13 to influence the large-scale loss of honey bees in interactive fashions as the en-
14 vironment changes [10]. Nutrition and viral infection are two broad factors that
15 pose heightened dangers to honey bee health in response to recent environmental
16 changes.

17 Pollen is the main source of nutrition (including proteins, amino acids, lipids,
18 sterols, starch, vitamins, and minerals) in honey bees [11, 12]. At the individual
19 level, pollen supplies most of the nutrients necessary for physiological development
20 [13] and is believed to have considerable impact on longevity [14]. At the colony level,
21 pollen enables young workers to produce jelly, which then nourishes larvae, drones,
22 older workers, and the queen [15, 16]. Various environmental changes (including
23 urbanization and monoculture crop production) have significantly altered the nutri-
24 tional profile available to honey bees. In particular, honey bees are confronted with
25 less diverse selections of pollen, which is of concern because mixed-pollen (polyflo-
26 ral) diets are generally considered healthier than single-pollen (monofloral) diets
27 [17, 18, 19]. Indeed, reported colony mortality rates are higher in developed land

28 areas compared to undeveloped land areas [20], and beekeepers rank poor nutrition
29 as one of the main reasons for colony losses [21]. Understanding how undiversified
30 diets affect honey bee health will be crucial to resolve problems that may arise as
31 agriculture continues to intensify throughout the world [22, 23].

32 Viral infection was a comparatively minor problem in honey bees until the last
33 century when Varroa destructor (an ectoparasitic mite) spread worldwide [24]. This
34 mite feeds on honey bee hemolymph [25], transmits cocktails of viruses, and sup-
35 ports replication of certain viruses [26, 27, 28]. More than 20 honey bee viruses have
36 been identified [29]. One of these viruses that has been linked to honey bee decline
37 is Israeli Acute Paralysis Virus (IAPV). A positive-sense RNA virus of the Dicistro-
38 viridae family [30], IAPV causes infected honey bees to display shivering wings, de-
39 creased locomotion, muscle spasms, and paralysis, and 80% of caged infected adult
40 honey bees die prematurely [31]. IAPV has demonstrated higher infectious capac-
41 ities than other honey bee viruses in certain conditions [32] and is more prevalent
42 in colonies that do not survive the winter [33]. Its role in the rising phenomenon of
43 “Colony Collapse Disorder” (in which the majority of worker bees disappear from
44 a hive) remains unclear: It has been implicated in some studies [34, 35] but not in
45 other studies [1, 30, 36]. Nonetheless, it is clear that IAPV reduces colony strength
46 and survival.

47 Although there is growing interest in how viruses and diet quality affect the health
48 and sustainability of honey bees, as well as a recognition that such factors might
49 operate interactively, there are only a small number of experimental studies thus
50 far directed toward elucidating the interactive effects of these two factors in honey
51 bees [37, 38, 39]. We recently used laboratory cages and nucleus hive experiments to
52 investigate the health effects of these two factors, and our results show the impor-
53 tance of the combined effects of both diet quality and virus infection. Specifically,

high quality pollen is able to mitigate virus-induced mortality to the level of diverse, polyfloral pollen [40].

Following up on these phenotypic findings from our previous study, we now aim to understand the corresponding underlying mechanisms by which high quality diets protect bees from virus-induced 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) [41]. 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 [42]. As it stands, there are only a small number of published experiments examining gene expression patterns related to diet effects [43] and IAPV infection effects [44] in honey bees. As far as we know, there are few to no studies investigating honey bee gene expression patterns specifically related to monofloral diets, and few to no studies investigating honey bee gene expression patterns related to the combined effects of diet in any broad sense and viral inoculation in any broad sense.

In this study, we examine how monofloral diets and viral inoculation influence gene expression patterns in honey bees by focusing on four treatment groups (low quality diet without IAPV exposure, high quality diet without IAPV exposure, low quality diet with IAPV exposure, and high quality diet with IAPV exposure). We conduct RNA-sequencing analysis on a randomly selected subset of the honey bees we used in our previous study (as is further described in our methods section). We then examine pairwise combinations of treatment groups, the main effect of

monofloral diet, the main effect of IAPV exposure, and the combined effect of the two factors on gene expression patterns.

We also compare the main effect of IAPV exposure in our dataset to that obtained in a previous study conducted by Galbraith and colleagues [44]. As RNA-sequencing data can be highly noisy, this comparison allowed us to characterize how repeatable and robust our RNA-sequencing results were in comparison to previous studies. Importantly, we use an in-depth data visualization approach to explore and corroborate our data, and suggest such an approach can be useful for cross-study comparisons and validation of noisy RNA-sequencing data in the future.

Results

Phenotypic results

We reanalyzed our previously published dataset with a subset more relevant to our RNA-sequencing approaches in the current study that have a more focused question regarding diet quality. We briefly show it again here to inform the RNA-sequencing comparison because we reduced the number of treatments (from eight to four) from the original published data [40] as a means to focus on diet quality effects.

As shown in Figure 1, mortality rates of honey bees 72 hour post-inoculation significantly differed among the treatment groups (mixed model ANOVA across all treatment groups, $df = 3, 54$; $F = 10.03$; $p < 2.30e-05$). The effect of virus treatment (mixed model ANOVA, $df = 1, 54$; $F = 24.73$; $p < 1.00e-05$) and diet treatment (mixed model ANOVA, $df = 1, 54$; $F = 5.32$; $p < 2.49e-02$) were significant, but the interaction between the two factors (mixed model ANOVA, $df = 1, 54$; $F = 4.72e-02$, $p = 8.29e-01$) was not significant. We compared mortality levels based on pairwise comparisons: For a given diet, honey bees exposed to the virus showed significantly higher mortality rate than honey bees not exposed to the virus. Namely, bees fed Rockrose pollen had significantly elevated mortality with virus infection compared to uninfected controls (Tukey HSD, $p < 1.18e-03$), and bees fed Chestnut

108 pollen similarly had significantly elevated mortality with virus infection compared
109 to controls (Tukey HSD, $p < 4.80\text{e-}03$) (Figure 1).

