

1 *Chapter 1*

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

5 **1.1 Introduction**

6 Commerically managed honeybees have undergone unusually large declines in the United
7 States and parts of Europe over the past decade (van Engelsdorp et al. 2009, Kulhanek et al.
8 2017, Laurent et al. 2016), with annual mortality rates exceeding what beekeepers consider
9 sustainable (Caron and Sagili 2011, Bond et al. 2014). More than 70 percent of major
10 global food crops (including fruits, vegatables, and nuts) at least benefit from pollination,
11 and yearly insect pollination services are valued wordwide at \$175 billion (Gallai et al.
12 2009). As honeybees are largely considered to be the leading pollinator of numerous crops,
13 their marked loss has considerable implications regarding agricultural sustainability (Klein
14 et al. 2007).

15 Honeybee declines have been associated with several factors, including pesticide use,
16 parasites, pathogens, habitat loss, and poor nutrition (Potts et al. 2010, Spivak et al. 2011).
17 Researchers generally agree that these stressors do not act in isolation; instead, they appear
18 to influence the large-scale loss of honeybees in interactive fashions as the environment
19 changes (Goulson et al. 2015). Nutrition and viral infection are two broad factors that pose
20 heightened dangers to honeybee health in response to recent environmental changes.

21 Pollen is the main source of nutrition (including proteins, amino acids, lipids, sterols,
22 starch, vitamins, and minerals) in honeybees (Roulston and Buchmann 2000, Stanley and
23 Linskens 1974). At the individual level, pollen supplies most of the nutrients necessary
24 for physiological development (Brodschneider and Crailsheim 2010) and is believed to
25 have considerable impact on longevity (Haydak 1970). At the colony level, pollen enables

26 young workers to produce jelly, which then nourishes larvae, drones, older workers, and the
27 queen (Crailsheim et al. 1992, Crailsheim 1992). Various environmental changes (including
28 urbanization and monoculture crop production) have significantly altered the nutritional
29 profile available to honeybees. In particular, honeybees are confronted with less diverse
30 selections of pollen, which is of concern because mixed-pollen (polyfloral) diets are generally
31 considered healthier than single-pollen (monofloral) diets (Schmidt 1984, Schmidt et al. 1987,
32 Alaux et al. 2010). Indeed, reported colony mortality rates are higher in developed land
33 areas compared to undeveloped land areas (Naug 2009), and beekeepers rank poor nutrition
34 as one of the main reasons for colony losses (Engelsdorp et al. 2008). Understanding how
35 undiversified diets affect honeybee health will be crucial to resolve problems that may arise
36 as agriculture continues to intensify throughout the world (Neumann and Carreck 2010,
37 Engelsdorp and Meixner 2010).

38 Viral infection was a comparatively minor problem in honeybees until the last century when
39 Varroa destructor (an ectoparasitic mite) spread worldwide (Rosenkranz et al. 2010). This
40 mite feeds on honeybee hemolymph (Weinberg and Madel 1985), transmits cocktails of
41 viruses, and supports replication of certain viruses (Shen et al. 2005, Yang and Cox-Foster
42 2007, Yang and Cox-Foster 2005). More than 20 honeybee viruses have been identified (Chen
43 and Siede 2007). One of these viruses that has been linked to honeybee decline is Israeli
44 Acute Paralysis Virus (IAPV). A positive-sense RNA virus of the Dicistroviridae family
45 (Miranda et al. 2010), IAPV causes infected honeybees to display shivering wings, decreased
46 locomotion, muscle spasms, and paralysis, and 80% of caged infected adult honeybees die
47 prematurely (Maori et al. 2009). IAPV has demonstrated higher infectious capacities
48 than other honeybee viruses in certain conditions (Carrillo-Tripp et al. 2016) and is more
49 prevalent in colonies that do not survive the winter (Chen et al. 2014). Its role in the rising
50 phenomenon of “Colony Collapse Disorder” (in which the majority of worker bees disappear
51 from a hive) remains unclear: It has been implicated in some studies (Cox-Foster et al.
52 2007, Hou et al. 2014) but not in other studies (van Engelsdorp et al. 2009, Cornman et al.
53 2012, Miranda et al. 2010). Nonetheless, it seems likely that IAPV reduces colony strength
54 and survival.

55 Although there is growing interest in how viruses and diet quality affect the health and
56 sustainability of honeybees, as well as a recognition that such factors might operate
57 interactively, there are only a small number of experimental studies thus far directed toward
58 elucidating the interactive effects of these two factors in honeybees (DeGrandi-Hoffman and
59 Chen 2015, DeGrandi-Hoffman et al. 2010, Conte et al. 2011). We recently used laboratory
60 cages and nucleus hive experiments to investigate the health effects of these two factors,
61 and our results show a significant interaction between diet quality and virus infection.
62 Specifically, high quality pollen is able to mitigate virus-induced mortality to the level of
63 diverse, polyfloral pollen (Dolezal et al. 2018).

Following up on these phenotypic findings from our previous study, we now aim to understand the corresponding underlying mechanisms by which high quality diets protect bees from virus-induced mortality. For example, it is not known whether the protective effect of good diet is due to direct, specific effects on immune function (resistance), or if it is due to indirect effects of good nutrition on energy availability and vigor (resilience). Transcriptomics is one means to achieve this goal. Transcriptomic analysis can help us identify 1) the genomic scale of transcriptomic response to diet and virus infection, 2) whether these factors interact in an additive or synergistic way on transcriptome function, and 3) the types of pathways affected by diet quality and viral infection. This information, heretofore lacking in the literature, can help us better understand how good nutrition may be able to serve as a "buffer" against other stressors (Dolezal and Toth 2018). As it stands, there are only a small number of published experiments examining gene expression patterns related to diet effects (Alaux et al. 2011) and IAPV infection effects (Galbraith et al. 2015) in honeybees. As far as we know, there are few to no studies investigating honeybee gene expression patterns specifically related to monofloral diets, and few to no studies investigating honeybee gene expression patterns related to the interaction effects of diet in any broad sense and viral inoculation in any broad sense.

In this study, we examine how monofloral diets and viral inoculation influence gene expression patterns in honeybees by focusing on four treatment groups (low quality diet without IAPV exposure, high quality diet without IAPV exposure, low quality diet with IAPV exposure, and high quality diet with IAPV exposure). We conduct RNA-sequencing analysis on a randomly selected subset of the honeybees 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 interactive effect of the two factors on gene expression patterns.

We also compare the main effect of IAPV exposure in our dataset to that obtained in a previous study conducted by Galbraith and colleagues (Galbraith et al. 2015). As RNA-sequencing data can be highly noisy, this comparison allowed us to characterize how repeatable and robust our RNA-seq results were in comparison to previous studies. Importantly, we use an in-depth data visualization approach to explore and validate our data, and suggest such an approach can be useful for cross-study comparisons of RNA-sequencing data in the future.

1.2 Methods

Details of the procedures we used to prepare virus inoculum, infect and feed caged honeybees, and quantify IAPV can be reviewed in our previous work (Dolezal et al. 2018). The statistical analysis we used to study the main and interaction effects of the two factors on mortality and IAPV titers is also described in our earlier report (Dolezal et al. 2018).

101 **1.2.1 Design of two-factor experiment**

102 There are several reasons why we focused only on diet quality (monofloral diets) as opposed
103 to diet diversity (monofloral diets versus polyfloral diets). First, when assessing diet
104 diversity, a sugar diet is often used as a control. However, such an experimental design
105 does not reflect real-world conditions for honeybees as they rarely face a total lack of pollen
106 ([Pasquale et al. 2013](#)). Second, in studies that compared honeybee health using monofloral
107 and polyfloral diets at the same time, if the polyfloral diet and one of the high-quality
108 monofloral diets both exhibited similarly beneficial effects, then it was difficult for the
109 authors to assess if the polyfloral diet was better than most of the monofloral diets because
110 of its diversity or because it contained as a subset the high-quality monofloral diet ([Pasquale](#)
111 [et al. 2013](#)). Third, colonies used for pollination in agricultural areas (monoculture) face
112 less diversified pollens (according to Brodschneider, 2010). Pollinating areas are currently
113 undergoing landscape alteration and agriculture intensification, and bees are increasingly
114 faced with less diversified diets (monoculture) ([Decourtey et al. 2010](#), [Brodschneider and](#)
115 [Crailsheim 2010](#)). As a result, there is a need to better understand how monofloral diets
116 affect honeybee health as a step toward mitigating the negative impact of human activity
117 on the honeybee population.

118 Consequently, for our nutrition factor, we examined two monofloral pollen diets, Cistus
119 (Rockrose) and Castanea (Chestnut). Cistus pollen is generally considered less nutritious
120 than Castanea pollen due to its lower levels of protein, amino acids, antioxidants, calcium,
121 and iron ([Pasquale et al. 2013](#), [Dolezal et al. 2018](#)). For our virus factor, one level contained
122 bees that were infected with IAPV and another level contained bees that were not infected
123 with IAPV. This experimental design resulted in four treatment groups (Cistus pollen
124 without IAPV exposure, Castanea pollen without IAPV exposure, Cistus pollen with IAPV
125 exposure, and Castanea pollen with IAPV exposure) that allowed us to assess main effects
126 and interactive effects between diet quality and IAPV infection in honeybees.

127 **1.2.2 RNA extraction**

128 Fifteen cages per treatment were originally sampled. Six live honeybees from each cage
129 were randomly selected 36 hours post inoculation and placed into tubes. Tubes were kept
130 on dry ice and then transferred into a -80C freezer until processing. Eight cages were
131 randomly selected from the original 15 cages, and 2 honeybees per cage were randomly
132 selected from the original six live honeybees per cage. Whole body RNA from each pool of
133 two honeybees were extracted using Qiagen RNeasy MiniKit followed by Qiagen DNase
134 treatment. Samples were suspended in water to 200-400 ng/ μ l. All samples were then
135 tested on a Bioanalyzer at the DNA core facility to ensure quality (RIN>8).

136 **1.2.3 Gene expression**

137 Samples were sequenced starting on January 14, 2016 at the Iowa State University DNA
138 Facility (Platform: Illumina HiSeq Sequencing; Category: Single End 100 cycle sequencing).
139 A standard Illumina mRNA library was prepared by the DNA facility. Reads were aligned
140 to the BeeBase Version 3.2 genome ([Consortium 2014](#)) from the Hymenoptera Genome
141 Database ([Elsik et al. 2016](#)) using the programs GMAP and GSNAp ([Wu et al. 2016](#)). We
142 tested all six pairwise combinations of treatments for DEGs (pairwise DEGs). We also
143 tested the diet main effect (diet DEGs), virus main effect (virus DEGs), and interaction
144 term for DEGs (interaction DEGs). We then also tested for virus main effect DEGs (virus
145 DEGs) in public data derived from a previous study exploring the gene expression of
146 IAPV virus infection in honeybees ([Galbraith et al. 2015](#)). We tested each DEG analysis
147 using recommended parameters with DESeq2 ([Love et al. 2014](#)), edgeR ([Robinson et al.
148 2010](#)), and LimmaVoom ([Ritchie et al. 2015](#)). In all cases, we used a false discovery rate
149 (FDR) threshold of 0.05 ([Benjamini and Hochberg 1995](#)). Fisher's exact test was used to
150 determine significant overlaps between DEG sets (whether from the same dataset but across
151 different analysis pipelines or from different datasets across the same analysis pipelines).
152 The `eulerr` shiny application was used to construct Venn diagram overlap images ([Larsson
153 2018](#)). In the main section of our paper and in subsequent analyses, we focus on the DEG
154 results from DESeq2 ([Love et al. 2014](#)) as this pipeline was also used in the Galbraith study
155 ([Galbraith et al. 2015](#)).

156 @@@ What percent of reads mapped? @@@ Total number of raw reads @@@ How many
157 lanes @@@ How many samples per lane

158 **1.2.4 Comparison to previous studies on transcriptomic response to viral
159 infection**

160 We also compare the main effect of IAPV exposure in our dataset to that obtained in
161 a previous study conducted by Galbraith and colleagues ([Galbraith et al. 2015](#)). While
162 our study examines honeybees from polyandrous colonies, the Galbraith study examined
163 honeybees from single-drone colonies. As a consequence, our honeybees will have an average
164 of about 75% genetic variance, and the honeybees from the Galbraith study will have an
165 average of about 25% genetic variance ([Page and Laidlaw 1988](#)). We should therefore expect
166 that the Galbraith study may generate data with lower signal:to:noise ratios than our data
167 due to the lower genetic variation between its replicates. At the same time, our honeybees
168 will be more likely to display the health benefits gained from increased genotypic variance
169 within colonies, including decreased parasitic load ([Sherman et al. 1988](#)), increased tolerance
170 to environmental changes ([Crozier and Page 1985](#)), and increasead colony performance
171 ([Mattila and Seeley 2007, Tarpy 2003](#)). Given that honeybees are naturally very polyandrous
172 ([Brodschneider et al. 2012](#)), our honeybees may also reflect more realistic environmental

173 and genetic simulations. Taken together, each study provides a different point of value: Our
174 study likely presents less artificial data while the Galbraith data likely presents less messy
175 data. We wish to explore how the gene expression effects of IAPV inoculation compare
176 between these two studies that used such different experimental designs. To achieve this
177 objective, we use visualization techniques to assess the signal:to:noise ratio between these
178 two datasets, and differential gene expression (DEG) analyses to determine any significantly
179 overlapping genes of interest between these two datasets. It is our hope that this aspect of
180 our study may shine light on how experimental designs that control genetic variability to
181 different extents might affect the resulting gene expression data in honeybees.

182 1.2.5 Visualization

183 We used @@@ visualization tools from @@@ and visual inference techniques to assess the
184 signal:to:noise ratio in the datasets and to assess the suitability of the DEG calls.

185 1.2.6 Gene Ontology

186 DEGs were uploaded as a background list to DAVID Bioinformatics Resources 6.7 ([Huang](#)
187 [et al. 2009a](#), [Huang et al. 2009b](#)). The overrepresented gene ontology (GO) terms of DEGs
188 were identified using the BEEBASE_ID identifier. To fine-tune the GO term list, only
189 significant terms ($FDR < 0.05$) and those correlating to Biological Processes were considered.
190 The refined GO term list was then imported into REVIGO ([Sipek et al. 2011](#)), which uses
191 semantic similarity measures to cluster long lists of GO terms. @@@ Pathways analysis
192 @@@

193 1.2.7 Detecting resilience versus resistance

194 To investigate whether the protective effect of good diet is due to direct, specific effects
195 on immune function (resistance), or if it is due to indirect effects of good nutrition on
196 energy availability and vigor (resilience), we created contrasts of interest (Table 1.10). In
197 particular, we assigned "resistance" genes to be the ones that were upregulated in the
198 Castanea group within the virus infected bees but not upregulated in the Castanea group
199 within the non-infected bees. We also assigned "resilience" genes to be the ones that were
200 upregulated in the Castanea group for both the virus infected bees and non-infected bees.
201 We then determined how many genes fell into these two categories and analyzed their GO
202 terminologies.

