

# **Visualization methods for genealogical and RNA-sequencing datasets**

Lindsay Rutter

Program of Study Committee:

Dianne Cook (Major Professor)

Amy Toth (Major Professor)

Heike Hofmann

Daniel Nettleton

James Reecy

May 1, 2018

---

# Contents

---

<b>Contents</b>	<b>2</b>
<b>1 Gene expression responses to diet quality and viral infection in <i>Apis mellifera</i></b>	<b>3</b>
1.1 Introduction . . . . .	3
1.2 Methods . . . . .	5
1.2.1 Design of two-factor experiment . . . . .	6
1.2.2 RNA extraction . . . . .	6
1.2.3 Gene expression . . . . .	7
1.2.4 Comparison to previous studies on transcriptomic response to viral infection . . . . .	7
1.2.5 Visualization . . . . .	8
1.2.6 Gene Ontology . . . . .	8
1.2.7 Probing tolerance versus resistance . . . . .	8
1.3 Results . . . . .	9
1.3.1 Phenotypic results . . . . .	9
1.3.2 Main effect DEG results . . . . .	11
1.3.3 Pairwise comparison of DEG results . . . . .	12
1.3.4 Comparison with Galbraith study . . . . .	13
1.3.5 Tolerance versus resistance . . . . .	16
1.4 Discussion . . . . .	18
<b>2 Supplementary material</b>	<b>21</b>
<b>Bibliography</b>	<b>42</b>

## *Chapter 1*

---

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

---

### **1.1 Introduction**

Commercially managed honey bees have undergone unusually large declines in the United States and parts of Europe over the past decade (van Engelsdorp et al. 2009, Kulhanek et al. 2017, Laurent et al. 2016), with annual mortality rates exceeding what beekeepers consider sustainable (Caron and Sagili 2011, Bond et al. 2014). 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 (Gallai et al. 2009). As honey bees are largely considered to be the leading pollinator of numerous crops, their marked loss has considerable implications regarding agricultural sustainability (Klein et al. 2007).

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

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

young workers to produce jelly, which then nourishes larvae, drones, older workers, and the queen (Crailsheim et al. 1992, Crailsheim 1992). 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 less diverse selections of pollen, which is of concern because mixed-pollen (polyfloral) diets are generally considered healthier than single-pollen (monofloral) diets (Schmidt 1984, Schmidt et al. 1987, Alaux et al. 2010). Indeed, reported colony mortality rates are higher in developed land areas compared to undeveloped land areas (Naug 2009), and beekeepers rank poor nutrition as one of the main reasons for colony losses (Engelsdorf et al. 2008). Understanding how undiversified diets affect honey bee health will be crucial to resolve problems that may arise as agriculture continues to intensify throughout the world (Neumann and Carreck 2010, Engelsdorf and Meixner 2010).

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

Although there is growing interest in how viruses and diet quality affect the health and sustainability of honey bees, as well as a recognition that such factors might operate interactively, there are only a small number of experimental studies thus far directed toward elucidating the interactive effects of these two factors in honey bees (DeGrandi-Hoffman and Chen 2015, DeGrandi-Hoffman et al. 2010, Conte et al. 2011). We recently used laboratory cages and nucleus hive experiments to investigate the health effects of these two factors, and our results show the importance of the combined effects of both diet quality and virus infection. Specifically, high quality pollen is able to mitigate virus-induced mortality to the level of diverse, polyfloral pollen (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 vigor (tolerance) ([Miller and Cotter 2017](#)). Transcriptomics is one means to better understand the mechanistic underpinnings of dietary and viral effects on honey bee health. Transcriptomic analysis can help us identify 1) the genomic scale of transcriptomic response to diet and virus infection, 2) whether these factors interact in an additive or synergistic way on transcriptome function, and 3) the types of pathways affected by diet quality and viral infection. This information, heretofore lacking in the literature, can help us better understand how good nutrition may be able to serve as a "buffer" against other stressors ([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 honey bees. As far as we know, there are few to no studies investigating honey bee gene expression patterns specifically related to monofloral diets, and few to no studies investigating honey bee gene expression patterns related to the combined effects of diet in any broad sense and viral inoculation in any broad sense.