110 As shown in Figure 2, IAPV titers of honey bees 72 hour post-inoculation sig-
111 nificantly differed among the treatment groups (mixed model ANOVA across all
112 treatment groups, $df = 3, 33$; $F = 6.10$; $p < 1.96\text{e-}03$). The effect of virus treatment
113 (mixed model ANOVA, $df = 1, 33$; $F = 15.04$; $p < 4.75\text{e-}04$) was significant, but
114 the diet treatment (mixed model ANOVA, $df = 1, 33$; $F = 2.55$; $p = 1.20\text{e-}01$)
115 and the interaction between the two factors (mixed model ANOVA, $df = 1, 33$; F
116 $= 7.02\text{e-}01$, $p = 4.08\text{e-}01$) were not significant. We compared IAPV titer volumes
117 based on pairwise comparisons: Bees fed Rockrose pollen had significantly elevated
118 IAPV titer volumes with virus infection compared to uninfected controls (Tukey
119 HSD, $p < 5.44\text{e-}03$). However, bees fed Chestnut pollen did not have significantly
120 elevated IAPV titer volumes with virus infection compared to uninfected controls
121 (Tukey HSD, $p = 1.11\text{e-}01$). Overall, we interpreted these findings to mean that
122 high-quality Chestnut pollen could “rescue” high virus titers resulting from the in-
123 oculation treatment, whereas low-quality Rockrose pollen could not do so (Figure
124 2).

125 Main effect DEG results

126 We observed a substantially larger number of DEGs in our diet main effect ($n =$
127 1,914) than in our virus main effect ($n = 43$) (Supplementary table 1 A and B,
128 Additional file 1). In the diet factor, there were more Chestnut-upregulated DEGs
129 ($n = 1033$) than Rockrose-upregulated DEGs ($n = 881$). In the virus factor, there
130 were more virus-upregulated DEGs ($n = 38$) than control-upregulated DEGs ($n =$
131 5). While these reported DEG counts are from the DESeq2 package, we saw similar
132 trends for the edgeR and limma package results (Supplementary table 1, Additional
133 file 1 and Additional file 18).

GO analysis of the Chestnut-upregulated DEGs revealed the following over-represented categories: Wnt signaling, hippo signaling, and dorso-ventral axis formation, as well as pathways related to circadian rhythm, mRNA surveillance, insulin resistance, inositol phosphate metabolism, FoxO signaling, ECM-receptor interaction, phototransduction, Notch signaling, JaK-STAT signaling, MAPK signaling, and carbon metabolism (Supplementary table 2, Additional file 1). GO analysis of the Rockrose DEGs revealed pathways related to terpenoid backbone biosynthesis, homologous recombination, SNARE interactions in vesicular transport, aminoacyl-tRNA biosynthesis, Fanconi anemia, and pyrimidine metabolism (Supplementary table 3, Additional file 1).

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

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

Pairwise comparison of DEG results

The number of DEGs across the six treatment pairings between the diet and virus factor ranged from 0 to 955 (Supplementary table 8, Additional file 1). Some of the trends observed in the main effect comparisons persisted: The diet level appeared to have greater influence on the number of DEGs than the virus level. Across every pair comparing the Chestnut and Rockrose levels, regardless of the virus level, the number of Chestnut-upregulated DEGs was higher than the number of Rockrose-upregulated DEGs (Supplementary table 8 C, D, E, F, Additional file 1). For the

161 pairs in which the diet level was controlled, the virus-exposed treatment showed
162 equal to or more DEGs than the control treatment (Supplementary table 8 A and
163 B, Additional file 1). There were no DEGs between the treatment pair controlling
164 for the Chestnut level of the virus effect (Supplementary table 8A, Additional file
165 1). These trends were observed for all three pipelines used (DESeq2, edgeR, and
166 limma).

167 Prior study comparison results

168 We wished to explore the signal:to:noise ratio between the Galbraith dataset and our
169 dataset. Basic PCA plots were constructed with the DESeq2 analysis pipeline and
170 showed that the Galbraith dataset may better separate the infected and uninfected
171 honey bees better than our dataset (Additional file 2). We also noted that the first
172 replicate of both treatment groups in the Galbraith data did not cluster as cleanly
173 in the PCA plots. However, through this automatically-generated plot, we can only
174 visualize information at the sample level. Wanting to learn more about the data at
175 the gene level, we continued with additional visualization techniques.

176 We used parallel coordinate lines superimposed onto boxplots to visualize the
177 DEGs associated with virus infection in the two studies. The background boxplot
178 represents the distribution of all genes in the data, and each parallel coordinate
179 line represents one DEG. To reduce overplotting of parallel coordinate lines, we
180 used hierarchical clustering techniques to separate DEGs into common patterns as
181 is described in the methods section.

182 We see that the 1,019 DEGs from the Galbraith dataset form relatively clean-
183 looking visual displays (Figure 3). We do see that the first replicate of the virus
184 group appears somewhat inconsistent with the other virus replicates in Cluster 1,
185 confirming that this trend in the data that we saw in the PCA plot carried through
186 into the DEG results. In contrast, we see that the 43 virus-related DEGs from our
187 dataset do not look as clean in their visual displays (Figure 4). The replicates appear

188 somewhat inconsistent in their estimated expression levels and there is not always
189 such a large difference between treatment groups. We see a similar finding when we
190 also examine a larger subset of 1,914 diet-related DEGs from our study (Additional
191 file 3).

192 We also used litre plots to examine the structure of individual DEGs: We see
193 that indeed the individual virus DEGs from our data (Additional file 4) show less
194 consistent replications and less differences between the treatment groups compared
195 to the individual virus DEGs from the Galbraith data (Additional files 5 and 6). For
196 the Galbraith data, we examined individual DEGs from the first cluster (Additional
197 file 5) and second cluster (Additional file 6) because the first cluster had previously
198 shown less consistency in the first replicate of the treatment group (Figure 3). We
199 verify this trend again in the litre plots with the DEG points in the first cluster
200 showing less tight cluster patterns (Additional files 5 and 6).

