RESEARCH

Transcriptomic responses to diet quality and viral infection in Apis mellifera

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

*Correspondence:
adolezal@illinois.edu
[7]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

Background

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

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

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

28 compared to undeveloped land areas [22], and beekeepers rank poor nutrition as

29 one of the main reasons for colony losses [23]. Understanding how low diversity di-

30 ets (i.e. monofloral diets) affect honey bee health will be crucial to resolve problems

31 that may arise as agriculture continues to intensify throughout the world [24, 25].

32 Indeed, differing qualities of monofloral diets have been shown to affect nurse bee

33 physiology and tolerance to parasites [26].

34 Viral infection was considered a comparatively minor problem in honey bees un-

35 til the last century when the ectoparasitic varroa mite (*Varroa destructor*) spread

36 worldwide [27, 28, 29]. This mite feeds on honey bee hemolymph and/or fat body

37 tissue [30, 31], and is believed to decrease lipid and glycogen reserves and reduce

38 protein synthesis in bees [32]. Additionally, it transmits multiple viruses and sup-

39 ports replication of some viruses [33, 34, 35, 36]. More than 20 honey bee viruses

40 have been identified [37]. One of these viruses that has been linked to honey bee

41 decline is Israeli acute paralysis virus (IAPV), a positive-sense RNA virus of the

42 family *Dicistroviridae* [38]. IAPV infection causes shivering wings, decreased loco-

43 motion, muscle spasms, paralysis, and high premature death percentages in caged

44 infected adult honey bees [39]. IAPV has demonstrated higher infectious capacity

45 [40] and is more prevalent in colonies that do not survive the winter [41].

46 Although there is growing interest in how viruses and diet quality affect the health

47 and sustainability of honey bees, as well as a recognition that such factors might

48 operate interactively, there are only a small number of experimental studies thus

49 far directed toward elucidating the interactive effects of these two factors in honey

50 bees [42, 43, 44, 45, 46]. We recently used laboratory cages and nucleus hive experi-

51 ments to investigate the health effects of these two factors, and our results show the

52 importance of the combined effects of both diet quality and virus infection. Specifi-

53 cally, ingestion by honey bees of high quality pollen is able to mitigate virus-induced

54 mortality to the level of diverse, polyfloral pollen [11].

Following up on these findings, 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 that reduces the pathogen load of the host (resistance, [47]) or if it is due to indirect effects of good nutrition on the ability of the host to reduce pathogen impacts without affecting pathogen load (tolerance, [47]). 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, which can help us generate candidate gene lists to further investigate the relative roles of tolerance and resistance. 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 [48] and virus infection effects [49, 50, 51, 52, 53] 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 [54, 55, 56, 48]. 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 [48, 51]. Previous transcriptomic studies have identified genes serving

links between metabolism and antiviral defense in honey bees [57, 58]; see [59]
for an overview. Numerous studies on the transcriptomic effects of virus infection
in model insect organisms have shown that RNA silencing, transcriptional paus-
ing, Toll pathways, IMD pathways, JAK/STAT pathways, and Toll-7 autophagy
pathways play substantial roles in virus-host systems [60, 61]. Studies of virus-bee
systems have revealed some of the antiviral defense pathways known in model or-
ganisms are conserved and also related to bee antiviral immune responses [62].

To our knowledge, there are few to no studies investigating honey bee gene expres-
sion patterns specifically related to monofloral diets, and few 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 [45]. 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
inoculation, high quality diet without IAPV inoculation, low quality diet with IAPV
inoculation, and high quality diet with IAPV inoculation). For our diet factor, we
examined two monofloral pollen diets, rockrose (*Cistus* sp.) and chestnut (*Castanea*
sp.). Rockrose pollen is generally considered less nutritious than chestnut pollen be-
cause it contains smaller amounts of protein, amino acids, antioxidants, calcium, and
iron [11, 26]. For specific quantitative differences between these two pollen groups,
please see [26]. Throughout this paper, we refer to our four treatment groups as
"NR" (non-inoculation and low quality pollen), "NC" (non-inoculation and high
quality pollen), "VR" (IAPV inoculation and low quality pollen), and "VC" (IAPV
inoculation and high quality pollen). 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.

Because RNA-seq data can be noisy and subject to high levels of inter-experiment variation, we further sought to validate our transcriptomic data via comparison to a previous RNA-seq study on honey bee responses to viral infection. To do this, we compare the main effect of IAPV exposure in our dataset to that obtained in a previous study conducted by Galbraith and colleagues [49]. While our study examines honey bees derived from naturally-mated queens, the Galbraith study examined honey bees derived from single-drone inseminated queens. 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 [63]. We note that the difference between these studies may be even greater than this as we 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 [64], increased tolerance to environmental changes [65], and increased colony performance [66, 67]. Given that honey bees are naturally very polyandrous [68], our naturally-mated honey bees may also reflect more realistic environmental and genetic conditions. 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 [69, 70, 71], 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 (including scatterplot matrices,

parallel coordinate plots, and litre plots from the bigPint software package [72]) can be useful for cross-study comparisons and validation of noisy RNA-sequencing data in the future.

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, 40]. In brief, our virus inoculum was prepared by injection of infectious virus particles (derived from infected adults) into white-eyed honey bee pupae; these pupae were then homogenized and virus particles enriched and resuspended. This inoculum was then characterized for presence of acute bee paralysis virus, black queen cell virus, deformed wing virus (DWV), IAPV, Kashmir bee virus, and sacbrood bee virus (SBV). Experimental infection tests of adult bees and honey bee cell cultures [40] showed that only IAPV is amplified in adult bees. To infect caged bees for these experiments, newly emerged bees from 15 healthy colonies at the Iowa State University research apiary were homogeneously mixed, then counted into clear acrylic cages in groups of 35 bees per cage. Cages were then presented with open feeders containing 30% sucrose solution (control) or 30% sucrose solution containing a 1:1000 dilution of viral inoculum (treatment). Dietary treatments were then added (described below). To quantify virus titers, two live bees were randomly sampled at 36 hpi from each of 9-10 randomly selected cages. Virus levels were then measured via RT-qPCR and quantified against a standard curve, identically to methods described in [40, 11].

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 fit to the data by restricted maximum likelihood (REML) using the "lme" function in the R package "nlme". A random (intercept) effect for experimental

163   setup was included in the model. Post-hoc pairwise comparisons of the four (diet

164   and virus combination) treatment groups were performed and Benjamini-Hochberg

165   adjusted p-values were calculated to limit familywise Type I error rates [73].

166   Design of two-factor experiment

167   For our nutrition factor, we examined two monofloral pollen diets, rockrose (*Cis-*

168   *tus* sp.) and chestnut (*Castanea* sp.). Rockrose pollen is generally considered less

169   nutritious than chestnut pollen due to its lower levels of protein, amino acids, an-

170   tioxidants, calcium, and iron [11, 26]. For our virus factor, one level contained bees

171   that were inoculated with IAPV and another level contained bees that were not

172   inoculated with IAPV. This experimental design resulted in four treatment groups

173   (low quality rockrose pollen without IAPV exposure, high quality chestnut pollen

174   without IAPV exposure, low quality rockrose pollen with IAPV exposure, and high

175   quality chestnut pollen with IAPV exposure) that allowed us to assess main effects

176   and interactive effects between diet quality and IAPV infection in honey bees.

177   There are several reasons why our design focused only on diet quality (monofloral

178   diets) as opposed to diet diversity (monofloral diets versus polyfloral diets). First,

179   when assessing diet diversity, a sugar diet is often used as a control. However,

180   such an experimental design does not reflect real-world conditions for honey bees

181   as they rarely face a total lack of pollen [26]. Moreover, younger larvae tend to

182   be fed pollen diets, whereas older larvae tend to be fed nectar diets. By focusing

183   on pollen diets, our study design reflects natural diet conditions for larvae of a

184   specific age category [74]. Second, in studies that compared honey bee health using

185   monofloral and polyfloral diets at the same time, if the polyfloral diet and one of

186   the high-quality monofloral diets both exhibited similarly beneficial effects, then it

187   was difficult for the authors to assess if the polyfloral diet was better than most of

188   the monofloral diets because of its diversity or because it contained as a subset the

189   high-quality monofloral diet [26]. Third, as was previously mentioned, honey bees

190 are now confronted with less diverse sources of pollen. As a result, there is a need

191 to better understand how monofloral diets affect honey bee health.

RNA extraction

193 Fifteen cages per treatment were originally produced for monitoring of mortality.

194 From these, six live honey bees were randomly selected from each cage 36 hours

195 post inoculation and placed into tubes [40]. In summary, 8 samples (representing two

196 bees each) were sequenced per experimental condition (i.e., 32 samples sequenced).