203 **1.3 Results**204 **1.3.1 Phenotypic results**

205 We reanalyzed our previously published dataset with a subset more relevant to our RNA-
206 sequencing approaches in the current study that have a more focused question regarding
207 diet quality. We briefly show it again here to inform the RNA-seq comparison because we
208 reduced the number of treatments (from eight to four) from the original published data
209 ([Dolezal et al. 2018](#)).

210 Mortality rates of honeybees 72 hour post-inoculation significantly differed among the
211 treatment groups (mixed model ANOVA across all treatment groups, $df=3, 55; F=10.07$;
212 $p<2.18e-05$). The effect of virus treatment (mixed model ANOVA, $df=1, 55; F=24.343$;
213 $p<7.84e-06$) and diet treatment (mixed model ANOVA, $df=1, 55; F=5.796; p<0.0194$)
214 were significant, but the interaction between the two factors (mixed model ANOVA, $df=1,$
215 $55; F=0.062, p=0.8039$) was not significant. The virus treatment was significant: For a
216 given diet, honeybees exposed to the virus showed significantly higher mortality rate than
217 honeybees not exposed to the virus (Tukey HSD, $p<0.05$). Without virus exposure, there
218 was only a numerical reduction in mortality rate for bees fed Castanea pollen compared to
219 Rockrose pollen (Tukey HSD, $p>0.05$). Similarly, with virus exposure, there was only a
220 numerical reduction in mortality rate for bees fed Castanea pollen compared to Rockrose
221 pollen (Tukey HSD, $p>0.05$). Overall, we discovered that the higher-quality Castanea diet
222 had the ability to numerically reduce mortality in the presence of IAPV infection compared
223 to the lower-quality Rockrose diet (Figure 1.1 A).

224 IAPV titer volumes of honeybees 72 hour post-inoculation significantly differed among the
225 treatment groups (mixed model ANOVA across all treatment groups, $df=3, 34; F=6.096$;
226 $p<0.00196$). The effect of virus treatment (mixed model ANOVA, $df=1, 34; F=15.686$;
227 $p<0.000362$) was significant, but the diet treatment (mixed model ANOVA, $df=1, 34;$
228 $F=1.898; p>0.05$) and the interaction between the two factors (mixed model ANOVA, $df=1,$
229 $34; F=0.702, p>0.05$) were not significant. Honeybees that were infected with the virus
230 and fed a poor-quality Rockrose diet showed significant increases in IAPV titer volumes
231 compared to honeybees that were not infected with the virus regardless of their diet quality
232 (Tukey HSD, $p<0.05$). Overall, we discovered that the higher-quality Castanea diet had
233 the ability to numerically reduce IAPV titer volume for both infected and non-infected
234 honeybees (Figure 1.1 B).

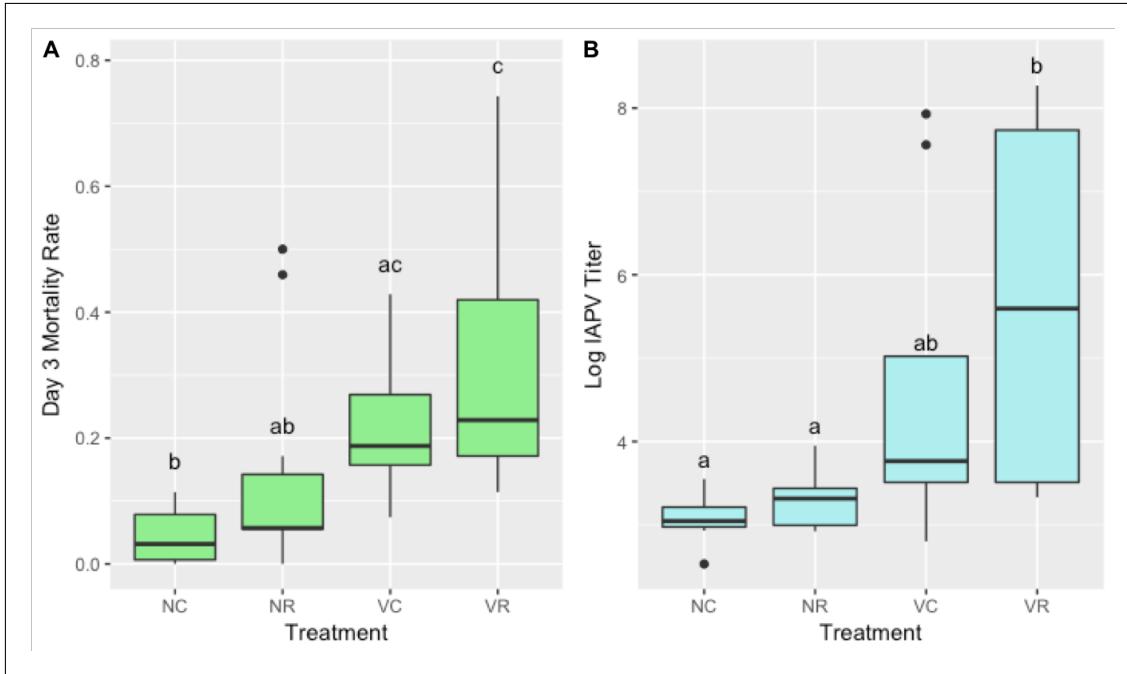


Figure 1.1: Mortality rates (A) and IAPV titers (B) for the four treatment groups.

235 1.3.2 Main effect DEG results

236 We observed a substantially larger number of DEGs in our diet main effect ($n = 1914$) than
237 in our virus main effect ($n = 43$) (Table 1.1A and B). In the diet factor, there were more
238 Castanea DEGs ($n = 1033$) than Rockrose DEGs ($n = 881$). In the virus factor, there
239 were more virus-exposed DEGs ($n = 38$) than control DEGs ($n = 5$). While these reported
240 DEGs numbers are from the DESeq2 package, we saw similar trends for the edgeR and
241 limma package results (Table 1.1A and B).

242 Pathway analysis of the Castanea DEGs revealed enriched (Benjamini correction < 0.05)
243 Wnt signaling, hippo signaling, and dorso-ventral axis formation, as well as pathways related
244 to circadian rhythm, mRNA surveillance, insulin resistance, inositol phosphate metabolism,
245 FoxO signaling, ECM-receptor interaction, phototransduction, Notch signaling, Jak-STAT
246 signaling, MAPK signaling, and carbon metabolism (Table 1.2). Pathway analysis of the
247 Rockrose DEGs revealed pathways related to terpenoid backbone biosynthesis, homologous
248 recombination, SNARE interactions in vesicular transport, aminoacyl-tRNA biosynthesis,
249 Fanconi anemia, and pyrimidine metabolism (Table 1.3).

250 It was difficult to perform a pathway analysis with so few DEGs ($n = 43$) in our virus main
251 effect study. As a result, we focused on individual genes and their known functionalities
252 (Table 1.5). Of the 43 virus-related DEGs, only 10 of them successfully mapped with
253 DAVID software. These genes had implications in the recognition of pathogen-related lipid
254 products and the cleaving of transcripts from viruses, as well as involvement in ubiquitin

255 and proteosome pathways, transcription pathways, apoptotic pathways, oxidoreductase
256 processes, and several more functions (Table ??).

257 1.3.3 Interaction DEG results

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

260 1.3.4 Pairwise comparison DEG results

261 The number of DEGs across the six treatment pairings between the diet and virus factor
262 ranged from 0 to 941 (Table 1.4). Some of the trends observed in the main effect comparisons
263 persisted: The diet level appeared to have greater influence on the number of DEGs than
264 the virus level. Across every pair comparing the Castanea and Rockrose levels, regardless
265 of the virus level, the number of Castanea-related DEGs was higher than the number of
266 Rockrose-related DEGs (Table 1.4 C, D, E, F). For the pairs in which the diet level was
267 controlled, the virus-exposed treatment showed equal to or more DEGs than the control
268 treatment (Table 1.4 A, B). There were no DEGs between the treatment pair controlling
269 for the control level of the virus effect (Table 1.4 A). These trends were observed for all
270 three pipelines used (DESeq2, edgeR, and limma).

271 1.3.5 Comparison with Galbraith study

272 We wished to explore the signal:to:noise ratio between the Galbraith dataset and our
273 dataset. Basic MDS plots are constructed during the DESeq2 analysis pipeline, and we
274 could immediately determine that the Galbraith dataset may better separate the infected
275 and uninfected honeybees better than our dataset (Figure 1.6). We also noted that the
276 first replicate of both treatment groups in the Galbraith data did not cluster as cleanly in
277 the MDS plots. However, through this automatically-generated plot, we can only visualize
278 information at the sample level. Wanting to learn more about the data at the gene level,
279 we continued with additional visualization techniques.

280 We used parallel coordinate lines superimposed onto boxplots to visualize the DEGs that
281 came out of the DESeq2 package. The background boxplot represents the distribution of
282 all genes in the data, and each parallel coordinate line represents one DEG. To reduce
283 overlapping of parallel coordinate lines, we often use hierarchical clustering techniques to
284 separate DEGs into common patterns. See more information about this plotting method
285 and the ideal visual structure of DEGs in our earlier chapter.

286 We see that the 1,019 DEGs from the Galbraith dataset form pretty clean-looking visual
287 displays (Figure 1.2). We do see that the first replicate of the virus group appears somewhat
288 inconsistent with the other virus replicates in Cluster 2, confirming that this trend in the
289 data that we saw in the MDS plot carried through into the DEG results. In contrast, we

290 see that the 43 virus-related DEGs from our dataset do not look as clean in their visual
291 displays (Figure 1.3). The replicates appear strikingly inconsistent and there is not always
292 such a large difference between treatment groups. We see a similar finding when we also
293 examine a larger subset of 1,914 diet-related DEGs from our study.

294 We also used litre plots to examine the structure of individual DEGs: We see that indeed
295 the individual DEGs from our data (Figure 1.9) show less consistent replications and
296 less differences between the treatment groups compared to the individual DEGs from the
297 Galbraith data (Figure 1.8 and Figure 1.7). For the Galbraith data, we examined individual
298 DEGs from the first cluster (Figure 1.8) and second cluster (Figure 1.7) because the second
299 cluster was a bit less ideal due to its inconsistent first replicate of the treatment group.

300 Finally, we looked at scatterplot matrices to assess the DEGs. We created standardized
301 scatterplot matrices for each of the four clusters (Figure 1.2) of the Galbraith data (Figures
302 1.10, 1.11, 1.12, and 1.13). We also created standardized scatterplot matrices for our data.
303 However, as our dataset contained 24 samples, we would need to include 276 scatterplots in
304 our matrix, which would not be helpful. As a result, we created four scatterplot matrices
305 of our data, each with subsets of 6 samples to be more comparable to the Galbraith data
306 (Figures 1.14, 1.15, 1.16, and 1.17). We can again confirm through these plots that the
307 DEGs from the Galbraith data appeared more as expected: Deviating more from the $x=y$
308 line in the treatment scatterplots while staying close to the $x=y$ line in replicate scatterplots.

309 Despite the DEGs from the Galbraith dataset displaying the expected patterns more than
310 those from our dataset, there was significant overlap in the DEGs between the two studies
311 (Figure 1.5).

312 1.3.6 Resilience versus resistance

313 To investigate whether

314 (Table 1.1

315 Figure 1.18

³¹⁶ **1.4 Discussion**

³¹⁷ **1.5 Appendix**

A	OUR DIET EFFECT	C higher	R higher	Total
DESeq2	1033	881	1914	
EdgeR	889	832	1721	
Limma	851	789	1640	

B	OUR VIRUS EFFECT	V higher	C higher	Total
DESeq2	38	5	43	
EdgeR	17	3	20	
Limma	0	0	0	

C	GALBRAITH VIRUS EFFECT	V higher	C higher	Total
DESeq2	795	224	1019	
EdgeR	580	150	730	
Limma	193	20	213	

Table 1.1: Number of DEGs across three analysis pipelines for (A) the diet effect in our study, (B) the virus main effect in our study, and (C) the virus main effect in the Galbraith study.

CHAPTER 1. GENE EXPRESSION RESPONSES TO DIET QUALITY AND VIRAL
12 INFECTION IN APIS MELLIFERA

Pathway Term	# of Genes	Benjamini	Example Genes
Wnt signaling pathway	15	2.20E-03	<i>1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase, armadillo segment polarity protein, calcium/calmodulin-dependent protein kinase II, casein kinase I-like, C-terminal-binding protein, division abnormally delayed protein, histone acetyltransferase p300-like, protein kinase, serine/threonine-protein kinase NLK, stress-activated protein kinase JNK</i>
Dorso-ventral axis formation	8	2.80E-02	<i>CUGBP Elav-like family member 2, ETS-like protein pointed, cytoplasmic polyadenylation element-binding protein 2, encore, epidermal growth factor receptor-like, neurogenic locus Notch protein, protein giant-lease, protein son of sevenless</i>
Hippo signaling pathway	12	3.00E-02	<i>actin, cadherin-related tumor suppressor, casein kinase I-like, cisks large tumor suppressor protein, division abnormally delayed protein, hemicentin-2, protein dachsous, protein expanded-like, stress-activated protein kinase JNK</i>
Circadian rhythm	4	2.40E-01	<i>casein kinase I-like, protein cycle, protein kinase shaggy, thyrotroph embryonic factor</i>
mRNA surveillance pathway	10	2.60E-01	<i>cleavage and polyadenylation specificity factor subunit CG7185, eukaryotic peptide chain release factor GTP-binding subunit ERF3A, heterogeneous nuclear ribonucleoprotein 27C, polyadenylate-binding protein 1, regulator of nonsense transcripts 1, serine/threonine-protein kinase SMG1, serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit gamma isoform-like, serine/threonine-protein phosphatase alpha-2 isoform</i>
Insulin resistance	8	2.80E-01	<i>insulin-like receptor-like (InR-2), long-chain fatty acid transport protein 1, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein kinase shaggy, serine/threonine-protein phosphatase alpha-2 isoform, stress-activated protein kinase JNK, tyrosine-protein phosphatase non-receptor type 61F-like</i>
Inositol phosphate metabolism	8	2.90E-01	<i>1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase classes I and II, inositol oxygenate, methylmalonate-semialdehyde dehydrogenase (acylating)-like protein, multiple inositol polyphosphate phosphatase 1-like, myotubularin-related protein 4, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, uncharacterized oxidoreductase YrbE-like</i>
FoxO signaling pathway	9	3.00E-01	<i>casein kinase I-like, epidermal growth factor receptor-like, histone acetyltransferase p300-like, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein son of seven less, serine/threonine-protein kinase NLK, stress-activated protein kinase JNK</i>
ECM-receptor interaction	5	3.20E-01	<i>agrin-like, collagen alpha-1 (IV) chain, collagen alpha-5 (IV) chain, dystroglycan, integrin beta-PS-like</i>
Phototransduction	6	3.30E-01	<i>1-phosphatidylinositol 4,5-biphosphate phosphodiesterase, actin muscle-like, calcium/calmodulin-dependent protein kinase II, G protein-coupled receptor kinase 1, protein kinase</i>
Notch signaling pathway	5	3.80E-01	<i>C-terminal-binding protein, histone acetyltransferase p300-like, neurogenic locus Notch protein, protein jagged-1, protein numb</i>
Jak-STAT signaling pathway	4	3.90E-01	<i>histone acetyltransferase p300-like, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein son of sevenless</i>
MAPK signaling pathway	4	4.40E-01	<i>epidermal growth factor receptor-like, ETS-like protein pointed, protein son of sevenless, proto-oncogene tyrosine-protein kinase ROS</i>
Carbon metabolism	12	4.50E-01	<i>2-oxoglutarate dehydrogenase, aminomethyltransferase, fructose-bisphosphate aldolase, glycine dehydrogenase (decarboxylating), L-threonine ammonia-lyase, methylmalonate-semialdehyde dehydrogenase [acylating]-like protein, NADP-dependent malic enzyme, probable aconitate hydratase, PTS-dependent dihydroxyacetone kinase, pyruvate carboxylase, succinate dehydrogenase [ubiquinone] iron-sulfur subunit</i>

Table 1.2: Pathways related to diet main effect Castanea DEGs.