201 Finally, we looked at scatterplot matrices to assess the DEGs. We created stan-
202 dardized scatterplot matrices for each of the four clusters (Figure 3) of the Galbraith
203 data (Additional files 7, 8, 9, and 10). We also created standardized scatterplot ma-
204 trices for our data. However, as our dataset contained 24 samples, we would need
205 to include 276 scatterplots in our matrix, which would be too numerous to allow
206 for efficient visual assessment of the data. As a result, we created four scatterplot
207 matrices of our data, each with subsets of 6 samples to be more comparable to the
208 Galbraith data (Additional files 11, 12, 13, and 14). We can again confirm through
209 these plots that the DEGs from the Galbraith data appeared more as expected:
210 Deviating more from the $x=y$ line in the treatment scatterplots while staying close
211 to the $x=y$ line in replicate scatterplots.

212 Despite the virus-related DEGs ($n = 1,019$) from the Galbraith dataset displaying
213 the expected patterns more than those from our dataset ($n = 43$), there was signif-
214 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 results

Using the contrasts specified in Table 2, we discovered 122 “tolerance” candidate DEGs and 125 “resistance” candidate DEGs. We again used parallel coordinate lines superimposed onto boxplots to visualize these candidate DEGs. 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).

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

242 found functions related to metabolism (such as carbohydrate metabolism, chitin
243 metabolism, oligosaccharide biosynthesis, and general metabolism) (Figure 5B).

244 Post hoc analysis results

245 In general, the R-squared values between gene read counts and physiological mea-
246 surements were low ($R\text{-squared} < 0.1$). However, some DEG clusters showed slightly
247 larger R-squared values than the non-DEG group (the rest of the data). One promi-
248 nent example of this includes the first and second cluster of the virus-related DEGs
249 (Additional file 19I). The Kruskal–Wallis test was used to determine if R-squared
250 populations of DEG clusters significantly differed from those of the rest of the data.
251 The p-values and Bonferroni correction values for each of the 36 tests (as described
252 in the methods section) is provided in Supplementary table 9, Additional file 1. An
253 overall trend emerges to suggest that DEGs may have significantly larger correla-
254 tion with the physiological measurements compared to non-DEGs. It is difficult to
255 interpret these results in light of the noisiness of this data, but it may be of interest
256 to conduct further studies examining differential expression between physiological
257 measurements.

258 Discussion

259 Challenges to honey bee health are a growing concern, in particular the combined,
260 interactive effects of nutritional stress and pathogens (Dolezal and Toth 2018). In
261 this study, we used RNA-sequencing to probe mechanisms underlying honey bee
262 responses to two effects, diet quality and infection with the prominent virus of
263 concern, IAPV. In general, we found a major nutritional transcriptomic response,
264 with nearly 2,000 transcripts changing in response to diet quality (rockrose/poor
265 diet versus chestnut/good diet). The majority of these genes were upregulated in
266 response to high quality diet, and these genes were over-represented for functions
267 (Supplementary table 2, Additional file 1) such as nutrient signaling metabolism (in-
268 sulin resistance) and immune response (Notch signaling and JaK-STAT pathways).

269 These data suggest high quality nutrition may allow bees to alter their metabolism,
270 favoring investment of energy into innate immune responses.

271 Somewhat surprisingly, the transcriptomic response to virus infection in our exper-
272 iment was fairly limited. We found only 43 transcripts to be differentially expressed,
273 some with known immune functions (Table 1) such as argonaute-2 and a gene with
274 similarity to MD-2 lipid recognition protein, as well as genes related to transcrip-
275 tional regulation and muscle contraction. The small number of DEGs in this study
276 may be partly explained by the large amount of noise in the data (Figure 4 and
277 Additional files 2B, 4, 11, 12, 13, and 14).

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

290 Data visualization is a useful method to identify noise and robustness in RNA-
291 sequencing data. In this study, we used extensive data visualization to improve the
292 interpretation of our RNA-sequencing results. For example, the DESeq2 package
293 comes with certain visualization options that are popular in RNA-sequencing anal-
294 ysis. One of these visualization is the principal component analysis (PCA) plot,
295 which allows users to visualize the similarity between samples within a dataset. We

could determine from this plot that indeed the Galbraith data may show more similarity between its replicates and differences between its treatments compared to our data (Additional file 2). However, the PCA plot only shows us information at the sample level. We wanted to investigate how these differences in the signal:to:noise ratios of the datasets would affect the structure of any resulting DEGs. As a result, we also used three plotting techniques from the bigPint package: We investigated the 1,019 virus-related DEGs from the Galbraith dataset and the 43 virus-related DEGs from our dataset using parallel coordinate lines, scatterplot matrices, and litre plots. To prevent overplotting issues in our graphics, we used a hierarchical clustering technique for the parallel coordinate lines to separate the set of DEGs into smaller groups. We also needed to examine four subsets of samples from our dataset to make effective use of the scatterplot matrices. After these tailorizations, we determined that the same patterns we saw in the PCA plots regarding the entire dataset extended down the pipeline analysis into the DEG calls: Even the DEGs from the Galbraith dataset showed more similarity between their replicates and differences between their treatments compared to those from our data. However, the 365 DEGs from the Galbraith data in Cluster 1 of Figure 3 showed an inconsistent first replicate in the treatment group (“V.1”), which was something we observed in the PCA plot. This indicates that this feature also extended down the analysis pipeline into DEG calls. Despite the differences in signal between these two datasets, there was substantial overlap in the resulting DEGs. We believe these visualization applications can be useful for future researchers analyzing RNA-sequencing data to quickly and effectively ensure that the DEG calls look reliable or at least overlap with DEG calls from similar studies that look reliable. We also expect this type of visualization exploration can be especially crucial when studying complex organisms that do not have genetic identicalness or similarity between replicates and/or when using experiments that may lack rigid design control.