197 Tubes were kept on dry ice and then transferred into a -80C freezer until processing.

198 From the fifteen possible cages, eight were randomly selected for RNA-sequencing.

199 From these eight cages, two of the honey bees per cage were randomly selected

200 from the original six live honey bees per cage. These two bees were combined to

201 form a pooled sample representing the cage. Whole body RNA from each pool

202 was extracted using Qiagen RNeasy MiniKit followed by Qiagen DNase treatment.

203 Samples were suspended in water to 200-400 ng/μl. All samples were then tested

204 on a Bioanalyzer at the Iowa State University DNA Facility to ensure quality (RIN

205 > 8).

Gene expression

207 Samples were sequenced starting on January 14, 2016 at the Iowa State University

208 DNA Facility (Platform: Illumina HiSeq Sequencing 2500 in rapid run mode; Cat-

209 egory: Single End 100 cycle sequencing). A standard Illumina mRNA library was

210 prepared by the DNA facility. Reads were aligned to the BeeBase Version 3.2 genome

211 [75] from the Hymenoptera Genome Database [76] using the programs GMAP and

212 GSNAP [77]. There were four lanes of sequencing with 24 samples per lane. Each

213 sample was run twice. Approximately 75-90% of reads were mapped to the honey

214 bee genome. Each lane produced around 13 million single-end 100 basepair reads.

215 We tested all six pairwise combinations of treatments for DEGs (pairwise DEGs:

216 NR versus NC, NR versus VR, NR versus VC, NC versus VR, NC versus VC,

217 and VR versus VC). We also tested the diet main effect (diet DEGs), virus main

218 effect (virus DEGs), and interaction term for DEGs (interaction DEGs). We then

219 also tested for virus main effect DEGs (virus DEGs) in public data derived from a

220 previous study exploring the gene expression of IAPV virus infection in honey bees

221 [49]. We tested each DEG analysis using recommended parameters with DESeq2

222 [78], edgeR [79], and LimmaVoom [80]. In all cases, we used a false discovery rate

223 (FDR) threshold of 0.05 [81]. Fisher's exact test was used to determine significant

224 overlaps between DEG sets (whether from the same dataset but across different

225 analysis pipelines or from different datasets across the same analysis pipelines).

226 The eulerr shiny application was used to construct Venn diagram overlap images

227 [82]. In the end, we focused on the DEG results from DESeq2 [78] as this pipeline

228 was also used in the Galbraith study [49]. We used the independent filtering process

229 built into the DESeq2 software that mitigates multiple comparison corrections on

230 genes with no power rather than defining one filtering threshold.

231 Comparison to prior studies on transcriptomic response to viral infection

232 We compare the main effect of IAPV exposure in our dataset to that obtained in a

233 previous study conducted by Galbraith and colleagues [49] who also addressed honey

234 bee transcriptomic responses to virus infection. We applied the same downstream

235 bioinformatics analyses between our count table and the count table provided in

236 the Galbraith study. When we applied our bioinformatics pipeline to the Galbraith

237 count table, we obtained different differential expression counts compared to the

238 results published in the Galbraith study. However, there was substantial overlap and

239 we considered this justification to use the differential expression list we obtained in

240 order to keep the downstream bioinformatics analyses as similar as possible between

241 the two datasets (Additional file 17).

242 We used honey bees from naturally-mated colonies, whereas Galbraith et al. [49]

243 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 [83] from the DESeq2 package, a well-known and established tool. Along with that, we used lesser-known multivariate visualization tools from our R package called bigPint [72]. Specifically, we used parallel coordinate plots [84], scatterplot matrices [85], and litre plots (which we recently developed based on "replicate line plots" [86]) 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 [86].

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 for its read counts across its samples [87, 88]. 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 [89, 90]. The overrepresented gene ontology (GO) terms of DEGs were determined

271 using the BEEBASE_ID identifier option (honey bee gene model) in the DAVID

272 software. To fine-tune the GO term list, only terms correlating to Biological Pro-

273 cesses were considered. The refined GO term list was then imported into REVIGO

274 [91], which uses semantic similarity measures to cluster long lists of GO terms.

275 Probing tolerance versus resistance

276 To investigate whether the protective effect of good diet is due to direct, specific

277 effects on immune function (resistance), or if it is due to indirect effects of good

278 nutrition on energy availability and vigor (tolerance), we created contrasts of in-

279 terest (Table 2). In particular, we assigned "resistance candidate DEGs" to be the

280 ones that were upregulated in the chestnut group within the virus inoculated bees

281 but not upregulated in the chestnut group within the non-inoculated bees. Our

282 interpretation of these genes is that they represent those that are only activated

283 in inoculated bees that are fed a high quality diet. We also assigned "tolerance

284 candidate DEGs" to be the ones that were upregulated in the chestnut group for

285 both the virus inoculated bees and non-inoculated bees. Our interpretation of these

286 genes is that they represent those that are constitutively activated in bees fed a

287 high quality diet, regardless of whether they are experiencing infection or not. We

288 then determined how many genes fell into these two categories and analyzed their

289 GO terminologies.

290 Post hoc analysis

291 We found considerable noisiness in our data and saw, through gene-level visual-

292 izations, that our DEGs contained outliers and inconsistent replicates. Hence, we

293 wanted to explore whether our DEG read counts correlated with pathogen response

294 metrics, including IAPV titers, SBV (also present in our inoculum [40, 11]), and

295 mortality rates. We explored correlation with SBV because our inoculum [40] does

296 contain SBV, and bees from both inoculated and non-inoculated groups do exhibit

297 detectable SBV titers. For this process, we considered virus main effect DEGs (Fig-

298 ure 4), "tolerance candidate" DEGs (Additional file 15), and "resistance candidate"

299 DEGs (Additional file 16). For each DEG in each cluster, we calculated a coefficient

300 of determination (R-squared) value to estimate the correlation between its raw read

301 counts and the pathogen response metrics across its 24 samples. We then used the

302 Kruskal–Wallis test to determine if the distribution of the R-squared values in any

303 of the DEG clusters significantly differed from those in the non-DEG genes (the rest

304 of the data). As there were four clusters for each of the nine combinations of DEG

305 lists ("tolerance" candidate DEGs, "resistance" candidate DEGs, and virus-related

306 DEGs) and pathogen response measurements (IAPV titer, SBV titer, and mortality

307 rate), this process resulted in 36 statistical tests.

308 Results

309 Mortality and virus titers

310 We reanalyzed our previously published dataset with a subset that focuses on diet

311 quality and is more relevant to the current study. We show the data subset here to

312 inform the RNA-sequencing comparison because we reduced the number of treat-

313 ments from the original published data (from eight to four) [11] as a means to focus

314 on diet quality effects.

315 As shown in Figure 1, mortality rates of honey bees 72 hours post-inoculation

316 significantly differed among the treatment groups (mixed model ANOVA across all

317 treatment groups, df = 3, 54; F = 10.03; p < 2.34e-05). The effect of virus treatment

318 (mixed model ANOVA, df = 1, 54; F = 24.73; p < 7.04e-06) and diet treatment

319 (mixed model ANOVA, df = 1, 54; F = 5.32; p < 2.49e-02) were significant, but

320 the interaction between the two factors (mixed model ANOVA, df = 1, 54; F =

321 4.72e-02, p = 8.29e-01) was not significant. We compared mortality levels based

322 on pairwise comparisons: For a given diet, honey bees exposed to the virus showed

323 significantly higher mortality rate than honey bees not exposed to the virus. Bees fed

324 rockrose pollen had significantly elevated mortality with virus infection compared to

325 non-inoculated controls (Benjamini-Hochberg, p < 1.53e-03), and bees fed chestnut

326 pollen similarly had significantly elevated mortality with virus infection compared

327 to controls (Benjamini-Hochberg, p < 3.12e-03) (Figure 1).

328 As shown in Figure 2, IAPV titers of honey bees 72 hours post-inoculation sig-

329 nificantly differed among the treatment groups (mixed model ANOVA across all

330 treatment groups, df = 3, 33; F = 6.10; p < 2.03e-03). The effect of virus treat-

331 ment (mixed model ANOVA, df = 1, 33; F = 15.04; p < 4.75e-04) was significant,

332 but the diet treatment (mixed model ANOVA, df = 1, 33; F = 2.55; p = 1.20e-

333 01) and the interaction between the two factors (mixed model ANOVA, df = 1,

334 33; F = 7.02e-01, p = 4.08e-01) were not significant. We compared IAPV titers

335 based on pairwise comparisons: Bees fed rockrose pollen had significantly elevated

336 IAPV titers with virus infection compared to non-inoculated controls (Benjamini

337 Hochberg, p < 7.56e-03). However, bees fed chestnut pollen did not have signifi-

338 cantly elevated IAPV titers with virus infection compared to non-inoculated controls

339 (Benjamini Hochberg, p = 6.29e-02). While many of the non-inoculated treatment

340 groups showed some RT-qPCR amplification (non-inoculated average Ct=33.92;

341 inoculated average Ct=24.9), and thus have virus titers calculable on a standard

342 curve, these Ct levels are similar to those deemed uninfected in previous studies [49].