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

We also compare the main effect of IAPV exposure in our dataset to that obtained in a previous study conducted by Galbraith and colleagues ([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 corroborate our data, and suggest such an approach can be useful for cross-study comparisons and validation of noisy RNA-sequencing data in the future.

## 1.2 Methods

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 ([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](#)).

### 1.2.1 Design of two-factor experiment

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

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

### 1.2.2 RNA extraction

Fifteen cages per treatment were originally sampled. Six live honey bees from each cage were randomly selected 36 hours post inoculation and placed into tubes ([Carrillo-Tripp et al. 2016](#)). Tubes were kept on dry ice and then transferred into a -80C freezer until processing. Eight cages were randomly selected from the original 15 cages, and 2 honey bees per cage were randomly selected from the original six live honey bees per cage. Whole body RNA from each pool of two honey bees were extracted using Qiagen RNeasy MiniKit followed by Qiagen DNase treatment. Samples were suspended in water to 200-400 ng/ $\mu$ l. All samples were then tested on a Bioanalyzer at the DNA core facility to ensure quality (RIN>8).

### 1.2.3 Gene expression

Samples were sequenced starting on January 14, 2016 at the Iowa State University DNA Facility (Platform: Illumina HiSeq Sequencing; Category: Single End 100 cycle sequencing). A standard Illumina mRNA library was prepared by the DNA facility. Reads were aligned to the BeeBase Version 3.2 genome ([Consortium 2014](#)) from the Hymenoptera Genome Database ([Elsik et al. 2016](#)) using the programs GMAP and GSNAp ([Wu et al. 2016](#)). There were four lanes of sequencing with 24 samples per lane. Each sample was run twice. Approximately 75-90% of reads were mapped to the honey bee genome. Each lane produced around 13 million single-end 100 basepair reads. We tested all six pairwise combinations of treatments for DEGs (pairwise DEGs). We also tested the diet main effect (diet DEGs), virus main effect (virus DEGs), and interaction term for DEGs (interaction DEGs). We then also tested for virus main effect DEGs (virus DEGs) in public data derived from a previous study exploring the gene expression of IAPV virus infection in honey bees ([Galbraith et al. 2015](#)). We tested each DEG analysis using recommended parameters with DESeq2 ([Love et al. 2014](#)), edgeR ([Robinson et al. 2010](#)), and LimmaVoom ([Ritchie et al. 2015](#)). In all cases, we used a false discovery rate (FDR) threshold of 0.05 ([Benjamini and Hochberg 1995](#)). Fisher's exact test was used to determine significant overlaps between DEG sets (whether from the same dataset but across different analysis pipelines or from different datasets across the same analysis pipelines). The eulerr shiny application was used to construct Venn diagram overlap images ([Larsson 2018](#)). In the main section of our paper and in subsequent analyses, we focus on the DEG results from DESeq2 ([Love et al. 2014](#)) as this pipeline was also used in the Galbraith study ([Galbraith et al. 2015](#)).

### 1.2.4 Comparison to previous studies on transcriptomic response to viral infection

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](#)) who also addressed honey bee transcriptomic responses to virus infection.

While our study examines honey bees from polyandrous colonies, the Galbraith study examined honey bees from single-drone colonies. As a consequence, the honey bees in our study will be on average 25% genetically identical, whereas honey bees from the Galbraith study will be on average 75% genetically identical ([Page and Laidlaw 1988](#)). We should therefore expect that the Galbraith study may generate data with lower signal:to:noise ratios than our data due to the lower genetic variation between its replicates. At the same time, our honey bees will be more likely to display the health benefits gained from increased genotypic variance within colonies, including decreased parasitic load ([Sherman et al. 1988](#)), increased tolerance to environmental changes ([Crozier and Page 1985](#)), and increasead colony performance ([Mattila and Seeley 2007](#), [Tarpy 2003](#)). Given that honey