One of the goals of this study was to use our RNA-sequencing data to assess whether transcriptomic responses to diet quality and virus infection provide insight into whether high quality diet can buffer bees from pathogen stress via mechanisms of “resistance” or “tolerance”. Recent evidence has suggested that overall immunity is determined by more than just “resistance” (the reduction of pathogen fitness within the host by mechanisms of avoidance and control) [45]. Instead, overall immunity is related to “resistance” in conjunction with “tolerance” (the reduction of adverse effects and disease resulting from pathogens by mechanisms of healing) [41, 45]. Immune-mediated resistance and diet-driven tolerance mechanisms are costly and may compete with each other [41, 46]. Data and models have suggested that selection can favor an optimum combination of both resistance and tolerance [47, 48, 49, 50]. We attempted to address this topic through specific gene expression contrasts (Table 2), accompanied by GO analysis of the associated gene lists. We found an approximately equal number of resistance ($n = 125$) and tolerance ($n = 122$) related candidate DEGs, suggesting both processes may be playing significant roles in dietary buffering from pathogen induced mortality. Resistance candidate DEGs had functions related to several forms of metabolism (chitin and carbohydrate), regulation of transcription, and cell adhesion. Tolerance candidate DEGs had functions related to carbohydrate metabolism and chitin metabolism; however, they also showed functions related to immune response, including RNA polymerase II transcription and regulation of response to reactive oxygen species (Figure 5A). Previous studies have shown that transcriptional pausing of RNA polymerase II may be an innate immune response in *D. melanogaster* that allows for a more rapid response by increasing the accessibility of promoter regions of virally induced genes [51]. Moreover, circulating haemocytes in insects encapsulate and nodulate pathogens by forming a barrier between the pathogen and the host tissues. This barrier undergoes apoptosis and melanization through the phenoloxidase

350 enzyme cascade, which produces reactive oxygen species [41, 52, 53]. These possi-
351 ble immunological defense mechanisms within our “tolerance” candidate DEGs and
352 metabolic processes within our “resistance” candidate DEGs may provide additional
353 evidence of feedbacks between diet and disease in honey bees [42].

354 There were several limitations in this study that could be improved upon in fu-
355 ture studies. For instance, our comparison between the Galbraith data (single-drone
356 colonies) and our data (polyandrous colonies) was limited by numerous extraneous
357 variables between these studies. In addition to different molecular pipelines and
358 bioinformatic preprocessing pipelines used between these studies, the Galbraith
359 study focused on one-day old worker honey bees that were fed sugar and MegaBee
360 diet, whereas our study focused on adult worker honey bees that were fed monofloral
361 diets. Furthermore, the Galbraith data used eviscerated abdomens with attached
362 fat bodies and only considered symptomatic honey bees for their infected treat-
363 ment group, whereas we used whole bodies and considered both asymptomatic and
364 symptomatic honey bees for our infected treatment group. Further differences be-
365 tween the studies can be found in their corresponding published methods sections
366 [40, 44]. Our comparative visualization assessment between these two datasets was
367 also somewhat limited because the virus effect in the Galbraith study used three
368 replicates for each level, whereas the virus effect in our study used twelve replicates
369 for each level that were actually further subdivided into six replicates for each diet
370 level. Hence the apparent reduction in noise observed in the Galbraith data com-
371 pared to our data in the PCA plots, parallel coordinate plots, scatterplot matrices,
372 and litre plots may be an inadvertent product of the smaller number of replicates
373 used and the lack of a secondary treatment group rather than solely the reduction in
374 genetic variability through the single-drone colony design itself. With this in mind,
375 while our current efforts may be a starting point, future studies can shed more
376 light on signal:to:noise and differential expression differences between polyandrous

colony designs and single-drone colony designs by controlling for extraneous factors more strictly than what we were able to do in the current line of work.

In addition, this study used a whole body RNA-sequencing approach. In future related studies, it may be informative to use tissue-specific methods. Recent evidence has suggested that RNA-sequencing approaches toward composite structures in honey bees leads to false negatives, implying that genes strongly differentially expressed in particular structures may not reach significance within the composite structure. On a similar note, recent studies have also found that within a composite extraction, structures therein may contain opposite patterns of differential expression. We can provide more detailed answers to our original transcriptomic questions if we were to repeat this same experimental design only now at a more refined tissue level [54]. Another future direction related to this work would be to integrate multiple omics datasets to investigate monofloral diet quality and IAPV infection in honey bees. Indeed, previous studies in honey bees have found that multiple omics datasets do not always align in a clear-cut manner, and hence may broaden our understanding of the molecular mechanisms being explored [44].

Conclusions

To the best of our knowledge, there are few to no studies investigating honey bee gene expression specifically related to monofloral diets, and few to no studies examining honey bee gene expression related to the combined effects of diet in any general sense and viral inoculation in any general sense. It also remains unknown whether the protective effects of good diet in honey bees is due to direct effects on immune function (resistance) or indirect effects of energy availability on vigor and health (tolerance). We attempted to address these unresolved areas by conducting a two-factor RNA-sequencing study that examined how monofloral diets and IAPV inoculation influence gene expression patterns in honey bees. Overall, our data suggest complex transcriptomic responses to multiple stressors in honey bees. Diet has

404 the capacity for large and profound effects on gene expression and may set up the
405 potential for both resistance and tolerance to viral infection, adding to previous
406 evidence of possible feedbacks between diet and disease in honey bees [42].

407 Moreover, this study also demonstrated the benefits of using data visualizations
408 and multiple datasets to address inherently messy biological data. For instance,
409 by verifying the substantial overlap in our DEG lists to those obtained in another
410 study that addressed a similar question using specimens with less genetic variability,
411 we were able to place much higher confidence in the differential gene expression
412 results from our otherwise noisy data. We also suggested that comparing results
413 derived from polyandrous colony designs (which are usually more natural) and
414 single-drone colony designs (which usually have more signal) may allow researchers
415 to identify transcriptomic patterns in honey bees that are concurrently more realistic
416 and less noisy. Altogether, we hope our results underline the merits of using data
417 visualization techniques and multiple datasets to understand and interpret RNA-
418 sequencing datasets.

419 **Methods**

420 Details of the procedures we used to prepare virus inoculum, infect and feed caged
421 honey bees, and quantify IAPV can be reviewed in our previous work [40]. The
422 statistical analysis we used to study the main and interaction effects of the two
423 factors on mortality and IAPV titers is also described in our earlier report [40].