343 Overall, we interpreted these findings to mean that high-quality chestnut pollen

344 could partially reduce high virus titers resulting from the inoculation treatment,

345 whereas low-quality rockrose pollen could not (Figure 2).

346 Transcriptomic responses to virus infection and diet

347 We observed a substantially larger number of differentially expressed genes (DEGs)

348 in our diet main effect ($n = 1,914$) than in our virus main effect ($n = 43$) (Supple-

349 mentary table 1 A and B, Additional file 1). There were only four genes that were

350 DEGs in both our diet main effect and our virus main effect (GB48747, GB47214,

351 GB42908, and GB42507). In the diet factor, more DEGs were upregulated in the

352 more-nutritious chestnut group ($n = 1,033$) than in the less-nutritious rockrose

353 group ($n = 881$). In the virus factor, there were more virus-upregulated DEGs ($n =$

354 38) than control-upregulated DEGs ($n = 5$). While these reported DEG counts are

355 from the DESeq2 package, we saw similar trends for the edgeR and limma package

356 results (Supplementary table 1, Additional file 1 and Additional file 18). For our

357 DEG analysis, we used R software version 3.3.3 [92].

358   We performed GO analysis to statistically assign our DEGs to predefined bins

359 based on their functional characteristics, allowing us to better understand the bi-

360 ological processes of our DEGs. GO analysis of the chestnut-upregulated DEGs

361 revealed the following over-represented biological functions: Wnt signaling, hippo

362 signaling, and dorso-ventral axis formation, as well as pathways related to circa-

363 dian rhythm, mRNA surveillance, insulin resistance, inositol phosphate metabolism,

364 FoxO signaling, ECM-receptor interaction, phototransduction, Notch signaling,

365 JaK-STAT signaling, MAPK signaling, and carbon metabolism (Supplementary

366 table 2, Additional file 1). These encompassed almost all of the overrepresented

367 biological functions in chestnut-upregulated DEGs conditioned on non-inoculation

368 (i.e. upregulated in the "NC" group compared to the "NR" group; Supplementary

369 table 4, Additional file 1) and inoculation (i.e. upregulated in the "VC" group com-

370 pared to the "VR" group; Supplementary table 6, Additional file 1). GO analysis of

371 the rockrose DEGs revealed pathways related to terpenoid backbone biosynthesis,

372 homologous recombination, SNARE interactions in vesicular transport, aminoacyl-

373 tRNA biosynthesis, Fanconi anemia, and pyrimidine metabolism (Supplementary

374 table 3, Additional file 1). We note that Fanconi anemia pathways was also the

375 only GO term discovered in rockrose DEGs conditioned on viral inoculation (i.e.

376 upregulated in the "VR" group compared to the "VC" group) (Supplementary ta-

377 ble 7, Additional file 1). However, Fanconi anemia pathways were not found in

378 rockrose DEGs conditioned on non-inoculation (i.e. upregulated in the "NR" group

379 compared to the "NC" group) (Supplementary table 5, Additional file 1).

380   With so few DEGs ($n = 43$) in our virus main effect comparison, we focused on in-

381 dividual genes and their known functionalities rather than GO over-representation

382 (Table 1). Of the 43 virus-related DEGs, only 10 had GO assignments within the

383 DAVID database. These genes had putative roles in the recognition of pathogen-

384 related lipid products and the cleaving of transcripts from viruses, as well as in-

385 volvement in ubiquitin and proteosome pathways, transcription pathways, apoptotic

386 pathways, oxidoreductase processes, and several more functions (Table 1).

387   No interaction DEGs were observed between the diet and virus factors of the

388 study, in any of the pipelines (DESeq2, edgeR, and limma).

389   The number of DEGs across the six treatment pairings between the diet and

390 virus factor ranged from 0 to 955 (Supplementary table 8, Additional file 1). Again,

391 diet level appeared to have greater influence on the number of DEGs than the virus

392 level. Across every pair comparing the chestnut and rockrose levels, regardless of the

393 virus level, the number of chestnut-upregulated DEGs was higher than the number

394 of rockrose-upregulated DEGs (Supplementary table 8 C, D, E, F, Additional file 1).

395 Virus-treated bees showed equal to or more upregulated genes relative to controls,

396 under both diet treatments (Supplementary table 8 A and B, Additional file 1).

397 These trends were observed for all three pipelines used (DESeq2, edgeR, and limma).

398   Supplementary tables 1-9 in Additional file 20 contain complete DEG lists for all

399 comparisons performed in this study.

400 Transcriptomic data visualization and comparison to a previous study

401 We wished to explore the signal:to:noise ratio between the Galbraith dataset and

402 our dataset. Note that the Galbraith dataset contained three individual bees per

403 treatment group as a single pooled sample, while our dataset contained 16 indi-

404 vidual bees per treatment group in 8 RNA-seq samples. Basic PCA plots were

constructed with the DESeq2 analysis pipeline and showed that the Galbraith dataset may separate the inoculated and non-inoculated honey bees better than our dataset (Additional file 2). Wanting to learn more about the data at the gene level, we continued with new visualization techniques that are available online [72]. For more information about the visualizations used here, please refer to (https://lindsayrutter.github.io/bigPint/articles/plotIntro.html).

We used parallel coordinate lines superimposed onto side-by-side boxplots to visualize the DEGs associated with virus infection in the two studies. The background side-by-side boxplot represents the distribution of *all* genes in the data (all 15,314 genes in our count table), and each parallel coordinate line represents one DEG. In a parallel coordinate line, connections between samples with positive correlations should be flat, while connections between samples with negative correlations should be crossed. We expect DEGs to show more variability between treatments than between replicates. This means the parallel coordinate lines should be flat between replicates but crossed between treatments. However, overplotting problems would obscure our visualization if we were to plot all DEGs onto the same side-by-side boxplot. Therefore, we graphed clustered subsets of the DEGs (based on hierarchical clustering).

The 1,019 DEGs from the Galbraith dataset form relatively clean-looking visual displays, with consistent replicates and differences between treatments. The few inconsistent replicates we observed (such as V.1 from Cluster 1 and V.2 from Cluster 4) were small enough that consistent differences between the treatment groups remained apparent across the samples (Figure 3). In contrast, we see that the 43 virus-related DEGs from our dataset do not look as clean in their visual displays (Figure 4). The replicates appear somewhat inconsistent in their estimated expression levels and there is not always such a large (or even consistent) difference

431 between treatment groups. We see a similar finding when we also examine a larger

432 subset of 1,914 diet-related DEGs from our study (Additional file 3).

433 We next used repLIcate TREatment ("litre") plots, which we recently developed

434 for our bigPint software package [72]. Litre plots allow users to visualize one DEG

435 onto the Cartesian coordinates of one scatterplot matrix. In the litre plot, each gene

436 in the data is plotted once for every combination of replicates between treatment

437 groups. We use hexagon bins to summarize this massive information. Once the

438 background of hexagons has been drawn to reveal the distribution of all between-

439 treatment sample pair combinations for *all* genes, the user can superimpose all

440 between-treatment sample pair combinations for one gene of interest.

441 Additional file 4 shows nine example litre plots for our dataset; each litre plot

442 shows the 144 between-treatment sample pair combinations for one DEG of inter-

443 est. Additional file 5 and 6 similarly each show nine example litre plots for the

444 Galbraith dataset; each litre plot shows the nine between-treatment sample pair

445 combinations for one DEG of interest. We see that indeed the virus DEGs from

446 our data (Additional file 4) show less consistent replications and less differences be-

447 tween the treatment groups compared to the virus DEGs from the Galbraith data

448 (Additional files 5 and 6). We also observe that, in the Galbraith dataset, the DEG

449 points in the first cluster show less tight cluster patterns than the DEG points in

450 the second cluster (Additional files 5 and 6), an observation we saw previously in

451 the parallel coordinate plots (Figure 3).

452 Finally, we used scatterplot matrices from the bigPint software to further assess

453 the DEGs [72]. A scatterplot matrix is another effective multivariate visualization

454 tool that plots read count distributions across all genes and samples. Specifically, it

455 represents every gene in the dataset as a black point in each scatterplot. DEGs can

456 be superimposed as colored points to assess their patterns against the full dataset.

457 We expect DEGs to mostly fall along the $x=y$ line in replicate scatterplots (denot-

458   ing replicate consistency) but deviate from the $x=y$ line in treatment scatterplots

459   (denoting significant treatment changes). The $x=y$ line is shown in red in our plots.