bees are naturally very polyandrous ([Brodschneider et al. 2012](#)), our honey bees may also reflect more realistic environmental and genetic simulations. Taken together, each study provides a different point of value: Our study likely presents less artificial data while the Galbraith data likely presents less messy data. We wish to explore how the gene expression effects of IAPV inoculation compare between these two studies that used such different experimental designs. To achieve this objective, we use visualization techniques to assess the signal:to:noise ratio between these two datasets, and differential gene expression (DEG) analyses to determine any significantly overlapping genes of interest between these two datasets. It is our hope that this aspect of our study may shine light on how experimental designs that control genetic variability to different extents might affect the resulting gene expression data in honey bees.

### 1.2.5 Visualization

We used an array visualization tools as part of our analysis. We first used popular tools (like the MDS plot) from the DESeq2 package. After that, we used multivariate visualization tools from our work-in-progress package called bigPint. Specifically, we used parallel coordinate plots, litre plots, and scatterplot matrices to assess the variability between the replicates and the treatments in our data and in the DEG outputs from applied models. We also used these plotting techniques to assess for normalization problems and other common problems in RNA-seq analysis pipelines.

### 1.2.6 Gene Ontology

DEGs were uploaded as a background list to DAVID Bioinformatics Resources 6.7 ([Huang et al. 2009a](#), [Huang et al. 2009b](#)). The overrepresented gene ontology (GO) terms of DEGs were determined using the BEEBASE\_ID identifier option (honey bee gene model) in the DAVID software. To fine-tune the GO term list, only terms correlating to Biological Processes were considered. The refined GO term list was then imported into REVIGO ([Supek et al. 2011](#)), which uses semantic similarity measures to cluster long lists of GO terms.

### 1.2.7 Probing tolerance versus resistance

To investigate whether the protective effect of good diet is due to direct, specific effects on immune function (resistance), or if it is due to indirect effects of good nutrition on energy availability and vigor (tolerance), we created contrasts of interest (Table 1.1). In particular, we assigned “resistance candidate genes” to be the ones that were upregulated in the Chestnut group within the virus infected bees but not upregulated in the Chestnut group within the non-infected bees. We also assigned “tolerance candidate genes” to be the ones that were upregulated in the Chestnut group for both the virus infected bees and non-infected bees. Our interpretation of these genes is that they represent genes that

are constitutively activated in bees fed a high quality diet, regardless of whether they are experiencing infection or not. We then determined how many genes fell into these two categories and analyzed their GO terminologies.

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 1.2
NC vs NR	941	Genes that change expression due to diet effect in uninfected bees	Supplementary tables 2.4 and 2.5
VC vs VR	376	Genes that change expression due to diet effect in infected bees	Supplementary tables 2.6 and 2.7
VC upregulated in VC vs VR overlapped with NC upregulated in NC vs NR	122	“Tolerance” genes that are turned on by good diet regardless of virus infection status in bees	Figure 1.5A
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 1.5B

Table 1.1: Contrasts in our study for assessing GO and pathways analysis.

## 1.3 Results

### 1.3.1 Phenotypic results

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

Mortality rates of honey bees 72 hour post-inoculation significantly differed among the treatment groups (mixed model ANOVA across all treatment groups,  $df=3, 54; F=10.03$ ;  $p<2.30e-05$ ). The effect of virus treatment (mixed model ANOVA,  $df=1, 54; F=24.73$ ;  $p<1.00e-05$ ) and diet treatment (mixed model ANOVA,  $df=1, 54; F=5.32$ ;  $p<2.49e-02$ ) were significant, but the interaction between the two factors (mixed model ANOVA,  $df=1, 54; F=4.72e-02$ ,  $p=8.29e-01$ ) was not significant. We compared mortality levels based on pairwise comparisons: For a given diet, honey bees exposed to the virus showed significantly higher mortality rate than honey bees not exposed to the virus. Namely, bees fed Rockrose pollen had significantly elevated mortality with virus infection compared to uninfected controls (Tukey HSD,  $p<1.18e-03$ ), and bees fed Chestnut pollen similarly had significantly elevated mortality with virus infection compared to controls (Tukey HSD,  $p<4.80e-03$ )

(Figure 1.1).