424 **Design of two-factor experiment**

425 There are several reasons why, in the current study, we focused only on diet quality
426 (monofloral diets) as opposed to diet diversity (monofloral diets versus polyfloral
427 diets). First, when assessing diet diversity, a sugar diet is often used as a control.
428 However, such an experimental design does not reflect real-world conditions for
429 honey bees as they rarely face a total lack of pollen [55]. Second, in studies that
430 compared honey bee health using monofloral and polyfloral diets at the same time,

431 if the polyfloral diet and one of the high-quality monofloral diets both exhibited
432 similarly beneficial effects, then it was difficult for the authors to assess if the
433 polyfloral diet was better than most of the monofloral diets because of its diversity
434 or because it contained as a subset the high-quality monofloral diet [55]. Third, as
435 was previously mentioned, honey bees are now confronted with less diverse sources
436 of pollen. As a result, there is a need to better understand how monofloral diets
437 affect honey bee health.

438 Consequently, for our nutrition factor, we examined two monofloral pollen diets,
439 Rockrose (*Cistus*) and Castanea (Chestnut). Rockrose pollen is generally considered
440 less nutritious than Chestnut pollen due to its lower levels of protein, amino acids,
441 antioxidants, calcium, and iron [55, 40]. For our virus factor, one level contained
442 bees that were infected with IAPV and another level contained bees that were not
443 infected with IAPV. This experimental design resulted in four treatment groups
444 (Rockrose pollen without IAPV exposure, Chestnut pollen without IAPV exposure,
445 Rockrose pollen with IAPV exposure, and Chestnut pollen with IAPV exposure)
446 that allowed us to assess main effects and interactive effects between diet quality
447 and IAPV infection in honey bees.

448 RNA extraction

449 Fifteen cages per treatment were originally sampled. Six live honey bees from each
450 cage were randomly selected 36 hours post inoculation and placed into tubes [32].
451 Tubes were kept on dry ice and then transferred into a -80C freezer until processing.
452 Eight cages were randomly selected from the original 15 cages, and 2 honey bees per
453 cage were randomly selected from the original six live honey bees per cage. Whole
454 body RNA from each pool of two honey bees were extracted using Qiagen RNeasy
455 MiniKit followed by Qiagen DNase treatment. Samples were suspended in water
456 to 200-400 ng/ μ l. All samples were then tested on a Bioanalyzer at the DNA core
457 facility to ensure quality (RIN > 8).

458 Gene expression

459 Samples were sequenced starting on January 14, 2016 at the Iowa State University
460 DNA Facility (Platform: Illumina HiSeq Sequencing; Category: Single End 100 cycle
461 sequencing). A standard Illumina mRNA library was prepared by the DNA facility.
462 Reads were aligned to the BeeBase Version 3.2 genome [56] from the Hymenoptera
463 Genome Database [57] using the programs GMAP and GSNAP [58]. There were
464 four lanes of sequencing with 24 samples per lane. Each sample was run twice.
465 Approximately 75-90% of reads were mapped to the honey bee genome. Each lane
466 produced around 13 million single-end 100 basepair reads.

467 We tested all six pairwise combinations of treatments for DEGs (pairwise DEGs).
468 We also tested the diet main effect (diet DEGs), virus main effect (virus DEGs), and
469 interaction term for DEGs (interaction DEGs). We then also tested for virus main
470 effect DEGs (virus DEGs) in public data derived from a previous study exploring
471 the gene expression of IAPV virus infection in honey bees [44]. We tested each
472 DEG analysis using recommended parameters with DESeq2 [59], edgeR [60], and
473 LimmaVoom [61]. In all cases, we used a false discovery rate (FDR) threshold of 0.05
474 [62]. Fisher's exact test was used to determine significant overlaps between DEG
475 sets (whether from the same dataset but across different analysis pipelines or from
476 different datasets across the same analysis pipelines). The eulerr shiny application
477 was used to construct Venn diagram overlap images [63]. In the end, we focused on
478 the DEG results from DESeq2 [59] as this pipeline was also used in the Galbraith
479 study [44]. We used the independent filtering process built into the DESeq2 software
480 that mitigates multiple comparison corrections on genes with no power rather than
481 defining one filtering threshold.

482 Comparison to prior studies on transcriptomic response to viral infection

483 We compare the main effect of IAPV exposure in our dataset to that obtained in a
484 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 paper. 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 paper. 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).

While our study examines honey bees from polyandrous colonies, the Galbraith study examined honey bees from single-drone colonies. As a consequence, the honey bees in our study will be on average 25% genetically identical, whereas honey bees from the Galbraith study will be on average 75% genetically identical [64]. We should therefore expect that the Galbraith study may generate data with lower signal:to:noise ratios than our data due to the 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 [65], increased tolerance to environmental changes [66], and increased colony performance [67, 68]. Given that honey bees are naturally very polyandrous [69], our honey bees may also reflect more realistic environmental and genetic simulations. 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. We wish to explore how the gene expression effects of IAPV inoculation compare between these two studies that used such different experimental designs. To achieve this objective, 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. It is our hope that this aspect of our study may shine light on

512 how experimental designs that control genetic variability to different extents might
513 affect the resulting gene expression data in honey bees.

514 Visualization

515 We used an array of visualization tools as part of our analysis. We first used popular
516 tools like the PCA plot [70] from the DESeq2 package. After that, we used mul-
517 tivariate visualization tools from our work-in-progress R package called bigPint.
518 Specifically, we used parallel coordinate plots [71], scatterplot matrices [72], and
519 litre plots (which we recently developed based on “replicate line plots” [73] (cite
520 bigPint too)) to assess the variability between the replicates and the treatments in
521 our data. We also used these plotting techniques to assess for normalization prob-
522 lems and other common problems in RNA-sequencing analysis pipelines [73] (cite
523 bigPint too).