460   We created standardized scatterplot matrices for each of the four clusters (from

461   Figure 3) of the Galbraith data (Additional files 7, 8, 9, and 10). We also created

462   standardized scatterplot matrices for our data. However, as our dataset contained

463   24 samples, we would need to include 276 scatterplots in our matrix, which would

464   be too numerous to allow for efficient visual assessment of the data. As a result,

465   we created four scatterplot matrices of our data, each with subsets of 6 samples

466   to be more comparable to the Galbraith data (Additional files 11, 12, 13, and 14).

467   Specifically, we arbitrarily subsetted the samples so each one was represented once

468   in each of these four files (i.e. Additional File 11 shows samples 1-3; Additional

469   File 12 shows samples 4-6; Additional File 13 shows samples 7-9; and Additional

470   File 14 shows samples 10-12). We can again confirm through these plots that the

471   DEGs from the Galbraith data appeared more as expected: They deviated more

472   from the $x=y$ line in the treatment scatterplots while staying close to the $x=y$ line

473   in replicate scatterplots.

474   Despite the virus-related DEGs ($n = 1,019$) from the Galbraith dataset displaying

475   the expected patterns more than those from our dataset ($n = 43$), there was signif-

476   icant overlap (p-value $< 2.2e-16$) in the DEGs between the two studies, with 26/38

477   (68%) of virus-upregulated DEGs from our study also showing virus-upregulated

478   response in the Galbraith study (Figure 6).

479   Tolerance versus resistance

480   Using the contrasts specified in Table 2, we discovered 122 "tolerance" candi-

481   date DEGs and 125 "resistance" candidate DEGs. Within our 122 "tolerance"

482   gene ontologies, we found functions related to metabolism (such as carbohydrate

483   metabolism, fructose metabolism, and chitin metabolism). However, we also discov-

484   ered gene ontologies related to RNA polymerase II transcription, immune response,

485 and regulation of response to reactive oxygen species (Figure 5A). Within our 125

486 "resistance" gene ontologies, we found functions related to metabolism (such as car-

487 bohydrate metabolism, chitin metabolism, oligosaccharide biosynthesis, and general

488 metabolism) (Figure 5B).

489   To visually explore gene expression patterns related to tolerance and resistance,

490 we used hierarchical clustering to separate candidate DEGs into common patterns,

491 and then visualized these clusters using parallel coordinate lines superimposed onto

492 side-by-side boxplots. To reduce overplotting of parallel coordinate lines, we again

493 used hierarchical clustering techniques to separate DEGs into common patterns.

494 Perhaps unsurprisingly, we still see a substantial amount of noise (inconsistency

495 between replicates) in our resulting candidate DEGs (Additional files 15 and 16).

496 However, the broad patterns we expect to see still emerge: For example, based on

497 the contrasts we created to obtain the 'tolerance' candidate DEGs, we expect them

498 to display larger count values in the "NC" group compared to the "NR" group and

499 larger count values in the "VC" group compared to the "VR" group. Indeed, we see

500 this pattern in the associated parallel coordinate plots (Additional file 15). Likewise,

501 based on the contrasts we created to obtain the 'resistance' candidate DEGs, we

502 still expect them to display larger count values in the "VC" group compared to

503 the "VR" group, but we no longer expect to see a difference between the "NC"

504 and "NR" groups. We do generally see these expected patterns in the associated

505 parallel coordinate plots: While there are large outliers in the "NC" group, the "NR"

506 replicates are no longer typically below a standardized count of zero (Additional file

507 16). The genes in Cluster 3 may follow the expected pattern the most distinctively

508 (Additional file 16).

509   Post hoc analysis

510   To better understand sources of transcriptomic noise, we explored whether pathogen

511   response measurements (virus titers and mortality), which varied widely across

512   samples, were correlated with observed patterns in gene expression.

513   The R-squared values between gene read counts and pathogen response measure-

514   ments were generally low (R-squared < 0.1) across our dataset (Supplementary

515   table 9, Additional file 1). We further explored whether clusters of DEGs showed

516   higher correlations with pathogen response measurements than non-DEGs (the lat-

517   ter serving as a control, where we do not expect a correlation). A Kruskal–Wallis

518   test was used to determine if R-squared distributions of DEG clusters significantly

519   differed from those in the rest of the data. The p-values and Bonferroni correction

520   values for each of the 36 tests (as described in the methods section) is provided

521   in Supplementary table 9, Additional file 1. An overall trend emerges to suggest

522   that DEGs may have significantly larger correlation with the pathogen response

523   measurements compared to non-DEGs.

Discussion

525   Challenges to honey bee health are a growing concern, in particular the combined,

526   interactive effects of nutritional stress and pathogens [12]. In this study, we used

527   RNA-sequencing to probe mechanisms underlying honey bee responses to two ef-

528   fects, diet quality and infection with the prominent virus of concern, IAPV. In

529   general, we found a major nutritional transcriptomic response, with nearly 2,000

530   transcripts changing in response to diet quality (rockrose/poor diet versus chest-

531   nut/good diet). The majority of these genes were upregulated in response to high

532   quality diet, and these genes were over-represented for functions such as nutrient

533   signaling metabolism (insulin resistance), immune response (Notch signaling and

534   JaK-STAT pathways), and carbon metabolism (Supplementary table 2, Additional

535 file 1). These data suggest high quality nutrition may allow bees to alter their

536 metabolism, favoring investment of energy into immune responses.

537 One of the few studies that has investigated transcriptomic response to nutrition in

538 honey bees similarly found that pollen upregulates genes related to macromolecule

539 metabolism, insulin pathways, and TOR pathways [48]. Diet effects on transcrip-

540 tomics have been more extensively studied in the insect model Drosophila. One

541 recent transcriptomic study in *Drosophila melanogaster* reported an overexpression

542 of genes related to immunity, metabolism, and hemocyanin in a high-fat diet and

543 overexpression of genes related to cell cycle activity, DNA binding and transcription,

544 and CHK kinase-like protein activity in a high-sugar diet [54]. This same study also

545 discovered an upregulation of genes related to peptide and carbohydrate processing

546 in both high-fat and high-sugar diets, a finding the authors attributed to a general

547 increase in caloric intake. Another recent study investigated the transcriptomic ef-

548 fects of diets high in protein relative to sugar, diets high in sugar relative to protein,

549 and diets with equal amounts of protein and sugar [55]. *Drosophila mojavensis* and

550 *Drosophila arizonae* showed substantial differential expression between the dietary

551 conditions: genes involved in carbohydrate and lipid metabolism were upregulated

552 in response to high sugar low protein diets and genes involved in juvenile hormone

553 (JH) and ecdysone were upregulated in response to low sugar high protein diets. In-

554 terestingly, prior studies have suggested that JH regulates body size by controlling

555 ecdysone production, which modifies insulin signaling [56]. As we saw in our study,

556 these studies generally suggest that diet differences may relate to gene expression

557 changes in metabolism and immune responses in honey bees.

558 While some insect systems have shown relatively low transcriptional responses

559 to dicistrovirus infection [93, 94], previous work on honey bees has revealed many

560 hundreds of DEGs [49]. Discrepancies between datasets may be due to noise and

561 complexity of the honey bee microbiome. The transcriptomic response to virus infec-

tion in our experiment was fairly limited. We found only 43 differentially expressed transcripts, some with known immune functions such as an MD-2 lipid recognition protein that is particularly implicated in the recognition of pathogen products and argonaute-2, a protein that plays a central role in RNA silencing (Table 1). We also found genes related to transcriptional regulation, including Histone H4, Rho guanine nucleotide exchange factor 11, and transcriptional regular Myc-B, which is a regular gene that codes for a transcription factor. We additionally found Tropomyosin-like, a gene involved in muscle contraction. The small number of DEGs in this study may be partly explained by the large amount of noise in the data (Figure 4 and Additional files 2B, 4, 11, 12, 13, and 14) and baseline viral titers observed in our control bees (Figure 2).

There have been numerous studies on the transcriptomic effects of virus infection in model organisms like fruit flies and mosquitoes that can provide a useful framework for interpreting virus responses in honey bees. These studies have showed that RNA silencing is a major antiviral strategy, along with transcriptional pausing, Toll pathways, IMD pathways, JAK/STAT pathways, and Toll-7-autophagy pathways [60, 61]. Recent transcriptomic studies in honey bees have shown similar hallmarks of these same antiviral defense mechanisms, including RNA silencing, Toll pathways, IMD pathways, JAK/STAT pathways, autophagy, and endocytosis [62]. It is important to note that general immune responses to viral infection in insects might be an indirect result of cellular damage [61]. In fact, every virus-host interaction has its own particularities derived from the diverse methods of replication and infection cycle evolved by different viruses. An intricate set of pro- and anti-virus host factors such as ribosomal proteins and autophagy pathways are involved, but the response depends on the virus species, as has been elucidated in Drosophila [60, 61]. In addition, a non-sequence-specific antiviral response mediated by unspecific dsRNA pathway was discovered in honey bees [95, 50]. 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 [51]. 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) [51]. 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 [49]. 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 [62]; 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 [49], 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-value $< 2.2e-16$) 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.