Pathway Term	# of Genes	Benjamini	Example Genes
Wnt signaling pathway	15	2.20E-03	<i>1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase, armadillo segment polarity protein, calcium/calmodulin-dependent protein kinase II, casein kinase I-like, C-terminal-binding protein, division abnormally delayed protein, histone acetyltransferase p300-like, protein kinase, serine/threonine-protein kinase NLK, stress-activated protein kinase JNK</i>
Dorso-ventral axis formation	8	2.80E-02	<i>CUGBP Elav-like family member 2, ETS-like protein pointed, cytoplasmic polyadenylation element-binding protein 2, encore, epidermal growth factor receptor-like, neurogenic locus Notch protein, protein giant-lease, protein son of sevenless</i>
Hippo signaling pathway	12	3.00E-02	<i>actin, cadherin-related tumor suppressor, casein kinase I-like, cisks large tumor suppressor protein, division abnormally delayed protein, hemicentin-2, protein dachsous, protein expanded-like, stress-activated protein kinase JNK</i>
Circadian rhythm	4	2.40E-01	<i>casein kinase I-like, protein cycle, protein kinase shaggy, thyrotroph embryonic factor</i>
mRNA surveillance pathway	10	2.60E-01	<i>cleavage and polyadenylation specificity factor subunit CG7185, eukaryotic peptide chain release factor GTP-binding subunit ERF3A, heterogeneous nuclear ribonucleoprotein 27C, polyadenylate-binding protein 1, regulator of nonsense transcripts 1, serine/threonine-protein kinase SMG1, serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit gamma isoform-like, serine/threonine-protein phosphatase alpha-2 isoform</i>
Insulin resistance	8	2.80E-01	<i>insulin-like receptor-like (InR-2), long-chain fatty acid transport protein 1, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein kinase shaggy, serine/threonine-protein phosphatase alpha-2 isoform, stress-activated protein kinase JNK, tyrosine-protein phosphatase non-receptor type 61F-like</i>
Inositol phosphate metabolism	8	2.90E-01	<i>1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase classes I and II, inositol oxygenate, methylmalonate-semialdehyde dehydrogenase (acylating)-like protein, multiple inositol polyphosphate phosphatase 1-like, myotubularin-related protein 4, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, uncharacterized oxidoreductase YrbE-like</i>
FoxO signaling pathway	9	3.00E-01	<i>casein kinase I-like, epidermal growth factor receptor-like, histone acetyltransferase p300-like, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein son of seven less, serine/threonine-protein kinase NLK, stress-activated protein kinase JNK</i>
ECM-receptor interaction	5	3.20E-01	<i>agrin-like, collagen alpha-1 (IV) chain, collagen alpha-5 (IV) chain, dystroglycan, integrin beta-PS-like</i>
Phototransduction	6	3.30E-01	<i>1-phosphatidylinositol 4,5-biphosphate phosphodiesterase, actin muscle-like, calcium/calmodulin-dependent protein kinase II, G protein-coupled receptor kinase 1, protein kinase</i>
Notch signaling pathway	5	3.80E-01	<i>C-terminal-binding protein, histone acetyltransferase p300-like, neurogenic locus Notch protein, protein jagged-1, protein numb</i>
Jak-STAT signaling pathway	4	3.90E-01	<i>histone acetyltransferase p300-like, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein son of sevenless</i>
MAPK signaling pathway	4	4.40E-01	<i>epidermal growth factor receptor-like, ETS-like protein pointed, protein son of sevenless, proto-oncogene tyrosine-protein kinase ROS</i>
Carbon metabolism	12	4.50E-01	<i>2-oxoglutarate dehydrogenase, aminomethyltransferase, fructose-bisphosphate aldolase, glycine dehydrogenase (decarboxylating), L-threonine ammonia-lyase, methylmalonate-semialdehyde dehydrogenase [acylating]-like protein, NADP-dependent malic enzyme, probable aconitate hydratase, PTS-dependent dihydroxyacetone kinase, pyruvate carboxylase, succinate dehydrogenase [ubiquinone] iron-sulfur subunit</i>

Table 1.3: Pathways related to diet main effect Rockrose DEGs.

A	OUR PAIRS (NC, VC)	NC higher	VC higher	Total
DESeq2		0	0	0
EdgeR		0	0	0
Limma		0	0	0

B	OUR PAIRS (NR, VR)	VR higher	NR higher	Total
DESeq2		152	26	178
EdgeR		87	9	96
Limma		0	0	0

C	OUR PAIRS (VC, VR)	VC higher	VR higher	Total
DESeq2		247	129	376
EdgeR		130	59	189
Limma		10	1	11

D	OUR PAIRS (NC, VR)	NC higher	VR higher	Total
DESeq2		496	278	774
EdgeR		320	215	535
Limma		108	47	155

E	OUR PAIRS (VC, NR)	VC higher	NR higher	Total
DESeq2		540	415	955
EdgeR		431	251	682
Limma		140	91	231

F	OUR PAIRS (NC, NR)	NC higher	NR higher	Total
DESeq2		601	340	941
EdgeR		502	295	797
Limma		219	139	358

Table 1.4: Number of DEGs across three analysis pipelines for all six treatment pair combinations between the diet and virus factor.

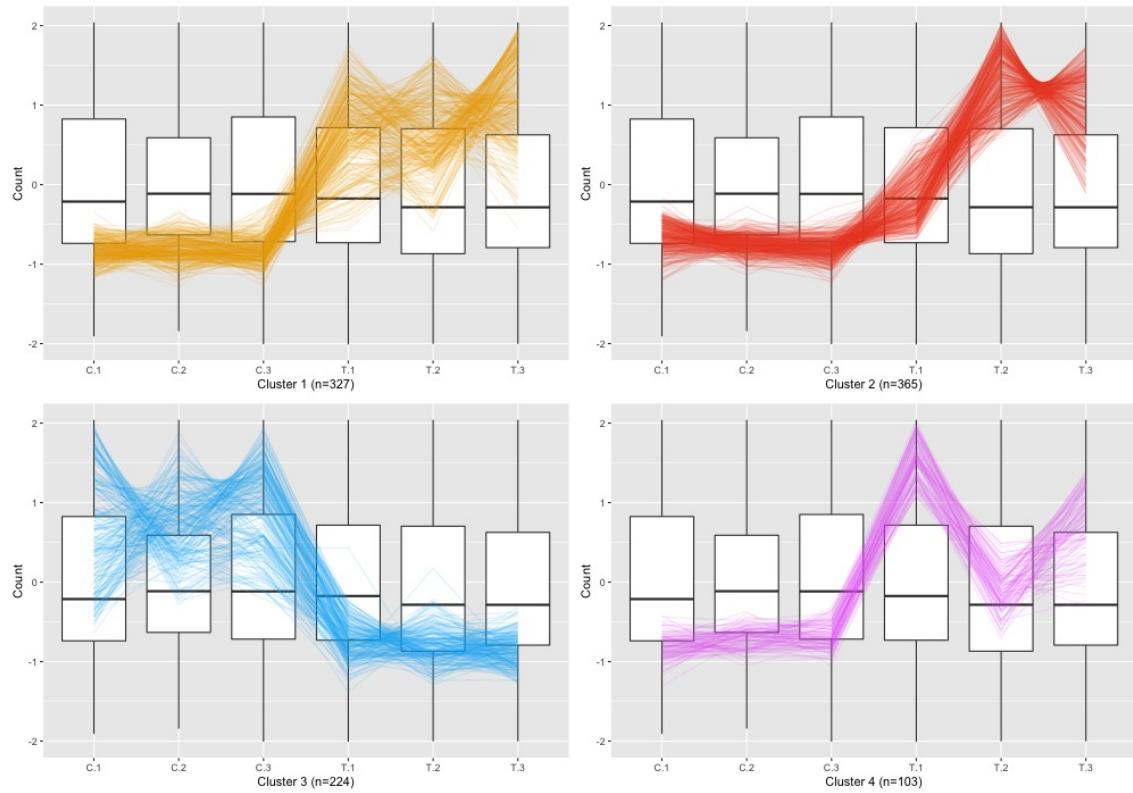


Figure 1.2: Parallel coordinate plots of DEGs between the virus-infected and control groups of the Galbraith study.

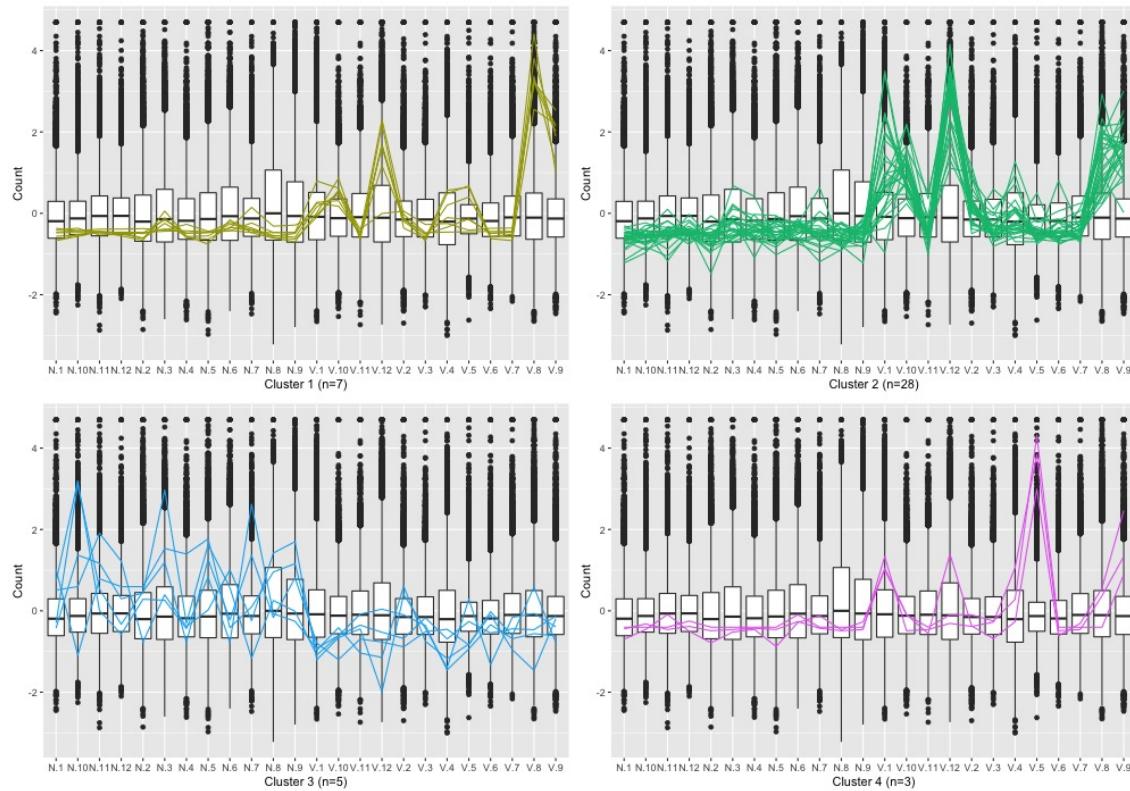


Figure 1.3: Parallel coordinate plots of DEGs between the virus-infected and control groups of our study.

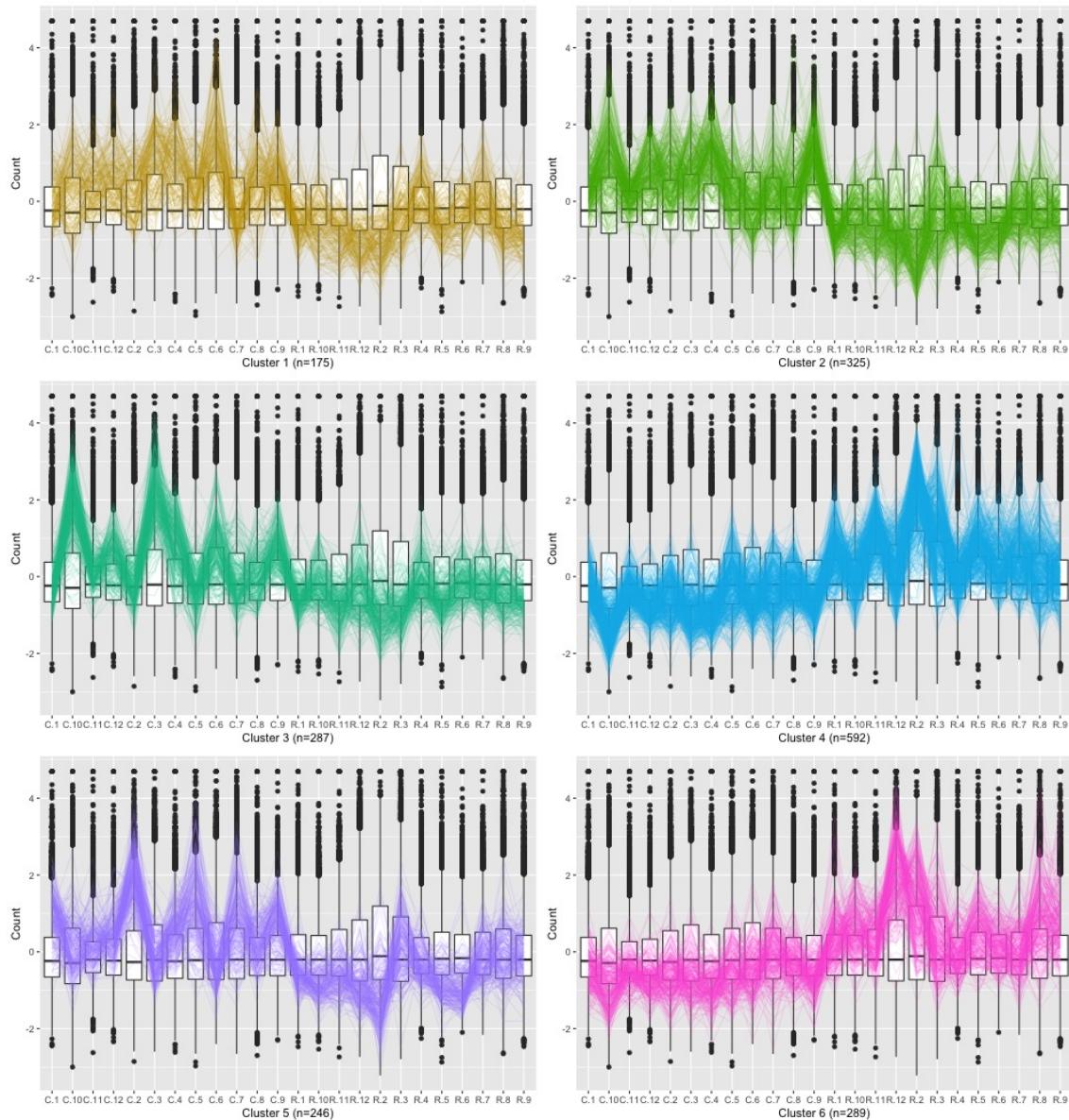


Figure 1.4: Parallel coordinate plots of DEGs between the Castanea and Rockrose groups of our study.