524 We also used statistical graphics to better understand patterns in our DEGs. How-
525 ever, in cases of large DEG lists, these visualization tools had overplotting problems
526 (where multiple objects are drawn on top of one another, making it impossible to
527 detect individual values). To remedy this problem, we first standardized each DEG
528 to have a mean of zero and standard deviation of unity [74, 75]. Then, we performed
529 hierarchical clustering on the standardized DEGs using Ward’s linkage. This process
530 divided large DEG lists into smaller clusters of similar patterns, which allowed us
531 to more efficiently visualize the different types of patterns within large DEG lists.

532 Gene ontology

533 DEGs were uploaded as a background list to DAVID Bioinformatics Resources 6.7
534 [76, 77]. The overrepresented gene ontology (GO) terms of DEGs were determined
535 using the BEEBASE.ID identifier option (honey bee gene model) in the DAVID
536 software. To fine-tune the GO term list, only terms correlating to Biological Pro-
537 cesses were considered. The refined GO term list was then imported into REVIGO
538 [78], which uses semantic similarity measures to cluster long lists of GO terms.

539 Probing tolerance versus resistance

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

552 Post hoc analysis

553 We found considerable noisiness in our data and saw, through gene-level visual-
554 izations, that our DEGs contained outliers and inconsistent replicates. Hence, we
555 wanted to explore whether our DEG read counts correlated with physiological met-
556 rics, including IAPV titers, Schmallenberg Virus (SBV) titers, and mortality rates.
557 For this process, we considered virus main effect DEGs (Figure 4), “tolerance can-
558 didate” DEGs (Additional file 15), and “resistance candidate” DEGs (Additional
559 file 16). For each DEG in each cluster, we calculated a coefficient of determination
560 (R-squared) value to estimate the correlation between its raw read counts and the
561 physiological metrics across its 24 samples. We then used the Kruskal–Wallis test
562 to determine if the distribution of the R-squared values in any of the DEG clusters
563 significantly differed from those in the non-DEG genes (the rest of the data). As
564 there were four clusters for each of the nine combinations of DEG lists (“tolerance”
565 candidate DEGs, “resistance” candidate DEGs, and virus-related DEGs) and phys-

iological measurements (IAPV titer, SBV titer, and mortality rate), this process
resulted in 36 statistical tests.

568 **Ethics approval and consent to participate**

569 Not applicable.

570 **Consent for publication**

571 Not applicable.

572 **Availability of data and materials**

573 The datasets generated and/or analysed during the current study are available in the [NAME] repository,
574 [PERSISTENT WEB LINK TO DATASETS]. Include our data, Galbraith data, scripts to reproduce tables and
575 figures (on GitHub).

576 **Competing interests**

577 The authors declare that they have no competing interests.

578 **Funding**

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

581 **Author's contributions**

582 Each author must write their authorsip contributions.

583 **Acknowledgements**

584 Who/what should we mention here? ...

585 **Author details**

586 ¹Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA 50011, USA. ²Econometrics
587 and Business Statistics, Monash University, Clayton, VIC 3800, Australia. ³Department of Entomology, Iowa State
588 University, Ames, IA 50011, USA. ⁴Department of Ecology, Evolution, and Organismal Biology, Iowa State
589 University, Ames, IA 50011, USA. ⁵Department of Entomology, University of Illinois at Urbana-Champaign, Urbana,
590 IL 61801, USA.

591 **References**

- 592 1. van Engelsdorp, D., Evans, J.D., Saegerman, C., Mullin, C., Haubruge, E., Nguyen, B.K., Frazier, M., Frazier,
593 J., Cox-Foster, D., Chen, Y., Underwood, R., Tarry, D.R., Pettis, J.S.: Colony collapse disorder: A descriptive
594 study. *PLoS ONE* **4**, 6481 (2009)
- 595 2. Kulhanek, K., Steinhauer, N., Rennich, K., Caron, D.M., Sagili, R.R., Pettis, J.S., Ellis, J.D., Wilson, M.E.,
596 Wilkes, J.T., Tarry, D.R., Rose, R., Lee, K., Rangel, J., vanEngelsdorp, D.: A national survey of managed
597 honey bee 2014–2015 annual colony losses in the USA. *Journal of Apicultural Research* **56**, 328–340 (2017)
- 598 3. Laurent, M., Hendrikx, P., Ribiere-Chabert, M., Chauzat, M.-P.: A pan-European epidemiological study on
599 honeybee colony losses 2012–2014. *Epilobee* **2013**, 44 (2016)
- 600 4. Caron, D., Sagili, R.: Honey bee colony mortality in the Pacific Northwest: Winter 2009/2010. *Am Bee J* **151**,
601 73–76 (2011)
- 602 5. Bond, J., Plattner, K., Hunt, K.: Fruit and Tree Nuts Outlook: Economic Insight U.S. Pollination- Services
603 Market. Economic Research Service Situation and Outlook FTS-357SA, USDA (2014)
- 604 6. Gallai, N., Salles, J.-M., Settele, J., Vaissière, B.B.: Economic valuation of the vulnerability of world agriculture
605 confronted with pollinator decline. *Ecol. Econ.* **68**, 810–821 (2009)
- 606 7. Klein, A.-M., Vaissière, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T.:
607 Importance of pollinators in changing landscapes for world crops. *Proc Biol Sci* **274**, 303–313 (2007)
- 608 8. Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., Kunin, W.E.: . Global pollinator
609 declines: trends, impacts and drivers **25**, 345–353 (2010)
- 610 9. Spivak, M., Mader, E., Vaughan, M., Euliss, N.H.: The Plight of the Bees. *Environ Sci Technol* **45**, 34–38
611 (2011)
- 612 10. Goulson, D., Nicholls, E., Botías, C., Rotheray, E.L.: Bee declines driven by combined stress from parasites,
613 pesticides, and lack of flowers. *Science* **347**, 1255957 (2015)
- 614 11. Roulston, T.H., Buchmann, S.L.: A phylogenetic reconsideration of the pollen starch-pollination correlation.
615 *Evol Ecol Res* **2**, 627–643 (2000)
- 616 12. Stanley, R.G., Linsens, H.F.: Pollen: Biology, Biochemistry, Management
- 617 13. Brodschneider, R., Crailsheim, K.: Nutrition and health in honey bees. *Apidologie* **41**, 278–294 (2010)
- 618 14. Haydak, M.H.: Honey bee nutrition. *Annu Rev Entomol* **15**, 143–156 (1970)
- 619 15. Crailsheim, K., Schneider, L.H.W., Hrassnigg, N., Bühlmann, G., Brosch, U., Gmeinbauer, R., Schöffmann, B.:
620 Pollen consumption and utilization in worker honeybees (*Apis mellifera carnica*): dependence on individual age
621 and function. *J Insect Physiol* **38**, 409–419 (1992)
- 622 16. Crailsheim, K.: The flow of jelly within a honeybee colony. *J Comp Physiol B* **162**, 681–689 (1992)
- 623 17. Schmidt, J.O.: Feeding preference of *Apis mellifera* L. (Hymenoptera: Apidae): Individual versus mixed pollen
624 species. *J. Kans. Entomol. Soc.* **57**, 323–327 (1984)
- 625 18. Schmidt, J.O., Thoenes, S.C., Levin, M.D.: Survival of honey bees, *Apis mellifera* (Hymenoptera: Apidae), fed
626 various pollen sources. *J. Econ. Entomol.* **80**, 176–183 (1987)