Data visualization is a useful method to identify noise and robustness in RNA-sequencing data [79]. In this study, we used extensive data visualization to improve the interpretation of our RNA-sequencing results. For example, the DESeq2 package comes with certain visualization options that are popular in RNA-sequencing analysis. One of these visualization is the principal component analysis (PCA) plot, 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

642 applications can be useful for future researchers analyzing RNA-sequencing data to

643 quickly and effectively ensure that the DEG calls look reliable or at least overlap

644 with DEG calls from similar studies that look reliable. We also expect this type of

645 visualization exploration can be especially crucial when studying wild populations

646 with high levels of genetic and environmental variation between replicates and/or

647 when using experiments that may lack rigid design control.

648   One of the goals of this study was to use our RNA-sequencing data to assess

649 whether transcriptomic responses to diet quality and virus infection provide insight

650 into whether high quality diet can buffer bees from pathogen stress via mechanisms

651 of "resistance" or "tolerance". Recent evidence has suggested that overall immu-

652 nity is determined by more than just "resistance" (the reduction of pathogen fitness

653 within the host by mechanisms of avoidance and control) [96]. Instead, overall im-

654 munity is related to "resistance" in conjunction with "tolerance" (the reduction

655 of adverse effects and disease resulting from pathogens by mechanisms of heal-

656 ing) [97, 96]. Immune-mediated resistance and diet-driven tolerance mechanisms

657 are costly and may compete with each other [97, 98]. Data and models have sug-

658 gested that selection can favor an optimum combination of both resistance and

659 tolerance [99, 100, 101, 102]. We attempted to address this topic through specific

660 gene expression contrasts (Table 2), accompanied by GO analysis of the associ-

661 ated gene lists. We found an approximately equal number of resistance ($n = 125$)

662 and tolerance ($n = 122$) related candidate DEGs, suggesting both processes may

663 be playing significant roles in dietary buffering from pathogen induced mortality.

664 Resistance candidate DEGs had functions related to several forms of metabolism

665 (chitin and carbohydrate), regulation of transcription, and cell adhesion (Figure

666 5B). Tolerance candidate DEGs had functions related to carbohydrate metabolism

667 and chitin metabolism; however, they also showed functions related to immune re-

668 sponse, including RNA polymerase II transcription (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 [103]. These possible immunological defense mechanisms within our "tolerance" candidate DEGs and metabolic processes within our "resistance" candidate DEGs may provide additional evidence of feedbacks between diet and disease in honey bees [12]. Thus, our study uses transcriptome data to generate lists of candidate genes that can be the focus of future investigations to better experimentally test putative roles of tolerance and resistance genes in this system.

There were several limitations in this study that could be improved upon in future studies. For instance, our comparison between the Galbraith data (single-drone colonies) and our data (naturally-mated colonies) was limited by numerous extraneous variables between these studies. In addition to different molecular pipelines and bioinformatic preprocessing pipelines used between these studies, the Galbraith study focused on worker honey bees that were fed sugar and artificial pollen diets, whereas our study focused on worker honey bees that were fed bee-collected monofloral diets. Also, Galbraith's bees were sampled at 24 hours while ours were sampled at 36 hours. Furthermore, the Galbraith data used eviscerated abdomens with attached fat bodies and observations to determine behaviorally symptomatic bees whereas we used whole bodies and categorized only into inoculated vs. non-inoculated groups. There are also differences in the hours post inoculation and possible differences in the inoculation amount between the studies. Further differences between the studies can be found in their corresponding published methods sections [11, 49]. The different factors between these two studies may be critical because particular antiviral factors in honey bees are linked to specific viruses, specific developmental stages, the analyzed tissue, the route of inoculation, and the time (post-inoculation) during which the study was performed. This was clearly

demonstrated when comparing honey bee responses to two related iflaviruses with very different infection dynamics, SBV vs. DWV [52]. Authors observed differences in induction of defensin and hymenoptaecin immune-related genes, and suggested the results reflect adaptations to the different routes of transmission [52].

Moreover, our comparative visualization assessment between these two datasets was also somewhat limited because the virus effect in the Galbraith study used three replicates for each level, whereas the virus effect in our study used twelve replicates for each level that were actually further subdivided into six replicates for each diet level. Hence the apparent reduction in noise observed in the Galbraith data compared to our data in the PCA plots, parallel coordinate plots, scatterplot matrices, and litre plots may be an inadvertent product of the smaller number of replicates used and the lack of a secondary treatment group rather than solely the reduction in genetic variability through the single-drone colony design itself. With this in mind, while our current efforts may be a starting point, future studies can shed more light on signal:to:noise and differential expression differences between naturally-mated 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. Previous work has shown that even though IAPV replication occurs in all honey bee tissues, it localizes more in gut and nerve tissues and in the hypopharyngeal glands. Likewise, the highest IAPV titers have been observed in gut tissues [41]. 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 [104]. These 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. 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 [49].

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 ex-

amining 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 sug-

gest complex transcriptomic responses to multiple stressors in honey bees. Diet has

the capacity for large and profound effects on gene expression and may set up the

potential for both resistance and tolerance to viral infection, adding to previous

evidence of possible feedbacks between diet and disease in honey bees [12].

Moreover, this study also demonstrated the benefits of using data visualizations

and multiple datasets to address inherently messy biological data. For instance, by

verifying the substantial overlap in our DEG lists to those obtained in another study

that addressed a similar question using specimens with less genetic variability, we

were able to place much higher confidence in the differential gene expression results

750 from our otherwise noisy data. We also suggested that comparing results derived

751 from multiple studies varying in level of genetic and environmental variability may

752 allow researchers to identify transcriptomic patterns that are concurrently more

753 realistic and less noisy. Altogether, we hope our results underline the merits of using

754 data visualization techniques and multiple datasets to understand and interpret

755 RNA-sequencing datasets.

Declarations

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

Consent for publication
Not applicable.

Availability of data and materials
The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus [105] and are accessible through GEO Series accession number GSE121885 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121885). Scripts to reproduce visualizations in this paper are available online (https://github.com/lindsayrutter/HoneyBeePaper). Information about bigPint visualizations, including how to intrepret them are available online (https://lindsayrutter.github.io/bigPint). Lists of BeeBase IDs for contrast DEGs from this study are found in Additional File 20.

Competing interests
The authors declare that they have no competing interests.

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

Author details
[1]Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA 50011, USA. [2]Department of Microbiology, Center for Scientific Research and Higher Education of Ensenada, Ensenada, Baja California 22860, Mexico. [3]Department of Entomology and Nematology, University of Florida, Gainesville, FL 32611, USA. [4]Econometrics and Business Statistics, Monash University, Clayton, VIC 3800, Australia. [5]Department of Entomology, Iowa State University, Ames, IA 50011, USA. [6]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA. [7]Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

References
1. van Engelsdorp, D., Evans, J.D., Saegerman, C., Mullin, C., Haubruge, E., Nguyen, B.K., Frazier, M., Frazier, J., Cox-Foster, D., Chen, Y., Underwood, R., Tarpy, D.R., Pettis, J.S.: Colony collapse disorder: A descriptive study. PLoS ONE **4**, 6481 (2009)
2. Kulhanek, K., Steinhauer, N., Rennich, K., Caron, D.M., Sagili, R.R., Pettis, J.S., Ellis, J.D., Wilson, M.E., Wilkes, J.T., Tarpy, D.R., Rose, R., Lee, K., Rangel, J., vanEngelsdorp, D.: A national survey of managed honey bee 2014–2015 annual colony losses in the USA. Journal of Apicultural Research **56**, 328–340 (2017)
3. Laurent, M., Hendrikx, P., Ribiere-Chabert, M., Chauzat, M.-P.: A pan-European epidemiological study on honeybee colony losses 2012–2014. Epilobee **2013**, 44 (2016)
4. Caron, D., Sagili, R.: Honey bee colony mortality in the Pacific Northwest: Winter 2009/2010. Am Bee J **151**, 73–76 (2011)
5. Bond, J., Plattner, K., Hunt, K.: Fruit and Tree Nuts Outlook: Economic Insight U.S. Pollination- Services Market. Economic Research Service Situation and Outlook FTS-357SA, USDA (2014)
6. Gallai, N., Salles, J.-M., Settele, J., Vaissière, B.B.: Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. Ecol. Econ. **68**, 810–821 (2009)