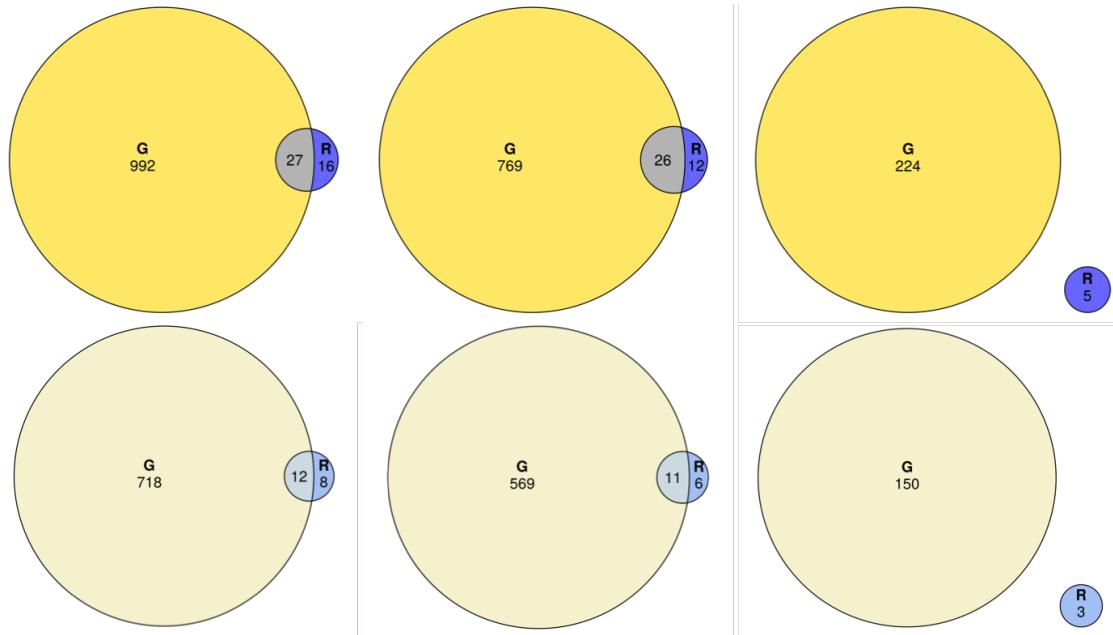


Figure 1.5: Venn diagrams comparing the DEG overlaps of the diet and treatment groups between the Galbraith study and our study. Top row left to right: DESeq results for all DEGs, virus DEGs, and control DEGs. Bottom row left to right: EdgeR results for all DEGs, virus DEGs, and control DEGs.

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

Table 1.5: 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.

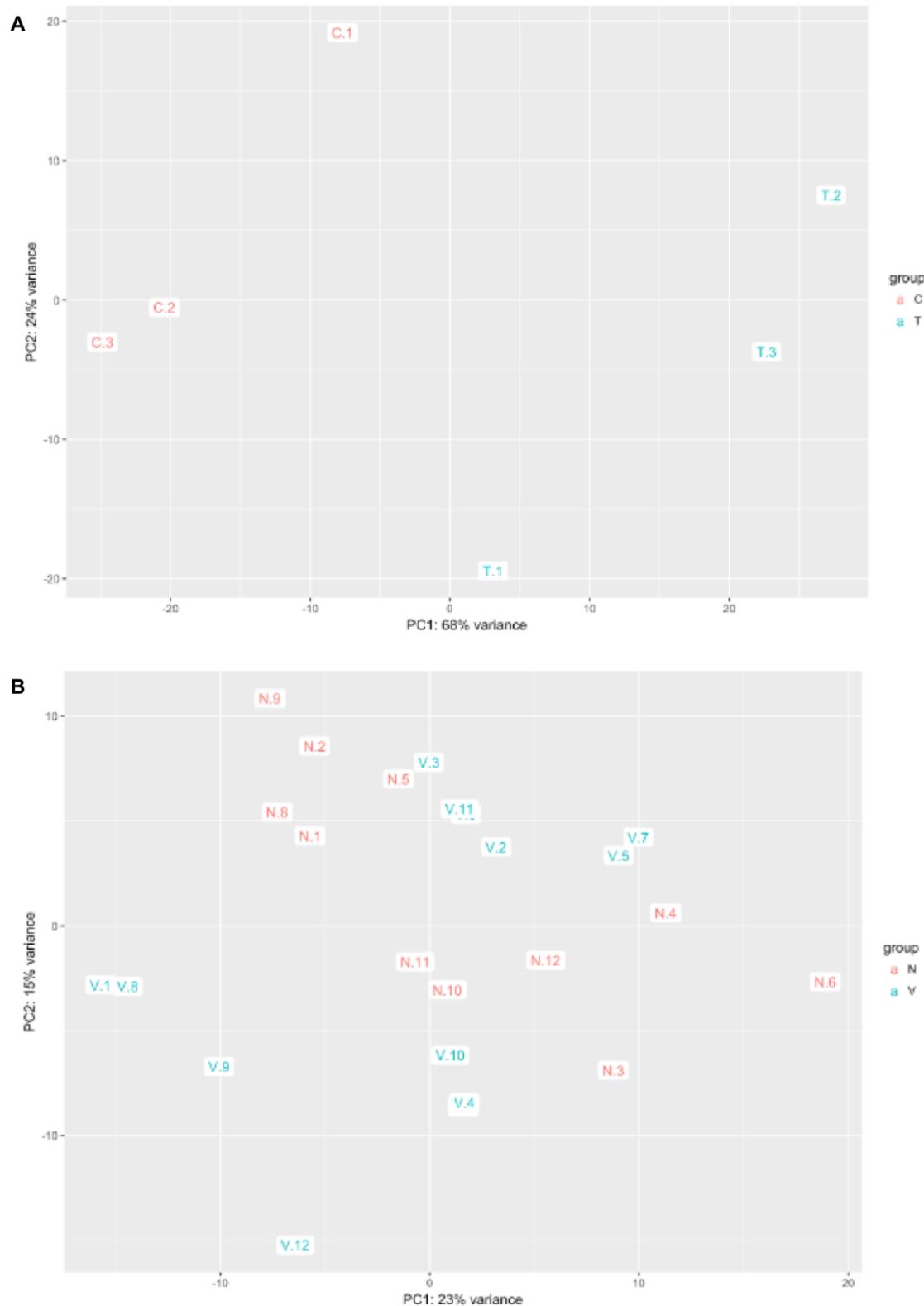


Figure 1.6: MDS plots constructed from DESeq2 package for the Galbraith dataset (A) and our dataset (B).

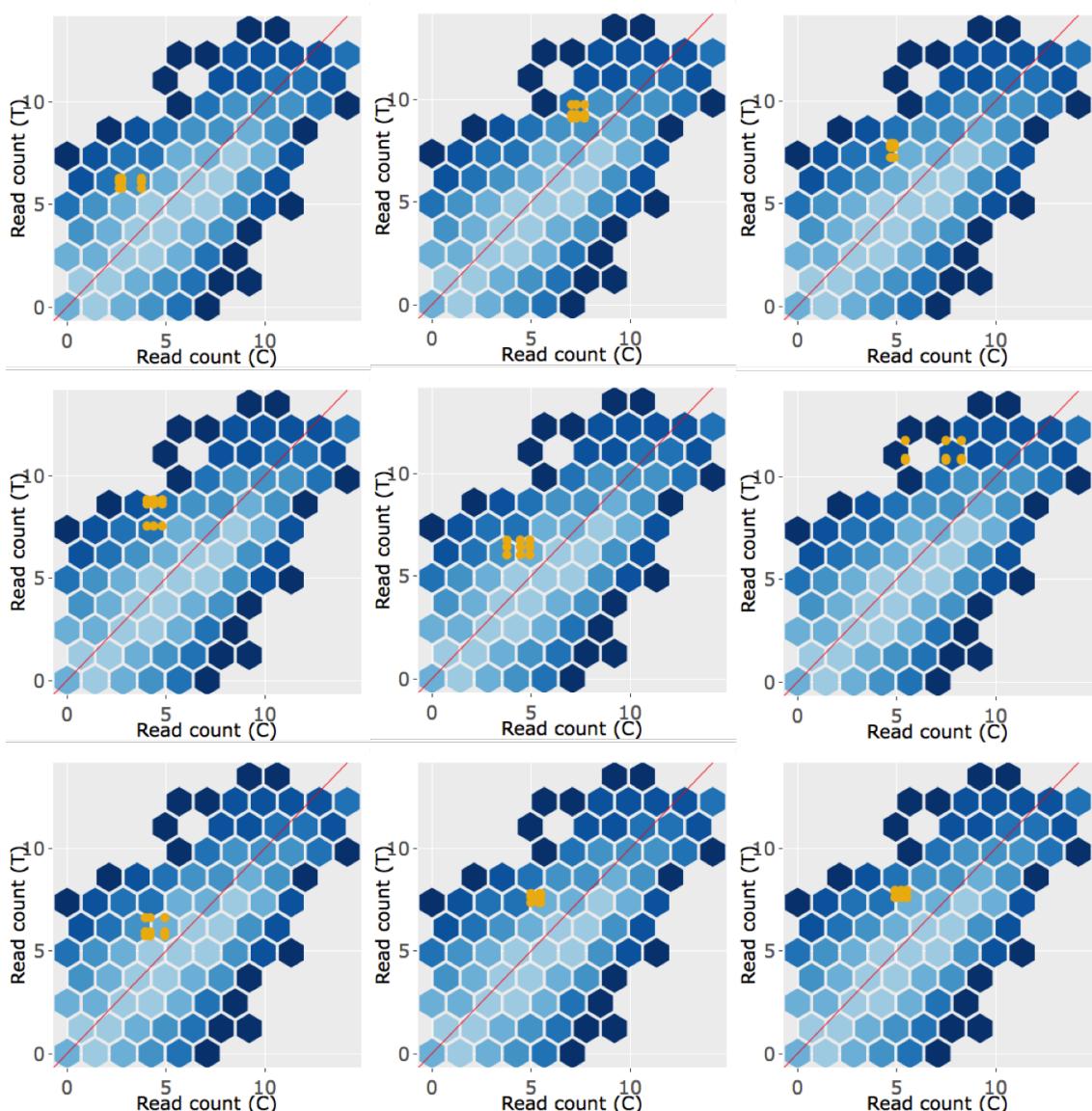


Figure 1.7: Example litre plots of the nine DEGs with the lowest FDR values from Cluster 1 (back in Figure 1.2) of the Galbraith dataset.

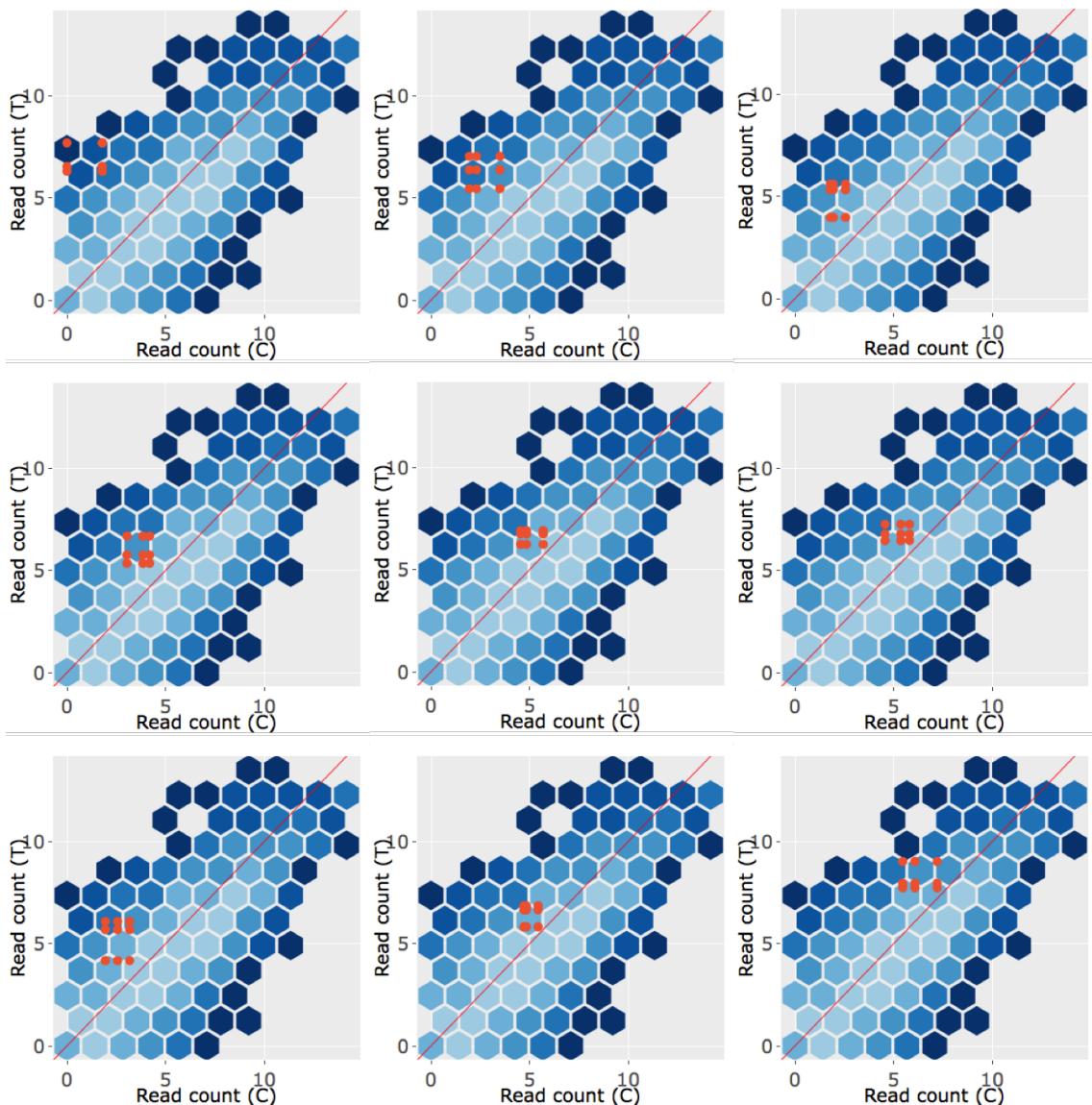


Figure 1.8: Example litre plots of the nine DEGs with the lowest FDR values from Cluster 2 (back in Figure 1.2) of the Galbraith dataset.