19. Alaux, C., Ducloz, F., Conte, D.C.Y.L.: Diet effects on honeybee immunocompetence. *Biol. Lett.* **6**, 562–565 (2010)
20. Naug, D.: Nutritional stress due to habitat loss may explain recent honeybee colony collapses. *Biol Conserv* **142**, 2369–2372 (2009)
21. Engelsdorp, D.V., Hayes, J.J., Underwood, R.M., Pettis, J.: A survey of honey bee colony losses in the U.S., fall 2007 to spring 2008. *PLoS ONE* **3**, 4071 (2008)
22. Neumann, P., Carreck, N.L.: Honey bee colony losses. *J Apicult Res* **49**, 1–6 (2010)
23. Engelsdorp, D.V., Meixner, M.D.: A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *J Invertebr Pathol* **103**, 80–95 (2010)
24. Rosenkranz, P., Aumeier, P., Ziegelmann, B.: Biology and control of *Varroa destructor*. *J Invertebr Pathol* **103**, 96–119 (2010)
25. Weinberg, K.P., Madel, G.: The influence of the mite *Varroa Jacobsoni* Oud. on the protein concentration and the haemolymph volume of the brood of worker bees and drones of the honey bee *Apis Mellifera* L. *Apidologie* **16**, 421–436 (1985)
26. Shen, M.Q., Cui, L.W., Ostiguy, N., Cox-Foster, D.: Intricate transmission routes and interactions between picorna-like viruses (Kashmir bee virus and sacbrood virus) with the honeybee host and the parasitic varroa mite. *J Gen Virol* **86**, 2281–2289 (2005)
27. Yang, X., Cox-Foster, D.: Effects of parasitization by *Varroa destructor* on survivorship and physiological traits of *Apis mellifera* in correlation with viral incidence and microbial challenge. *Parasitology* **134**, 405–412 (2007)
28. 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)
29. Chen, Y.P., Siede, R.: Honey bee viruses. *Adv Virus Res* **70**, 33–80 (2007)
30. Miranda, J.R.D., Cordon, G., Budge, G.: The acute bee paralysis virus-Kashmir bee virus-Israeli acute paralysis virus complex. *J Invertebr Pathol* **103**, 30–47 (2010)
31. 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)
32. 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)
33. 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)
34. Cox-Foster, D.L., Conlan, S., Holmes, E.C., Palacios, G., Evans, J.D., Moran, N.A., Quan, P.-L., Briese, T., Hornig, M., Geiser, D.M., Martinson, V., vanEngelsdorp, D., Kalkstein, A.L., Drysdale, A., Hui, J., Zhai, J., Cui, L., Hutchison, S.K., Simons, J.F., Egholm, M., Pettis, J.S., Lipkin, W.I.: A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**, 283–287 (2007)
35. Hou, C., Rivkin, H., Slabezki, Y., Chejanovsky, N.: Dynamics of the presence of Israeli acute paralysis virus in honey bee colonies with colony collapse disorder. *Viruses* **6**, 2012–2027 (2014)
36. Cornman, R.S., Tapy, D.R., Chen, Y., Jeffreys, L., Lopez, D., Pettis, J.S.: Pathogen webs in collapsing honey bee colonies. *PLoS ONE* **7**, 43562 (2012)
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. Dolezal, A.G., Carrillo-Tripp, J., Judd, T., Miller, A., Bonning, B., Toth, A.: Interacting stressors matter: Diet quality and virus infection in honey bee health. In prep (2018)
41. 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)
42. Dolezal, A.G., Toth, A.L.: Feedbacks between nutrition and disease in honey bee health. *Current Opinion in Insect Science* **26**, 114–119 (2018)
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. Carval, D., Ferriere, R.: A unified model for the coevolution of resistance, tolerance, and virulence. *Evolution* **64**, 2988–3009 (2010)
46. 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)
47. 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)
48. 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)
49. 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)
50. Chambers, M.C., Schneider, D.S.: Balancing resistance and infection tolerance through metabolic means. *PNAS* **109**, 13886–13887 (2012)
51. 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)
52. Cerenius, L., Söderhäll, K.: The prophenoloxidase-activating system in invertebrates. *Immunological Reviews* **198**, 116–126 (2004)
53. Sadd, B.M., Siva-Jothy, M.R.: Self-harm caused by an insect's innate immunity. *Proceedings of the Royal Society B: Biological Sciences* **273**, 2571–2574 (2006)
54. 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)
55. 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)
56. Consortium, H.B.G.S.: Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* **15**, 86 (2014)
57. 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)
58. 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)
59. 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)
60. 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)
61. 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)
62. 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)
63. 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>
64. Page, R.E., Laidlaw, H.H.: Full sisters and supersisters: A terminological paradigm. *Anim. Behav.* **36**, 944–945 (1988)
65. Sherman, P.W., Seeley, T.D., Reeve, H.K.: Parasites, pathogens, and polyandry in social Hymenoptera. *Am. Nat* **131**, 602–610 (1988)
66. 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)
67. Mattila, H.R., Seeley, T.D.: Genetic diversity in honey bee colonies enhances productivity and fitness. *Science* **317**, 362–364 (2007)
68. Tarpy, D.R.: Genetic diversity within honeybee colonies prevents severe infections and promotes colony growth. *Proc. R. Soc. Lond. B* **270**, 99–103 (2003)
69. Brodschneider, R., Arnold, G., Hrasnigg, N., Crailsheim, K.: Does patriline composition change over a honey bee queen's lifetime? *Insects* **3**, 857–869 (2012)
70. Jolliffe, I.T.: *Principal Component Analysis*. Springer, ??? (2002)
71. Inselberg, A.: The plane with parallel coordinates. *The Visual Computer* **1**, 69–91 (1985)
72. Cleveland, W.S.: *Visualizing Data*. Summit, New Jersey: Hobart Press, ??? (1993)
73. 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)
74. Chandrasekhar, T., Thangavel, K., Elayaraja, E.: Effective Clustering Algorithms for Gene Expression Data. *International Journal of Computer Applications* **32**, 4 (2011)
75. de Souto D. de Araujo, M., Costa, I., Soares, R., Ludermir, T., Schliep, A.: Comparative Study on Normalization Procedures for Cluster Analysis of Gene Expression Datasets. *International Joint Conference on Neural Networks*, 2793–2799 (2008)
76. 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)
77. 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)
78. Supek, F., Bošnjak, M., Škunca, N., Šmuc, T.: REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS ONE* **6**, 21800 (2011)