804   7. Klein, A.-M., Vaissière, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke,
805   T.: Importance of pollinators in changing landscapes for world crops. Proc Biol Sci **274**, 303–313 (2007)
806   8. Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., Kunin, W.E.: . Global pollinator
807   declines: trends, impacts and drivers **25**, 345–353 (2010)
808   9. Spivak, M., Mader, E., Vaughan, M., Euliss, N.H.: The Plight of the Bees. Environ Sci Technol **45**, 34–38
809   (2011)
810   10. Goulson, D., Nicholls, E., Botías, C., Rotheray, E.L.: Bee declines driven by combined stress from parasites,
811   pesticides, and lack of flowers. Science **347**, 1255957 (2015)
812   11. Dolezal, A.G., Carrillo-Tripp, J., Judd, T.M., Miller, W.A., Bonning, B.C., Toth, A.L.: Interacting stressors
813   matter: diet quality and virus infection in honeybee health. Royal society open science **6**, 181803 (2019)
814   12. Dolezal, A.G., Toth, A.L.: Feedbacks between nutrition and disease in honey bee health. Current Opinion in
815   Insect Science **26**, 114–119 (2018)
816   13. Roulston, T.H., Buchmann, S.L.: A phylogenetic reconsideration of the pollen starch-pollination correlation.
817   Evol Ecol Res **2**, 627–643 (2000)
818   14. Stanley, R.G., Linskens, H.F.: Pollen: Biology, Biochemistry, Management
819   15. Brodschneider, R., Crailsheim, K.: Nutrition and health in honey bees. Apidologie **41**, 278–294 (2010)
820   16. Haydak, M.H.: Honey bee nutrition. Annu Rev Entomol **15**, 143–156 (1970)
821   17. Crailsheim, K., Schneider, L.H.W., Hrassnigg, N., Bühlmann, G., Brosch, U., Gmeinbauer, R., Schöffmann, B.:
822   Pollen consumption and utilization in worker honeybees (Apis mellifera carnica): dependence on individual age
823   and function. J Insect Physiol **38**, 409–419 (1992)
824   18. Crailsheim, K.: The flow of jelly within a honeybee colony. J Comp Physiol B **162**, 681–689 (1992)
825   19. Schmidt, J.O.: Feeding preference of Apis mellifera L. (Hymenoptera: Apidae): Individual versus mixed pollen
826   species. J. Kans. Entomol. Soc. **57**, 323–327 (1984)
827   20. Schmidt, J.O., Thoenes, S.C., Levin, M.D.: Survival of honey bees, Apis mellifera (Hymenoptera: Apidae), fed
828   various pollen sources. J. Econ. Entomol. **80**, 176–183 (1987)
829   21. Alaux, C., Ducloz, F., Conte, D.C.Y.L.: Diet effects on honeybee immunocompetence. Biol. Lett. **6**, 562–565
830   (2010)
831   22. Naug, D.: Nutritional stress due to habitat loss may explain recent honeybee colony collapses. Biol Conserv
832   **142**, 2369–2372 (2009)
833   23. Engelsdorp, D.V., Hayes, J.J., Underwood, R.M., Pettis, J.: A survey of honey bee colony losses in the U.S.,
834   fall 2007 to spring 2008. PLoS ONE **3**, 4071 (2008)
835   24. Neumann, P., Carreck, N.L.: Honey bee colony losses. J Apicult Res **49**, 1–6 (2010)
836   25. Engelsdorp, D.V., Meixner, M.D.: A historical review of managed honey bee populations in Europe and the
837   United States and the factors that may affect them. J Invertebr Pathol **103**, 80–95 (2010)
838   26. Pasquale, G.D., Salignon, M., Conte, Y.L., Belzunces, L.P., Decourtye, A., Kretzschmar, A., Suchail, S.,
839   Brunet, J.-L., Alaux, C.: Influence of pollen nutrition on honey bee health: Do pollen quality and diversity
840   matter? PLoS ONE **8**, 72016 (2013)
841   27. Rosenkranz, P., Aumeier, P., Ziegelmann, B.: Biology and control of Varroa destructor. Journal of invertebrate
842   pathology. Journal of invertebrate pathology **103**, 96–119 (2010)
843   28. Allen, M.F., Ball, B.V., White, R.F., Antoniw, J.F.: The detection of acute paralysis virus in Varroa jacobsoni
844   by the use of a simple indirect ELISA. Apicult. Res. **25**, 100–105 (1986)
845   29. Yue, C., Genersch, E.: RT-PCR analysis of deformed wing virus in honeybees (Apis mellifera) and mites
846   (Varroa destructor). J. Gen. Virol. **86**, 3419–3424 (2005)
847   30. Weinberg, K.P., Madel, G.: The influence of the mite Varroa Jacobsoni Oud. on the protein concentration and
848   the haemolymph volume of the brood of worker bees and drones of the honey bee Apis Mellifera L. Apidologie
849   **16**, 421–436 (1985)
850   31. Ramsey, S.D., Ochoa, R., Bauchan, G., Gulbronson, C., Mowery, J.D., Cohen, A., Lim, D., Joklik, J., Cicero,
851   J.M., Ellis, J.D., Hawthorne, D., vanEngelsdorp, D.: Varroa destructor feeds primarily on honey bee fat body
852   tissue and not hemolymph. PNAS (2019)
853   32. Aronstein, K.A., Saldivar, E., Vega, R., Westmiller, S., Douglas, A.E.: How Varroa Parasitism Affects the
854   Immunological and Nutritional Status of the Honey Bee, Apis mellifera. Insects **3**, 601–615 (2012)
855   33. Shen, M.Q., Cui, L.W., Ostiguy, N., Cox-Foster, D.: Intricate transmission routes and interactions between
856   picorna-like viruses (Kashmir bee virus and sacbrood virus) with the honeybee host and the parasitic varroa
857   mite. J Gen Virol **86**, 2281–2289 (2005)
858   34. Yang, X., Cox-Foster, D.: Effects of parasitization by Varroa destructor on survivorship and physiological traits
859   of Apis mellifera in correlation with viral incidence and microbial challenge. Parasitology **134**, 405–412 (2007)
860   35. Yang, X.L., Cox-Foster, D.L.: Impact of an ectoparasite on the immunity and pathology of an invertebrate:
861   Evidence for host immunosuppression and viral amplification. P Natl Acad Sci USA **102**, 7470–7475 (2005)
862   36. Emsen, B., Hamiduzzaman, M.M., Goodwin, P.H., Guzman-Novoa, E.: Lower virus infections in Varroa
863   destructor-infested and uninfested brood and adult honey bees (Apis mellifera) of a low mite population
864   growth colony compared to a high mite population growth colony. PLoS ONE **10**, 0118885 (2015)
865   37. Chen, Y.P., Siede, R.: Honey bee viruses. Adv Virus Res **70**, 33–80 (2007)
866   38. Bonning, B.C., Miller, W.A.: Dicistroviruses. Annu Rev Entomol **55**, 129–150 (2010)
867   39. Maori, E., Paldi, N., Shafir, S., Kalev, H., Tsur, E., Glick, E., Sela, I.: IAPV, a bee-affecting virus associated
868   with Colony Collapse Disorder can be silenced by dsRNA ingestion. Insect Mol Biol **18**, 55–60 (2009)
869   40. Carrillo-Tripp, J., Dolezal, A.G., Goblirsch, M.J., Miller, W.A., Toth, A.L., Bonning, B.C.: In vivo and in vitro
870   infection dynamics of honey bee viruses. Sci Rep **6**, 22265 (2016)
871   41. Chen, Y.P., Pettis, J.S., Corona, M., Chen, W.P., Li, C.J., Spivak, M., Visscher, P.K., DeGrandi-Hoffman, G.,
872   Boncristiani, H., Zhao, Y., van Engelsdorp, D., Delaplane, K., Solter, L., Drummond, F., Kramer, M., Lipkin,
873   W.I., Palacios, G., Hamilton, M.C., Smith, B., Huang, S.K., Zheng, H.Q., Li, J.L., Zhang, X., Zhou, X.F.,
874   Wu, L.Y., Zhou, J.Z., Lee, M.-L., Teixeira, E.W., Li, Z.G., Evans, J.D.: Israeli acute paralysis virus:
875   Epidemiology, pathogenesis and implications for honey bee health. PLoS Pathog **10**, 1004261 (2014)

42. DeGrandi-Hoffman, G., Chen, Y.: Nutrition, immunity and viral infections in honey bees. Current Opinion in Insect Science **10**, 170–176 (2015)

43. 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)

44. Le Conte, Y., BRUNET, J.-L., McDonnell, C., Dussaubat, C., Alaux, C.: Interactions Between Risk Factors in Honey Bees

45. 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)

46. 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)

47. Medzhitov, R., Schneider, D.S., Soares, M.P.: Disease tolerance as a defense strategy. Science **335**, 936–941 (2012)

48. 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)

49. 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)

50. 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)

51. 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)

52. 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)