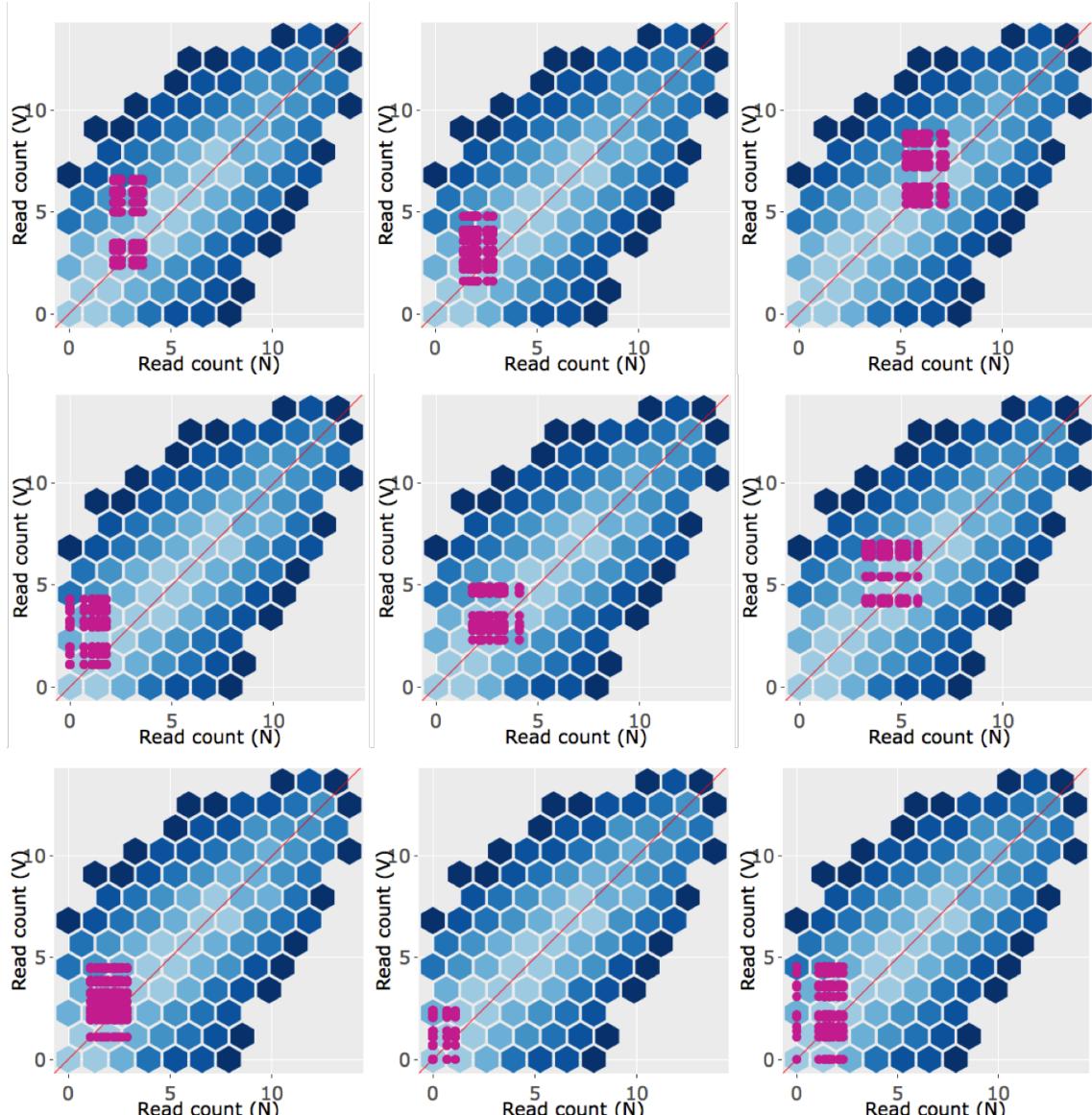


Figure 1.9: Example litre plots of the nine DEGs with the lowest FDR values from the 43 DEGs of our dataset.

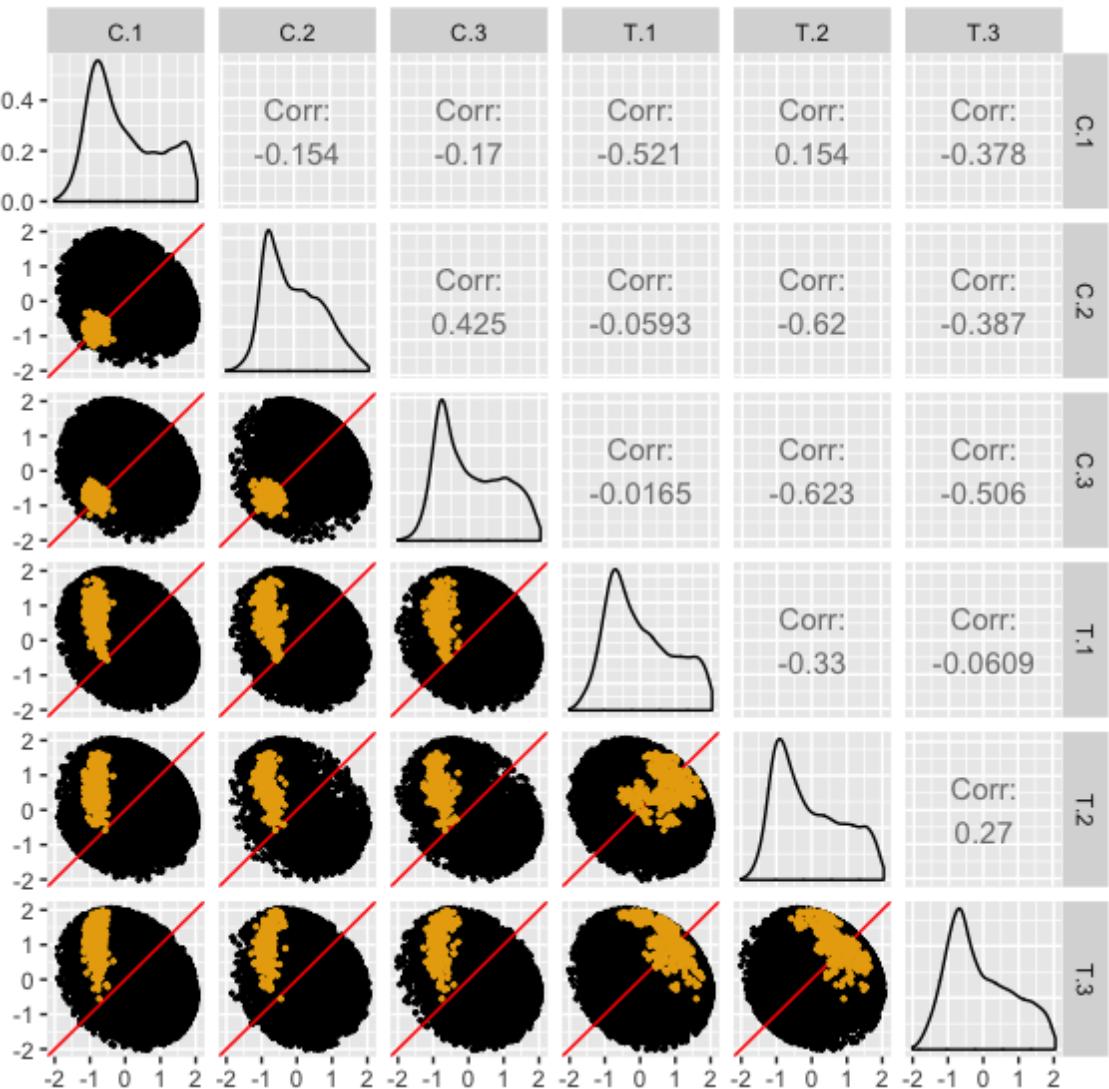


Figure 1.10: The DEGs from the first cluster of the Galbraith dataset (back in Figure 1.2) superimposed onto all genes in the form of a scatterplot matrix. The data has been standardized.

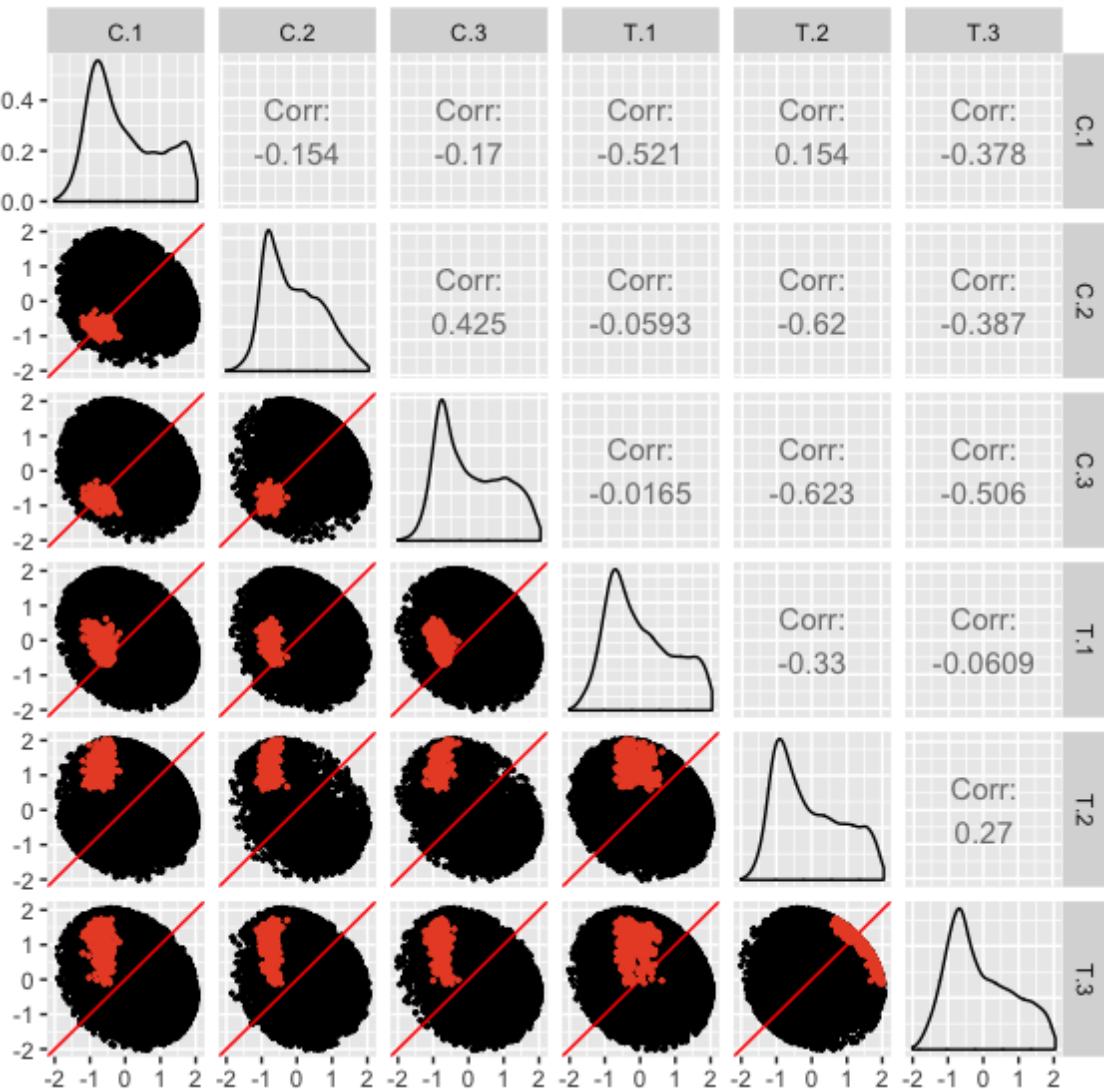


Figure 1.11: The DEGs from the second cluster of the Galbraith dataset (back in Figure 1.2) superimposed onto all genes in the form of a scatterplot matrix. The data has been standardized.

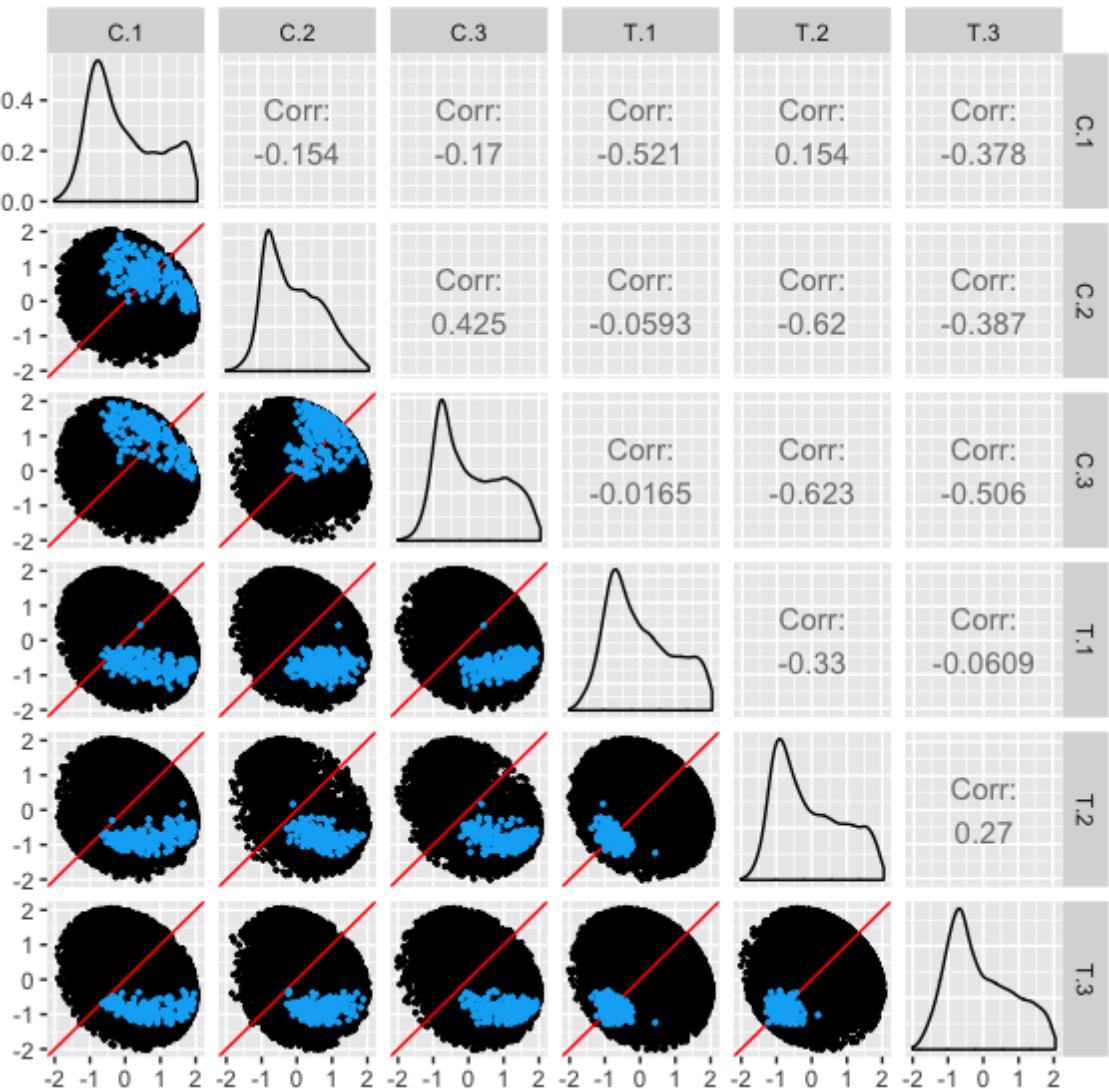


Figure 1.12: The DEGs from the third cluster of the Galbraith dataset (back in Figure 1.2) superimposed onto all genes in the form of a scatterplot matrix. The data has been standardized.