753 **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 Tukey honest significant differences with a confidence level of 95%.

Figure 2 IAPV titer volumes for the four treatment groups, two virus groups, and two diet groups. Left to right: IAPV titer volumes 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 Tukey honest significant differences with a confidence level of 95%.

Figure 3 Parallel coordinate plots of the 1,019 DEGs after hierarchical clustering of size four between the virus-infected and control groups 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 DEGs after hierarchical clustering of size four between the virus-infected and control groups of our study. 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 analysis results for the 122 DEGs related to our “tolerance” hypothesis and for the 125 DEGs related to our “resistance” hypothesis. GO analysis results for the 122 DEGs related to our “tolerance” hypothesis (A) and for the 125 DEGs related to our “resistance” hypothesis (B).

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.

754 **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, physiological 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).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

858 Additional file 14 — Scatterplot matrix of virus-related DEGs from our dataset, showing only replicates 10, 11, and
 859 12.
 860 The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only
 861 replicates 10, 11, and 12 are shown from both treatment groups. The data has been standardized. “N” represents
 862 non-infected control samples and “V” represents virus-treated samples. We see that, compared to the scatterplot
 863 matrices from certain clusters of the Galbraith data, the 43 DEGs from this subset of six samples from our data do
 864 not paint as clear of a picture, and most of them unexpectedly deviate from the $x=y$ line in the virus-related
 865 replicate plots (PNG).

866 Additional file 15 — Parallel coordinate plots of the “tolerance” candidate DEGs.
 867 Parallel coordinate plots of the 122 DEGs after hierarchical clustering of size four between the “tolerance” candidate
 868 DEGs. Here “N” represents non-infected control group, “V” represents treatment of virus, “C” represents
 869 high-quality Chestnut diet, and “R” represents low-quality Rockrose diet. The vertical red line indicates the
 870 distinction between treatment groups. We see there is considerable noise in the data (non-consistent replicate
 871 values), but that the general patterns of the DEGs follow what we expect based on our “tolerance” contrast (PNG).

872 Additional file 16 — Parallel coordinate plots of the “resistance” candidate DEGs.
 873 Parallel coordinate plots of the 125 DEGs after hierarchical clustering of size four between the “resistance”
 874 candidate DEGs. Here “N” represents non-infected control group, “V” represents treatment of virus, “C” represents
 875 high-quality Chestnut diet, and “R” represents low-quality Rockrose diet. The vertical red line indicates the
 876 distinction between treatment groups. We see there is considerable noise in the data (non-consistent replicate
 877 values), but that the general patterns of the DEGs follow what we expect based on our “resistance” contrasts
 878 (PNG).

879 Additional file 17 — Venn diagrams comparing the virus-related DEG overlaps in the Galbraith data using our
 880 pipeline and the pipeline used by Galbraith *et al.*
 881 Venn diagrams comparing the virus-related DEG overlaps of the Galbraith data from the DESeq2 bioinformatics
 882 pipelines used in the Galbraith study (labeled as “G.O.”) and the DESeq2 bioinformatics pipelines used in our study
 883 (labeled as “G.R.”). While we were not able to fully replicate the DEG list published in the Galbraith study, our DEG
 884 list maintained significant overlaps with their DEG list. From left to right: Total virus-related DEGs (subplot A),
 885 virus-upregulated DEGs (subplot B), control-upregulated DEGs (subplot C) (PNG).

886 Additional file 18 — Venn diagrams of main effect DEG overlaps across DESeq2, edgeR, and limma
 887 Venn diagrams comparing DEG overlaps across DESeq2, edgeR, and limma for our diet main effect (top row), our
 888 virus main effect (middle row), and the Galbraith virus main effect (bottom row). Within a given subplot, “D”
 889 represents DESeq2, “E” represents edgeR, and “L” represents limma. From left to right on top row: Total
 890 diet-related DEGs (subplot A), Castanea-upregulated DEGs (subplot B), Rockrose-upregulated DEGs (subplot C).
 891 From left to right on middle row: Total virus-related DEGs (subplot D), virus-upregulated DEGs (subplot E),
 892 control-upregulated DEGs in our data (subplot F). From left to right on bottom row: Total virus-related DEGs
 893 (subplot G), virus-upregulated DEGs (subplot H), control-upregulated DEGs in the Galbraith data (subplot I)
 894 (PNG). With the exception of the limma pipeline resulting in zero DEGs in our virus main effect analysis, we found
 895 significant overlaps between DEG lists across the different pipelines (DESeq2, edgeR, and limma). In general,
 896 DESeq2 resulted in the largest number of DEGs and limma resulted in the least number of DEGs (PNG).

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