53. Vincent, D., Poeschl, Y., Gogol-Döring, A., Alaux, C., Annoscia, D., Aurori, C., Barribeau, S.M., *et al.*: Unity in defence: honeybee workers exhibit conserved molecular responses to diverse pathogens. BMC Genomics **18**, 207 (2017)

54. 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)

55. Nazario-Yepiz, N.O., Loustalot-Laclette, M.R., Carpinteyro-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)

56. 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)

57. O'Neal, S.T., Swale, D.R., Anderson, T.D.: ATP-sensitive inwardly rectifying potassium channel regulation of viral infections in honey bees. Sci. Rep. **7**, 8668 (2017)

58. O'Neal, S.T., Brewster, C.C., Bloomquist, J.R., Anderson, T.D.: Amitraz and its metabolite modulate honey bee cardiac function and tolerance to viral infection. J. Invertebr. Pathol. **149**, 119–126 (2017)

59. Grozinger, C.M., Flenniken, M.L.: Bee viruses: Ecology, pathogenicity, and impacts. Annual review of entomology **64**, 205–226 (2019)

60. Xu, J., Cherry, S.: Viruses and antiviral immunity in Drosophila. Dev Comp Immunol **42**, 67–84 (2014)

61. Swevers, L., Liu, J., Smagghe, G.: Defense Mechanisms against Viral Infection in Drosophila: RNAi and Non-RNAi. Viruses **10**, 230 (2018)

62. 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)

63. Page, R.E., Laidlaw, H.H.: Full sisters and supersisters: A terminological paradigm. Anim. Behav. **36**, 944–945 (1988)

64. Sherman, P.W., Seeley, T.D., Reeve, H.K.: Parasites, pathogens, and polyandry in social Hymenoptera. Am. Nat **131**, 602–610 (1988)

65. 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)

66. Mattila, H.R., Seeley, T.D.: Genetic diversity in honey bee colonies enhances productivity and fitness. Science **317**, 362–364 (2007)

67. Tarpy, D.R.: Genetic diversity within honeybee colonies prevents severe infections and promotes colony growth. Proc. R. Soc. Lond. B **270**, 99–103 (2003)

68. Brodschneider, R., Arnold, G., Hrassnigg, N., Crailsheim, K.: Does patriline composition change over a honey bee queen's lifetime? Insects **3**, 857–869 (2012)

69. Hansen, K.D., Brenner, S.E., Dudoit, S.: Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Research **38**, 131 (2010)

70. Oshlack, A., Robinson, M.D., Young, M.D.: From RNA-seq reads to differential expression results. Genome Biology **11**, 220 (2010)

71. 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)

72. Rutter, L., Cook, D.: bigPint: Big multivariate data plotted interactively. https://github.com/lindsayrutter/bigPint. Accessed 22 November 2018 (2018)

73. 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)

948   74. Traynor, K.S., Conte, Y.L., Page, R.E.: Age matters: pheromone profiles of larvae differentially influence
949   foraging behaviour in the honeybee, Apis mellifera. Animal behaviour **99**, 1–8 (2015)
950   75. Consortium, H.B.G.S.: Finding the missing honey bee genes: lessons learned from a genome upgrade. BMC
951   Genomics **15**, 86 (2014)
952   76. Elsik, C.G., Tayal, A., Diesh, C.M., Unni, D.R., Emery, M.L., Nguyen, H.N., Hagen, D.E.: Hymenoptera
953   Genome Database: integrating genome annotations in HymenopteraMine. Nucleic Acids Research **4**, 793–800
954   (2016)
955   77. Wu, T.D., Reeder, J., Lawrence, M., Becker, G., Brauer, M.J.: GMAP and GSNAP for genomic sequence
956   alignment: Enhancements to speed, accuracy, and functionality. Methods Mol Biol **1418**, 283–334 (2016)
957   78. Love, M.I., Huber, W., Anders, S.: Moderated estimation of fold change and dispersion for RNA-seq data with
958   DESeq2. Genome Biology **15**, 550 (2014)
959   79. Robinson, M.D., McCarthy, D.J., Smyth, G.K.: edger: a bioconductor package for differential expression
960   analysis of digital gene expression data. Bioinformatics **26**, 139–140 (2010)
961   80. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K.: limma powers differential
962   expression analyses for rna-sequencing and microarray studies. Nucleic Acids Research **43**(7), 47 (2015)
963   81. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: A practical and powerful approach to
964   multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) **57**, 289–300 (1995)
965   82. Larsson, J.: eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. (2018). R package version 4.0.0.
966   https://cran.r-project.org/package=eulerr
967   83. I.T. Jolliffe: Principal Component Analysis. Springer, Berlin, Heidelberg (2002)
968   84. Inselberg, A.: The plane with parallel coordinates. The Visual Computer **1**, 69–91 (1985)
969   85. W.S. Cleveland: Visualizing Data. Hobart Press, Summit, New Jersey (1993)
970   86. Cook, D., Hofmann, H., Lee, E., Yang, H., Nikolau, B., Wurtele, E.: Exploring gene expression data, using
971   plots. Journal of Data Science **5**, 151–182 (2007)
972   87. Chandrasekhar, T., Thangavel, K., Elayaraja, E.: Effective Clustering Algorithms for Gene Expression Data.
973   International Journal of Computer Applications **32**, 4 (2011)
974   88. de Souto D. de Araujo, M., Costa, I., Soares, R., Ludermir, T., Schliep, A.: Comparative Study on
975   Normalization Procedures for Cluster Analysis of Gene Expression Datasets. International Joint Conference on
976   Neural Networks, 2793–2799 (2008)
977   89. Huang, D.W., Sherman, B.T., Lempicki, R.: Systematic and integrative analysis of large gene lists using
978   DAVID bioinformatics resources. Nat Protoc **4**, 44–57 (2009)
979   90. Huang, D.W., Sherman, B.T., Lempicki, R.A.: Bioinformatics enrichment tools: paths toward the
980   comprehensive functional analysis of large gene lists. Nucleic Acids Res **37**, 1–13 (2009)
981   91. Supek, F., Bošnjak, M., Škunca, N., Šmuc, T.: REVIGO summarizes and visualizes long lists of Gene
982   Ontology terms. PLoS ONE **6**, 21800 (2011)
983   92. R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for
984   Statistical Computing, Vienna, Austria (2008). R Foundation for Statistical Computing. ISBN 3-900051-07-0.
985   http://www.R-project.org
986   93. Merkling, S.H., Overheul, G.J., van Mierlo, J.T., Arends, D., Gilissen, C., van Rij, R.P.: The heat shock
987   response restricts virus infection in Drosophila. Scientific Reports **5**, 12758 (2015)
988   94. Dostert, C., Jouanguy, E., Irving, P., Troxler, L., Galiana, D., Hetru, C., Hoffmann, J.A., Imler, J.-L.: The
989   JAK-STAT signaling pathway is required but not sufficient for the antiviral response of Drosophila. Nature
990   Immunology **6**, 946–953 (2005)
991   95. Flenniken, M.L., Andino, R.: Non-specific dsRNA-mediated antiviral response in the honey bee. PLoS ONE **8**,
992   77263 (2013)
993   96. Carval, D., Ferriere, R.: A unified model for the coevolution of resistance, tolerance, and virulence. Evolution
994   **64**, 2988–3009 (2010)
995   97. Miller, C.V.L., Cotter, S.C.: Resistance and tolerance: The role of nutrients on pathogen dynamics and
996   infection outcomes in an insect host. Journal of Animal Ecology **87**, 500–510 (2017)
997   98. Moret, Y.: Trans-generational immune priming: Specific enhancement of the antimicrobial immune response in
998   the mealworm beetle, Tenebrio molitor. Proceedings of the Royal Society B: Biological Sciences **273**,
999   1399–1405 (2006)
1000 99. Mauricio, R., Rausher, M.D., Burdick, D.S.: Variation in the defense strategies of plants: are resistance and
1001   tolerance mutually exclusive? Ecology **78**, 1301–1310 (1997)
1002 100. Fornoni, J., Nunez-Farfan, J., Valverde, P.L., Rausher, M.D.: Evolution of mixed plant defense allocation
1003   against natural enemies. Evolution **58**, 1685–1695 (2004)
1004 101. Restif, O., Koella, J.C.: Shared control of epidemiological traits in a coevolutionary model of host-parasite
1005   interactions. The American Naturalist **161**, 827–836 (2003)
1006 102. Chambers, M.C., Schneider, D.S.: Balancing resistance and infection tolerance through metabolic means.
1007   PNAS **109**, 13886–13887 (2012)
1008 103. Xu, J., Grant, G., Sabin, L.R., Gordesky-Gold, B., Yasunaga, A., Tudor, M., Cherry, S.: Transcriptional pausing
1009   controls a rapid antiviral innate immune response in Drosophila. Cell Host Microbe **12**, 531–543 (2012)
1010 104. Johnson, B.R., Atallah, J., Plachetzki, D.C.: The importance of tissue specificity for RNA-seq: highlighting the
1011   errors of composite structure extractions. BMC Genomics **14**, 586 (2013)
1012 105. Edgar, R., Domrachev, M., Lash, A.E.: Gene Expression Omnibus: NCBI gene expression and hybridization
1013   array data repository. Nucleic Acids Res **30**, 207–210 (2002)
1014 106. Schlicker, A., Domingues, F.S., Rahnenfuhrer, J., Lengauer, T.: A new measure for functional similarity of
1015   gene products based on Gene Ontology. BMC Bioinformatics **7**, 302 (2006)

1016 **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 [49].
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. We also note a deviation of the second replicate from the
virus-treated samples ("V.2") from the other virus-treated replicates in Cluster 4.