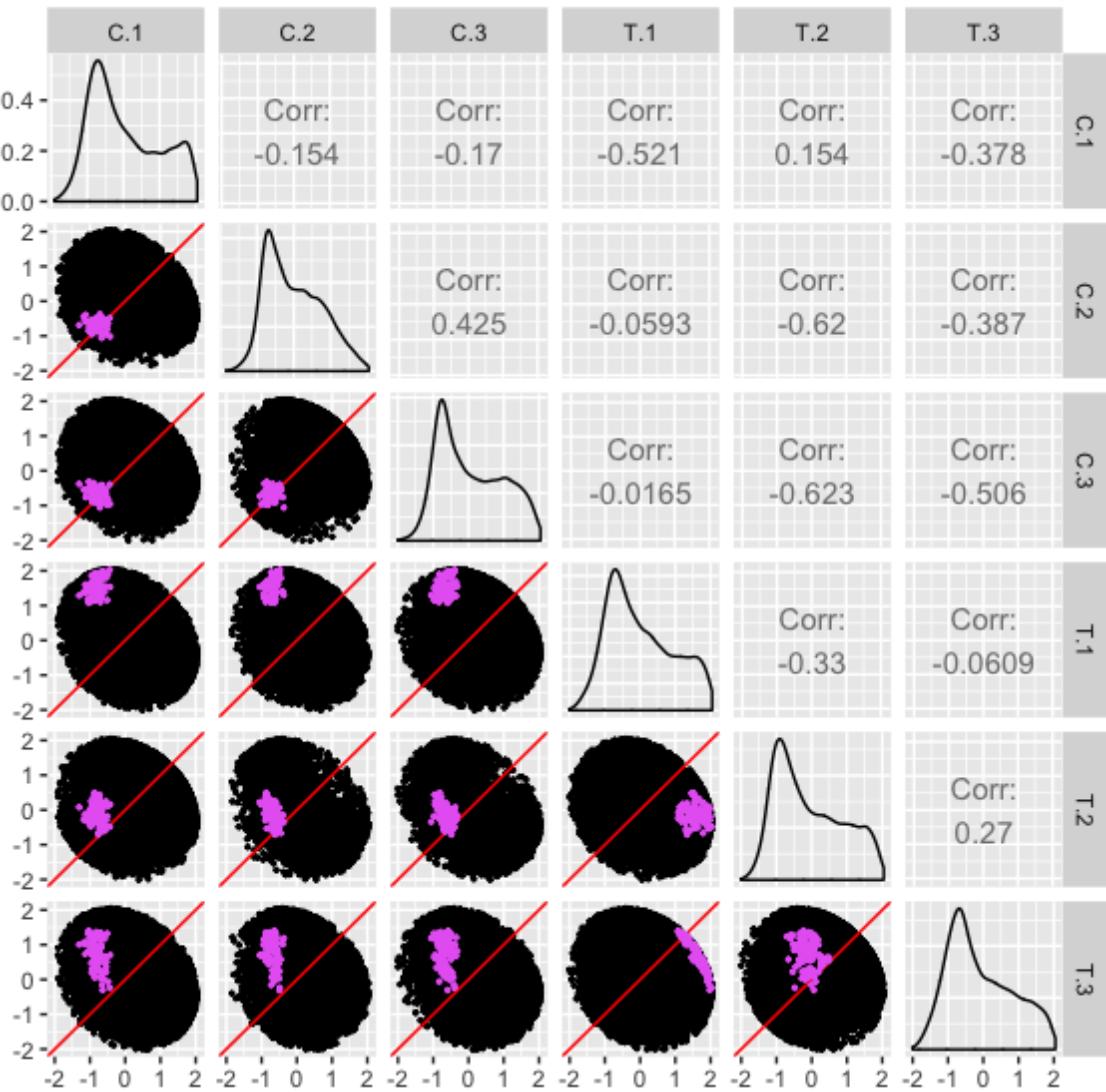


Figure 1.13: The DEGs from the fourth cluster of the Galbraith dataset (back in Figure 1.2) superimposed onto all genes in the form of a scatterplot matrix. The data has been standardized.

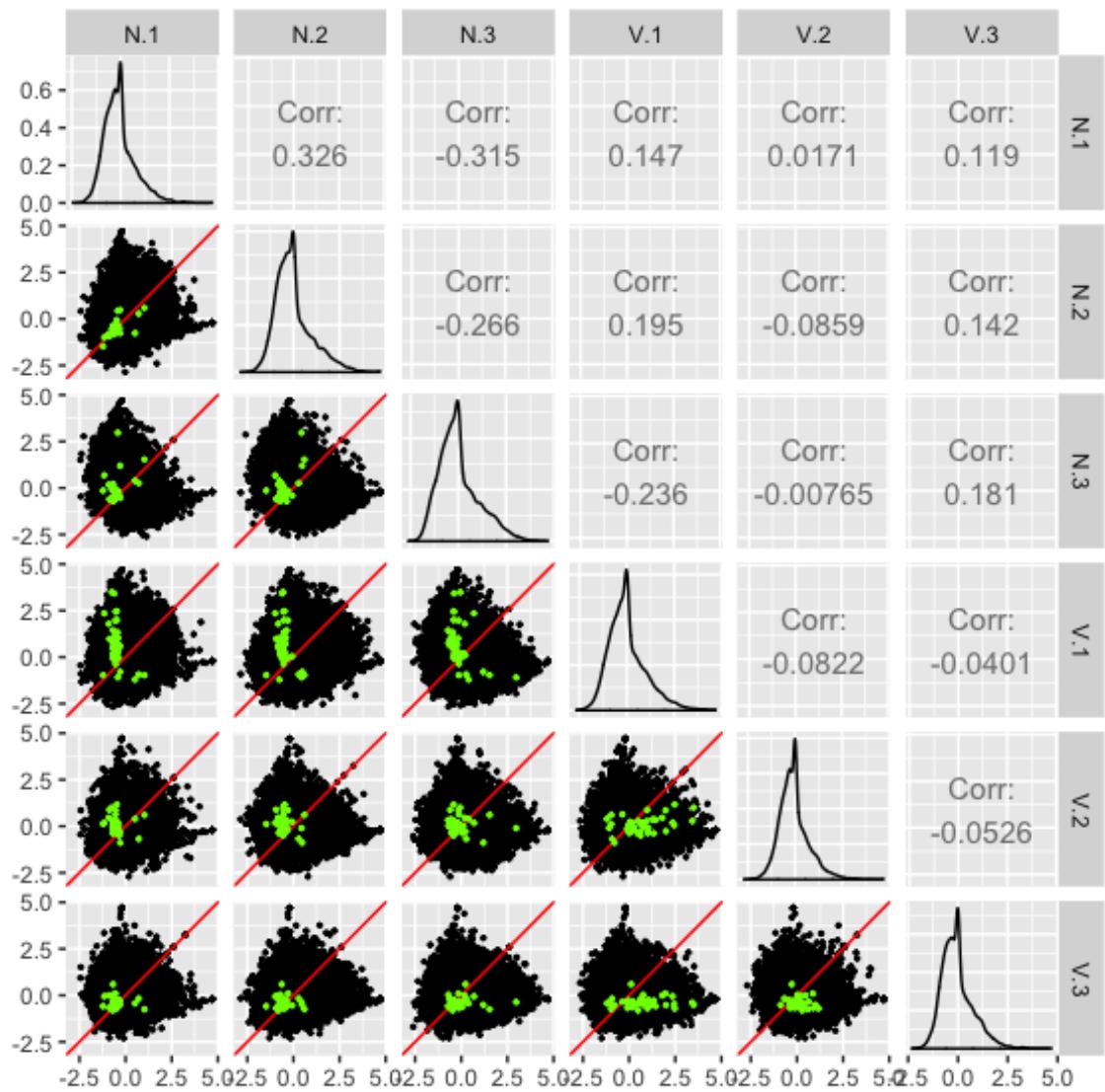


Figure 1.14: The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only replicates 1, 2, and 3 are shown from both treatment groups. The data has been standardized.

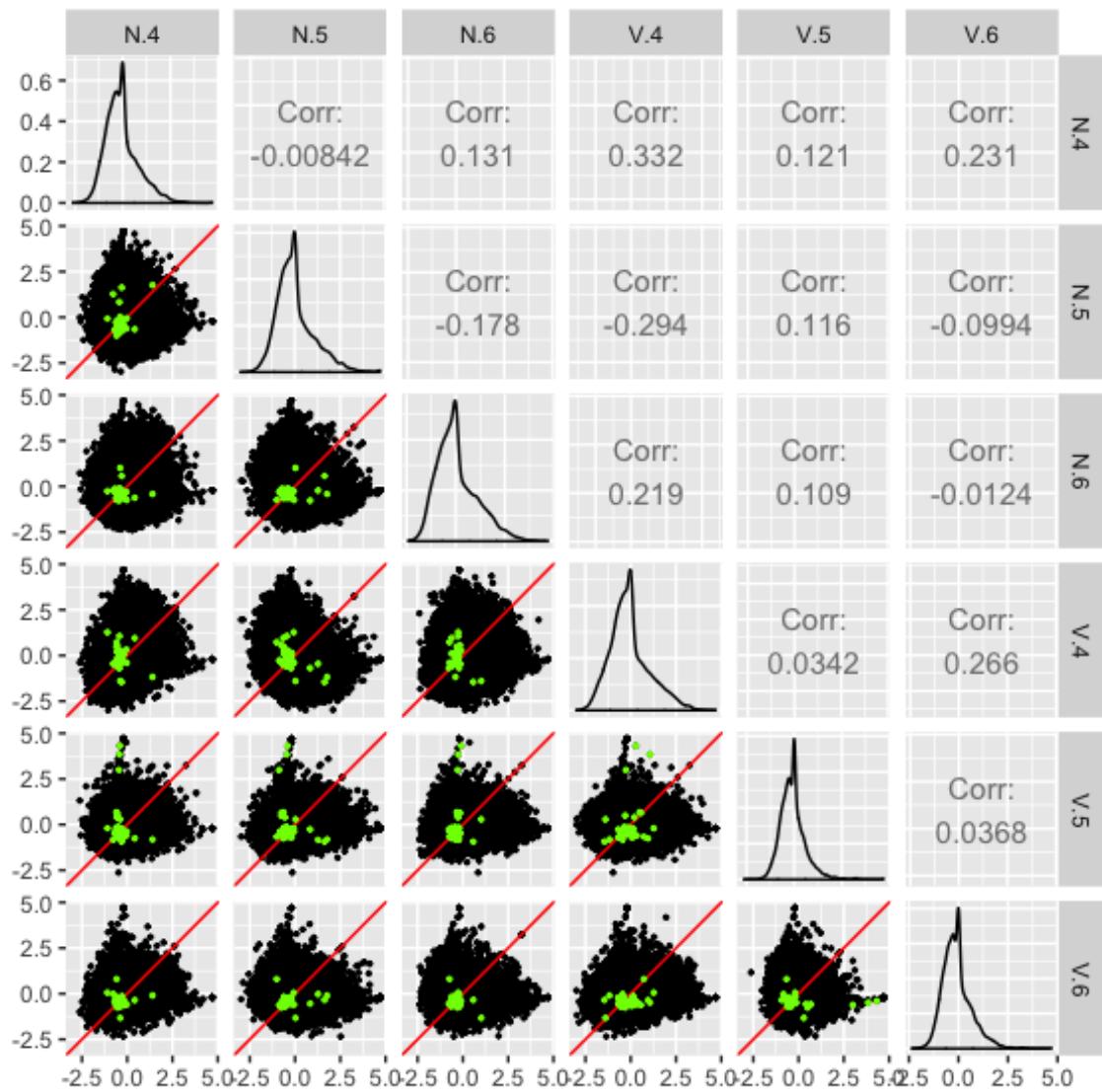


Figure 1.15: The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only replicates 4, 5, and 6 are shown from both treatment groups. The data has been standardized.

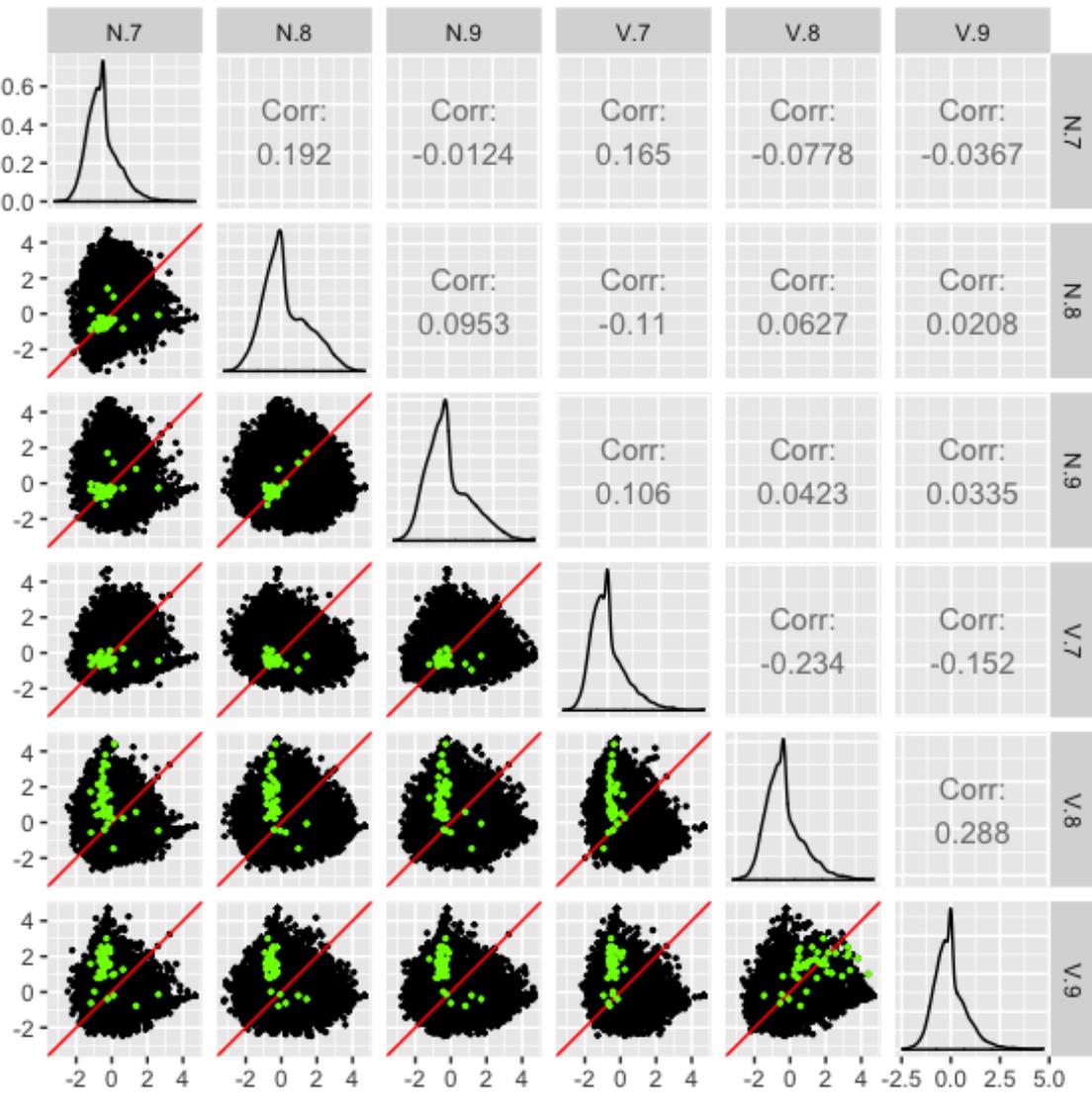


Figure 1.16: The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only replicates 7, 8, and 9 are shown from both treatment groups. The data has been standardized.

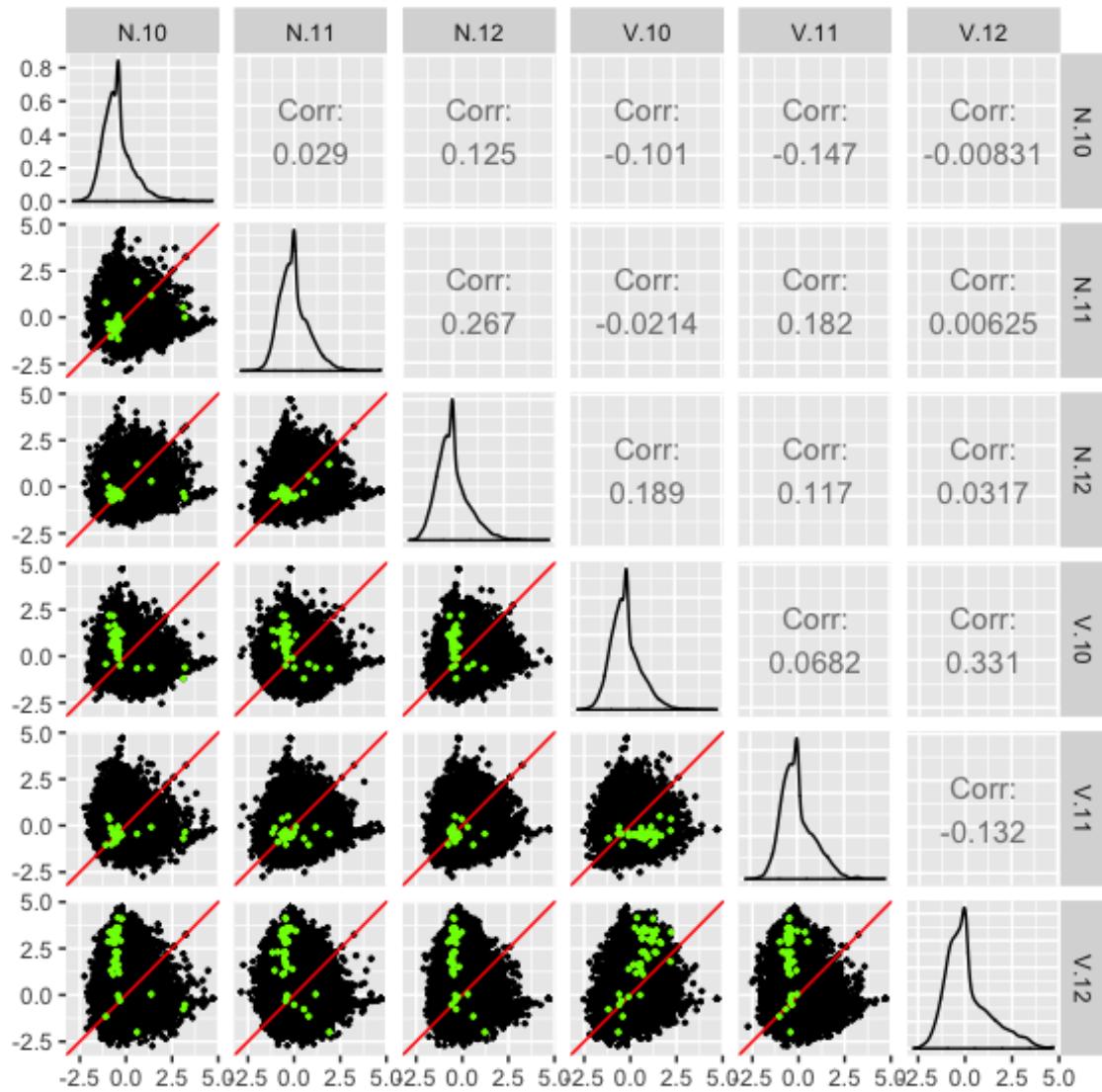


Figure 1.17: The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only replicates 10, 11, and 12 are shown from both treatment groups. The data has been standardized.

Pathway Term	# of Genes	Benjamini	Example Genes
Wnt signaling pathway	11	2.20E-04	<i>1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase, C-terminal-binding protein, calcium/calmodulin-dependent protein kinase II, casein kinase I-like, division abnormally delayed protein, histone acetyltransferase p300-like, protein kinase C, protein kinase shaggy, protein prickle-like, serine/threonine-protein kinase NLK</i>
Circadian rhythm	4	2.40E-02	<i>casein kinase I-like, period circadian protein, protein kinase shaggy, thyrotroph embryonic factor</i>
Hippo signaling pathway	7	5.60E-02	<i>actin, muscle-like, casein kinase I-like, division abnormally delayed protein, hemicentin-2, protein dachsous, serine/threonine-protein kinase Warts</i>
Phototransduction	5	7.30E-02	<i>1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase, G protein-coupled receptor kinase 1, actin (muscle-like), calcium/calmodulin-dependent protein kinase II, protein kinase C</i>
FoxO signaling pathway	6	1.50E-01	<i>casein kinase I-like, histone acetyltransferase p300-like, insulin-like receptor-like (InR-2), phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, serine/threonine-protein kinase NLK</i>
Notch signaling pathway	4	1.80E-01	<i>C-terminal-binding protein, histone acetyltransferase p300-like, protein jagged-1, protein numb</i>
Insulin resistance	5	2.10E-01	<i>insulin-like receptor-like (InR-2), phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein kinase shaggy, serine/threonine-protein phosphatase alpha-2 isoform</i>
mRNA surveillance pathway	6	2.30E-01	<i>cleavage and polyadenylation specificity factor subunit CG7185, heterogeneous nuclear ribonucleoprotein 27C, serine/threonine-protein kinase SMG1, serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit gamma isoform-like, serine/threonine-protein phosphatase alpha-2 isoform</i>
Jak-STAT signaling pathway	3	2.50E-01	<i>histone acetyltransferase p300-like, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform</i>
Phosphatidylinositol signaling system	5	2.70E-01	<i>1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase, diacylglycerol kinase theta, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein kinase C</i>

Table 1.6: Pathways analysis results for the 601 DEGs that were upregulated in the NC treatment in the NC versus NR treatment pair analysis.

Pathway Term	# of Genes	Benjamini	Example Genes
Sphingolipid metabolism	4	6.00E-01	<i>alkaline ceramidase, putative neutral sphingomyelinase, serine palmitoyltransferase 1, sphingosine-1-phosphate phosphatase 1-like</i>
SNARE interactions in vesicular transport	4	7.00E-01	<i>BET1 homolog, Golgi SNAP receptor complex member 2, syntaxin-7, vesicle transport protein USE1</i>
Basal transcription factors	4	7.30E-01	<i>cyclin-dependent kinase 7, general transcription factor IIF subunit 2, transcription initiation factor IIE subunit beta, transcription initiation factor TFIID subunit 10-like</i>

Table 1.7: Pathways analysis results for the 340 DEGs that were upregulated in the NR treatment in the NC versus NR treatment pair analysis.

Pathway Term	# of Genes	Benjamini	Example Genes
Hippo signaling pathway	5	7.50E-02	<i>actin (muscle-like), cadherin-related tumor suppressor, casein kinase I-like, hemicentin-2, stress-activated protein kinase JNK</i>
Wnt signaling pathway	4	3.00E-01	<i>1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase, armadillo segment polarity protein, casein kinase I-like, stress-activated protein kinase JNK</i>
Circadian rhythm	2	5.50E-01	<i>casein kinase I-like, thyrotroph embryonic factor</i>

Table 1.8: Pathways analysis results for the 247 DEGs that were upregulated in the VC treatment in the VC versus VR treatment pair analysis.

Pathway Term	# of Genes	Benjamini	Example Genes
Fanconi anemia pathway	4	1.60E-02	<i>breast cancer type 2 susceptibility protein homolog, DNA polymerase eta, E3 ubiquitin-protein ligase FANCL, Fanconi anemia group M protein</i>

Table 1.9: Pathways analysis results for the 129 DEGs that were upregulated in the VR treatment in the VC versus VR treatment pair analysis.

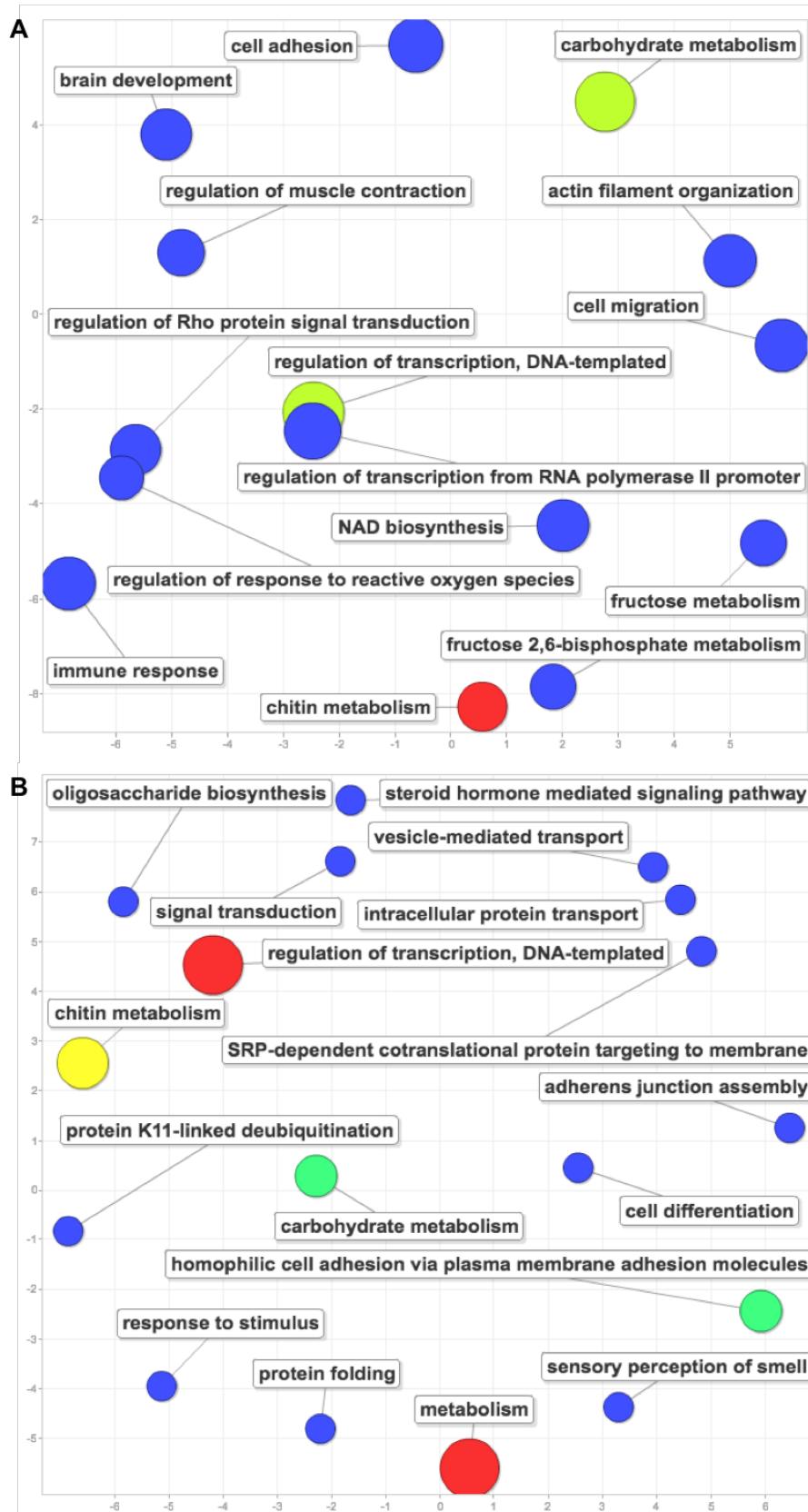


Figure 1.18: Gene ontology results for the 122 DEGs related to our “resilience” hypothesis (A) and for the 125 DEGs related to our “resistance” hypothesis (B).