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-inoculated and
control groups of our study. "N" represents non-inoculated 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 [106].

**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-value < 2.2e-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.

1017 **Tables**

BeeBase ID, NCBI Gene ID	Gene Name	Known functions	Us	Galbraith
GB41545, 409187	MD-2-related lipid-recognition protein-like	Implicated in lipid recognition, particularly in the recognition of pathogen related products	N	-
GB50955, 411577	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, 727455	UBA-like domain-containing protein 2	Found in diverse proteins involved in ubiquitin/proteasome pathways	V	V
GB47407, 406132	Histone H4	Capable of affecting transcription, DNA repair, and DNA replication when post-transcriptionally modified	V	V
GB42313, 409923	Leishmanolysin-like peptidase	Encodes a protein involved in cell migration and invasion; implicated in mitotic progression in D. melanogaster	V	V
GB50813, 410127	Rho guanine nucleotide exchange factor 11	Implicated in regulation of apoptopic processes, cell growth, signal transduction, and transcription	V	V
GB54503, 411255	Thioredoxin domain-containing protein	Serves as a general protein disulphide oxidoreductase	N	-
GB53500, 100576392	Transcriptional regulator Myc-B	Regulator gene that codes for a transcription factor	V	V
GB51305, 551252	Tropomyosin-like	Related to protein involved in muscle contraction	N	N
GB50178, 726905	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 non-inoculated bees	Supplementary tables 4 and 5, Additional file 1
VC vs VR	376	Genes that change expression due to diet effect in inoculated 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 inoculated 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-inoculated 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-inoculated 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 inoculated 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 inoculated 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-inoculated 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).

1057 Additional file 5 — Example litre plots of DEGs from Cluster 1 of the Galbraith dataset.
1058 Example litre plots of the nine DEGs with the lowest FDR values from the 365 DEGs in Cluster 1 (originally shown
1059 in Figure 3) of the Galbraith dataset. "N" represents non-inoculated control samples and "V" represents
1060 virus-treated samples. Most of the light orange points (representing the nine combinations of samples between
1061 treatment groups for a given DEG) deviate from the $x{=}y$ line in a tight bundle as expected (PNG).

1062 Additional file 6 — Example litre plots of DEGs from Cluster 2 of the Galbraith dataset.
1063 Example litre plots of the nine DEGs with the lowest FDR values from the 327 DEGs in Cluster 2 (originally shown in
1064 Figure 3) of the Galbraith dataset. "N" represents non-inoculated control samples and "V" represents virus-treated
1065 samples. Most of the dark orange points (representing the nine combinations of samples between treatment groups
1066 for a given DEG) deviate from the $x{=}y$ line in a compact clump as expected. However, they are not as tightly
1067 bunched together compared to what we saw in the example litre plots of Cluster 1 (shown in Additional file 5). As a
1068 result, what we see in these litre plots reflects what we saw in the parallel coordinate lines of Figure 3: The replicate
1069 consistency in the Cluster 1 DEGs is not as clean as that in the Cluster 2 DEGs, but is still relatively clean (PNG).

1070 Additional file 7 — Scatterplot matrix of DEGs from Cluster 1 of the Galbraith dataset.
1071 The 365 DEGs from the first cluster of the Galbraith dataset (originally shown in Figure 3) superimposed as light
1072 orange dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. "N"
1073 represents non-inoculated control samples and "V" represents virus-treated samples. We confirm that the DEGs
1074 mostly follow the expected structure, with their placement deviating from the $x{=}y$ line in the treatment
1075 scatterplots, but adhering to the $x{=}y$ line in the replicate scatterplots. However, we do see that sample "V.1" may
1076 be somewhat inconsistent in these DEGs, as its presence in the replicate scatterplots shows DEGs deviating from
1077 the $x{=}y$ line more than expected and its presence in the treatment scatterplots shows DEGs adhering to the $x{=}y$
1078 line more than expected. This inconsistent sample was something we observed in Figure 3 (PNG).

1079 Additional file 8 — Scatterplot matrix of DEGs from Cluster 2 of the Galbraith dataset.
1080 The 327 DEGs from the second cluster of the Galbraith dataset (originally shown in Figure 3) superimposed as dark
1081 orange dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. "N"
1082 represents non-inoculated control samples and "V" represents virus-treated samples. We confirm that the DEGs
1083 mostly follow the expected structure, with their placement deviating from the $x{=}y$ line in the treatment
1084 scatterplots, but adhering to the $x{=}y$ line in the replicate scatterplots (PNG).

1085 Additional file 9 — Scatterplot matrix of DEGs from Cluster 3 of the Galbraith dataset.
1086 The 224 DEGs from the third cluster of the Galbraith dataset (originally shown in Figure 3) superimposed as
1087 turquoise dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. "N"
1088 represents non-inoculated control samples and "V" represents virus-treated samples. We confirm that the DEGs
1089 mostly follow the expected structure, with their placement deviating from the $x{=}y$ line in the treatment
1090 scatterplots, but adhering to the $x{=}y$ line in the replicate scatterplots (PNG).

1091 Additional file 10 — Scatterplot matrix of DEGs from Cluster 4 of the Galbraith dataset.
1092 The 103 DEGs from the fourth cluster of the Galbraith dataset (originally shown in Figure 3) superimposed as pink
1093 dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. "N" represents
1094 non-inoculated control samples and "V" represents virus-treated samples. We confirm that the DEGs mostly follow
1095 the expected structure, with their placement deviating from the $x{=}y$ line in the treatment scatterplots, but adhering
1096 to the $x{=}y$ line in the replicate scatterplots. We also see that the second replicate from the virus-treated sample
1097 ("V.2") may be somewhat inconsistent in these DEGs, as its presence in the replicate scatterplots results in the
1098 DEGs unexpectedly deviating from the $x{=}y$ line and its presence in the treatment scatterplots results in the DEGs
1099 unexpectedly adhering to the $x{=}y$ line. This inconsistent sample was something we observed in Figure 3 (PNG).

1100 Additional file 11 — Scatterplot matrix of virus-related DEGs from our dataset, showing only replicates 1, 2, and 3.
1101 The 43 virus-related DEGs from our dataset superimposed as magenta dots onto all genes in the form of a
1102 scatterplot matrix. Only replicates 1, 2, and 3 are shown from both treatment groups. The data has been
1103 standardized. "N" represents non-inoculated control samples and "V" represents virus-treated samples. We see that,
1104 compared to the scatterplot matrices from certain clusters of the Galbraith data, the 43 DEGs from this subset of
1105 six samples from our data do not paint as clear of a picture, sometimes unexpectedly deviating from the $x{=}y$ line in
1106 the replicate plots and sometimes unexpectedly adhering to the $x{=}y$ line in the treatment plots (PNG).

1107 Additional file 12 — Scatterplot matrix of virus-related DEGs from our dataset, showing only replicates 4, 5, and 6.
1108 The 43 virus-related DEGs from our dataset superimposed as magenta dots onto all genes in the form of a
1109 scatterplot matrix. Only replicates 4, 5, and 6 are shown from both treatment groups. The data has been
1110 standardized. "N" represents non-inoculated control samples and "V" represents virus-treated samples. We see that,
1111 compared to the scatterplot matrices from certain clusters of the Galbraith data, the 43 DEGs from this subset of
1112 six samples from our data do not paint as clear of a picture, and most of them unexpectedly adhere to the $x{=}y$ line
1113 in the treatment plots (PNG).

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

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

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

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

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

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

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

Additional file 20 — Tables listing DEGs for contrasts.
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 of 178 DEGs in our NR versus VR contrast. **Table 4:** IDs of 376 DEGs in our VC versus VR contrast. **Table 5:** IDs of 774 DEGs in our NC versus VR contrast. **Table 6:** IDs of 955 DEGs in our VC versus NR contrast. **Table 7:** IDs of 941 DEGs in our NC versus NR contrast. **Table 8:** IDs of 125 resistance candidate genes. **Table 9:** IDs of 122 tolerance candidate genes. (XLS).