Contrast	DEGs	Interpretation	Results
V vs N	43	Genes that change expression due to virus effect regardless of diet status in bees	Table 1.5
NC vs NR	941	Genes that change expression due to diet effect in uninfected bees	Tables 1.6 and 1.7
VC vs VR	376	Genes that change expression due to diet effect in infected bees	Tables 1.8 and 1.9
VC upregulated in VC vs VR overlaped with NC upregulated in NC vs NR	122	“Resilience” genes that are turned on by good diet regardless of virus infection status in bees	Figure 1.18A
VC upregulated in VC vs VR but NC is not upregulated in NC vs NR	125	“Resistance” genes that are turned on by good diet only in infected bees	Figure 1.18B

Table 1.10: Contrasts in our study for assessing GO and pathways analysis.

Bibliography

- 319 C. Alaux, F. Ducloz, and D. Crauser Y. Le Conte. Diet effects on honeybee immunocompe-
320 tence. *Biol. Lett.*, 6:562–565, 2010.
- 321 C. Alaux, C. Dantec, H. Parrinello, and Y. Le Conte. Nutrigenomics in honey bees: digital
322 gene expression analysis of pollen’s nutritive effects on healthy and varroa-parasitized
323 bees. *BMC Genomics*, 12:496, 2011.
- 324 Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and
325 powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*
326 (*Methodological*), 57:289–300, 1995.
- 327 J. Bond, K. Plattner, and K. Hunt. *Fruit and Tree Nuts Outlook: Economic Insight U.S.*
328 *Pollination- Services Market*. USDA. Economic Research Service Situation and Outlook
329 FTS-357SA, 2014.
- 330 R. Brodschneider and K. Crailsheim. Nutrition and health in honey bees. *Apidologie*, 41:
331 278–294, 2010.
- 332 R. Brodschneider, G. Arnold, N. Hrassnigg, and K. Crailsheim. Does patriline composition
333 change over a honey bee queen’s lifetime? *Insects*, 3:857–869, 2012.
- 334 D. Caron and R. Sagili. Honey bee colony mortality in the Pacific Northwest: Winter
335 2009/2010. *Am Bee J*, 151:73–76, 2011.
- 336 J. Carrillo-Tripp, A.G. Dolezal, M.J. Goblirsch, W.A. Miller, A.L. Toth, and B.C. Bonning.
337 In vivo and in vitro infection dynamics of honey bee viruses. *Sci Rep*, 6:22265, 2016.
- 338 Y.P. Chen and R. Siede. Honey bee viruses. *Adv Virus Res*, 70:33–80, 2007.
- 339 Y.P. Chen, J.S. Pettis, M. Corona, W.P. Chen, C.J. Li, M. Spivak, P.K. Visscher,
340 G. DeGrandi-Hoffman, H. Boncristiani, Y. Zhao, D. van Engelsdorp, K. Delaplane,
341 L. Solter, F. Drummond, M. Kramer, W.I. Lipkin, G. Palacios, M.C. Hamilton, B. Smith,

- 342 S.K. Huang, H.Q. Zheng, J.L. Li, X. Zhang, X.F. Zhou, L.Y. Wu, J.Z. Zhou, M-L. Lee,
343 E.W. Teixeira, Z.G. Li, and J.D. Evans. Israeli acute paralysis virus: Epidemiology,
344 pathogenesis and implications for honey bee health. *Plos Pathog*, 10:e1004261, 2014.
- 345 Honey Bee Genome Sequencing Consortium. Finding the missing honey bee genes: lessons
346 learned from a genome upgrade. *BMC Genomics*, 15:86, 2014.
- 347 Y. Le Conte, J-L. Brunet, C. McDonnell, and C. Alaux. *Interactions between risk factors
348 in honey bees*. CRC Press, 2011.
- 349 R.S. Cornman, D.R. Tarpy, Y. Chen, L. Jeffreys, D. Lopez, and J.S. Pettis. Pathogen webs
350 in collapsing honey bee colonies. *Plos One*, 7:e43562, 2012.
- 351 D.L. Cox-Foster, S. Conlan, E.C. Holmes, G. Palacios, J.D. Evans, N.A. Moran, P-L. Quan,
352 T. Brieske, M. Hornig, D.M. Geiser, V. Martinson, D. vanEngelsdorp, A.L. Kalkstein,
353 A. Drysdale, J. Hui, J. Zhai, L. Cui, S.K. Hutchison, J.F. Simons, M. Egholm, J.S. Pettis,
354 and W.I. Lipkin. A metagenomic survey of microbes in honey bee colony collapse disorder.
355 *Science*, 318:283–287, 2007.
- 356 K. Crailsheim. The flow of jelly within a honeybee colony. *J Comp Physiol B*, 162:681–689,
357 1992.
- 358 K. Crailsheim, L.H.W Schneider, N. Hrassnigg, G. Bühlmann, U. Brosch, R. Gmeinbauer,
359 and B. Schöffmann. Pollen consumption and utilization in worker honeybees (*Apis
360 mellifera carnica*): dependence on individual age and function. *J Insect Physiol*, 38:
361 409–419, 1992.
- 362 R.H. Crozier and R.E. Page. On being the right size: Male contributions and multiple
363 mating in social Hymenoptera. *Behav. Ecol. Sociobiol.*, 18:105–115, 1985.
- 364 A. Decourtye, E. Mader, and N. Desneux. Landscape enhancement of floral resources for
365 honey bees in agro-ecosystems. *Apidologie*, 41:264–277, 2010.
- 366 G. DeGrandi-Hoffman and Y. Chen. Nutrition, immunity and viral infections in honey
367 bees. *Current Opinion in Insect Science*, 10:170–176, 2015.
- 368 G. DeGrandi-Hoffman, Y. Chen, E. Huang, and M.H Huang. The effect of diet on protein
369 concentration, hypopharyngeal gland development and virus load in worker honey bees
370 (*Apis mellifera L.*). *J Insect Physiol*, 56:1184–1191, 2010.
- 371 A.G. Dolezal and A.L. Toth. Feedbacks between nutrition and disease in honey bee health.
372 *Current Opinion in Insect Science*, 26:114–119, 2018.
- 373 A.G. Dolezal, J. Carrillo-Tripp, T. Judd, A. Miller, B. Bonning, and A. Toth. Interacting
374 stressors matter: Diet quality and virus infection in honey bee health. *In prep*, 2018.

- 375 C.G. Elsik, A. Tayal, C.M. Diesh, D.R. Unni, M.L. Emery, H.N. Nguyen, and D.E. Hagen.
376 Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine.
377 *Nucleic Acids Research*, 4:D793–800, 2016.
- 378 D. Van Engelsdorp and M.D. Meixner. A historical review of managed honey bee populations
379 in Europe and the United States and the factors that may affect them. *J Invertebr Pathol*,
380 103:S80–S95, 2010.
- 381 D. Van Engelsdorp, J. Jr. Hayes, R.M Underwood, and J. Pettis. A survey of honey bee
382 colony losses in the U.S., fall 2007 to spring 2008. *Plos One*, 3:e4071, 2008.
- 383 D.A. Galbraith, X. Yang, E.L. Niño, S. Yi, and C. Grozinger. Parallel epigenomic and
384 transcriptomic responses to viral infection in honey bees (*Apis mellifera*). *Plos Pathogens*,
385 11:e1004713, 2015.
- 386 N. Gallai, J-M. Salles, J. Settele, and B.B. Vaissière. Economic valuation of the vulnerability
387 of world agriculture confronted with pollinator decline. *Ecol. Econ.*, 68:810–821, 2009.
- 388 D. Goulson, E. Nicholls, C. Botías, and E.L. Rotheray. Bee declines driven by combined
389 stress from parasites, pesticides, and lack of flowers. *Science*, 347:1255957, 2015.
- 390 M.H. Haydak. Honey bee nutrition. *Annu Rev Entomol*, 15:143–156, 1970.
- 391 C. Hou, H. Rivkin, Y. Slabezki, and N. Chejanovsky. Dynamics of the presence of israeli
392 acute paralysis virus in honey bee colonies with colony collapse disorder. *Viruses*, 6:
393 2012–2027, 2014.
- 394 D.W. Huang, B.T. Sherman, and R. Lempicki. Systematic and integrative analysis of large
395 gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4:44–57, 2009a.
- 396 D.W. Huang, B.T. Sherman, and R.A. Lempicki. Bioinformatics enrichment tools: paths
397 toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37:
398 1–13, 2009b.
- 399 A-M. Klein, B.E. Vaissière, J.H. Cane, I. Steffan-Dewenter, S.A. Cunningham, C. Kremen,
400 and T. Tscharntke. Importance of pollinators in changing landscapes for world crops.
401 *Proc Biol Sci*, 274:303–313, 2007.
- 402 K. Kulhanek, N. Steinhauer, K. Rennich, D.M. Caron, R.R. Sagili, J.S. Pettis, J.D. Ellis,
403 M.E. Wilson, J.T. Wilkes, D.R. Tarpy, R. Rose, K. Lee, J. Rangel, and D. vanEngelsdorp.
404 A national survey of managed honey bee 2014–2015 annual colony losses in the USA.
405 *Journal of Apicultural Research*, 56:328–340, 2017.
- 406 Johan Larsson. *eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses*, 2018.
407 URL <https://cran.r-project.org/package=eulerr>. R package version 4.0.0.

- 408 M. Laurent, P. Hendrikx, M. Ribiere-Chabert, and M-P. Chauzat. A pan-European
409 epidemiological study on honeybee colony losses 2012–2014. *Epilobee*, 2013:44, 2016.
- 410 M.I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion
411 for RNA-seq data with DESeq2. *Genome Biology*, 15:550, 2014.
- 412 E. Maori, N. Paldi, S. Shafir, H. Kaled, E. Tsur, E. Glick, and I. Sela. IAPV, a bee-affecting
413 virus associated with Colony Collapse Disorder can be silenced by dsRNA ingestion.
414 *Insect Mol Biol*, 18:55–60, 2009.
- 415 H.R. Mattila and T.D. Seeley. Genetic diversity in honey bee colonies enhances productivity
416 and fitness. *Science*, 317:362–364, 2007.
- 417 J.R. De Miranda, G. Cordoni, and G. Budge. The acute bee paralysis virus-Kashmir bee
418 virus-Israeli acute paralysis virus complex. *J Invertebr Pathol*, 103:S30–47, 2010.
- 419 D. Naug. Nutritional stress due to habitat loss may explain recent honeybee colony collapses.
420 *Biol Conserv*, 142:2369–2372, 2009.
- 421 P. Neumann and N.L. Carreck. Honey bee colony losses. *J Apicult Res*, 49:1–6, 2010.
- 422 R.E. Page and H.H. Laidlaw. Full sisters and supersisters: A terminological paradigm.
423 *Anim. Behav.*, 36:944–945, 1988.
- 424 G.D. Pasquale, M. Salignon, Y.L. Conte, L.P. Belzunces, A. Decourtey, A. Kretzschmar,
425 S. Suchail, J-L. Brunet, and C. Alaux. Influence of pollen nutrition on honey bee health:
426 Do pollen quality and diversity matter? *Plos One*, 8:e72016, 2013.
- 427 S.G. Potts, J.C. Biesmeijer, C. Kremen, P. Neumann, O. Schweiger, and W.E. Kunin. .
428 *Global pollinator declines: trends, impacts and drivers*, 25:345–353, 2010.
- 429 M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, and G.K. Smyth. limma
430 powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic
431 Acids Research*, 43(7):e47, 2015.
- 432 M.D. Robinson, D.J. McCarthy, and G.K. Smyth. edger: a bioconductor package for
433 differential expression analysis of digital gene expression data. *Bioinformatics*, 26:139–
434 140, 2010.
- 435 P. Rosenkranz, P. Aumeier, and B. Ziegelmann. Biology and control of Varroa destructor.
436 *J Invertebr Pathol*, 103:S96–S119, 2010.
- 437 T.H. Roulston and S.L. Buchmann. A phylogenetic reconsideration of the pollen starch-
438 pollination correlation. *Evol Ecol Res*, 2:627–643, 2000.
- 439 J.O. Schmidt. Feeding preference of *Apis mellifera* L. (Hymenoptera: Apidae): Individual
440 versus mixed pollen species. *J. Kans. Entomol. Soc.*, 57:323–327, 1984.

- 441 J.O. Schmidt, S.C. Thoenes, and M.D. Levin. Survival of honey bees, *Apis mellifera*
442 (Hymenoptera: Apidae), fed various pollen sources. *J. Econ. Entomol.*, 80:176–183, 1987.
- 443 M.Q. Shen, L.W. Cui, N. Ostiguy, and D. Cox-Foster. Intricate transmission routes and
444 interactions between picorna-like viruses (Kashmir bee virus and sacbrood virus) with
445 the honeybee host and the parasitic varroa mite. *J Gen Virol*, 86:2281–2289, 2005.
- 446 P.W. Sherman, T.D. Seeley, and H.K. Reeve. Parasites, pathogens, and polyandry in social
447 Hymenoptera. *Am. Nat*, 131:602–610, 1988.
- 448 M. Spivak, E. Mader, M. Vaughan, and N.H. Euliss. The Plight of the Bees. *Environ Sci
449 Technol*, 45:34–38, 2011.
- 450 R.G. Stanley and H.F. Linskens. *Pollen: Biology, biochemistry, management*. Springer
451 Verlag, 1974.
- 452 F. Supek, M. Bošnjak, N. Škunca, and T. Šmuc. REVIGO summarizes and visualizes long
453 lists of Gene Ontology terms. *Plos ONE*, 6:e21800, 2011.
- 454 D.R. Tarpy. Genetic diversity within honeybee colonies prevents severe infections and
455 promotes colony growth. *Proc. R. Soc. Lond. B*, 270:99–103, 2003.
- 456 D. van Engelsdorp, J.D. Evans, C. Saegerman, C. Mullin, E. Haubrige, B.K. Nguyen,
457 M. Frazier, J. Frazier, D. Cox-Foster, Y. Chen, R. Underwood, D.R. Tarpy, and J.S.
458 Pettis. Colony collapse disorder: A descriptive study. *PLoS One*, 4:e6481, 2009.
- 459 K.P. Weinberg and G. Madel. The influence of the mite *Varroa Jacobsoni Oud.* on the
460 protein concentration and the haemolymph volume of the brood of worker bees and
461 drones of the honey bee *Apis Mellifera L.* *Apidologie*, 16:421–436, 1985.
- 462 T.D. Wu, J. Reeder, M. Lawrence, G. Becker, and M.J. Brauer. GMAP and GSNAp
463 for genomic sequence alignment: Enhancements to speed, accuracy, and functionality.
464 *Methods Mol Biol*, 1418:283–334, 2016.
- 465 X. Yang and D. Cox-Foster. Effects of parasitization by *Varroa destructor* on survivorship
466 and physiological traits of *Apis mellifera* in correlation with viral incidence and microbial
467 challenge. *Parasitology*, 134:405–412, 2007.
- 468 X.L. Yang and D.L. Cox-Foster. Impact of an ectoparasite on the immunity and pathology
469 of an invertebrate: Evidence for host immunosuppression and viral amplification. *P Natl
470 Acad Sci USA*, 102:7470–7475, 2005.