IAPV titers of honey bees 72 hour post-inoculation significantly differed among the treatment groups (mixed model ANOVA across all treatment groups,  $df=3, 33; F=6.10; p<1.96e-03$ ). The effect of virus treatment (mixed model ANOVA,  $df=1, 33; F=15.04; p<4.75e-04$ ) was significant, but the diet treatment (mixed model ANOVA,  $df=1, 33; F=2.55; p=1.20e-01$ ) and the interaction between the two factors (mixed model ANOVA,  $df=1, 33; F=7.02e-01, p=4.08e-01$ ) were not significant. We compared IAPV titer volumes based on pairwise comparisons: Bees fed Rockrose pollen had significantly elevated IAPV titer volumes with virus infection compared to uninfected controls (Tukey HSD,  $p<5.44e-03$ ). However, bees fed Chestnut pollen did not have significantly elevated IAPV titer volumes with virus infection compared to uninfected controls (Tukey HSD,  $p=1.11e-01$ ). Overall, we interpreted these findings to mean that high-quality Chestnut pollen could “rescue” high virus titers resulting from the inoculation treatment, whereas low-quality Rockrose pollen could not do so (Figure 1.2).

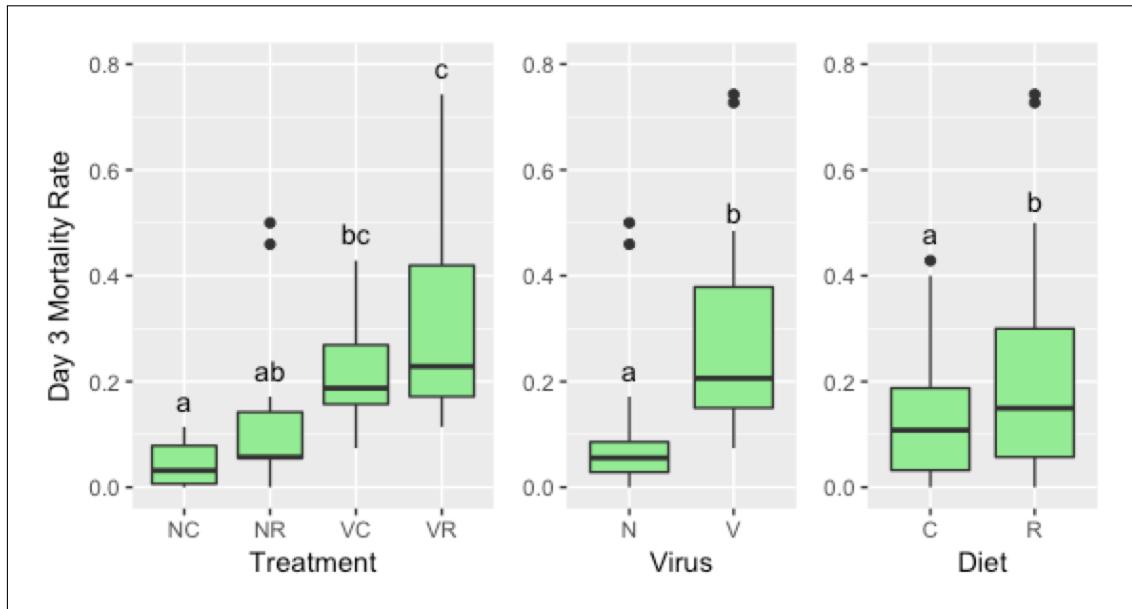


Figure 1.1: Left to right: Mortality rates for the four treatment groups, two virus groups, and two diet groups. “N” represents non-inoculation, “V” represents viral inoculation, “C” represents Chestnut pollen, and “R” represents Rockrose pollen. The mortality rate data included 59 samples with 15 replicates per treatment group, except for the “NC” group having 14 replicates. ANOVA values and p-values for the statistical tests are listed in the text of the paper. The letters above the bars represent Tukey honest significant differences with a confidence level of 95%.

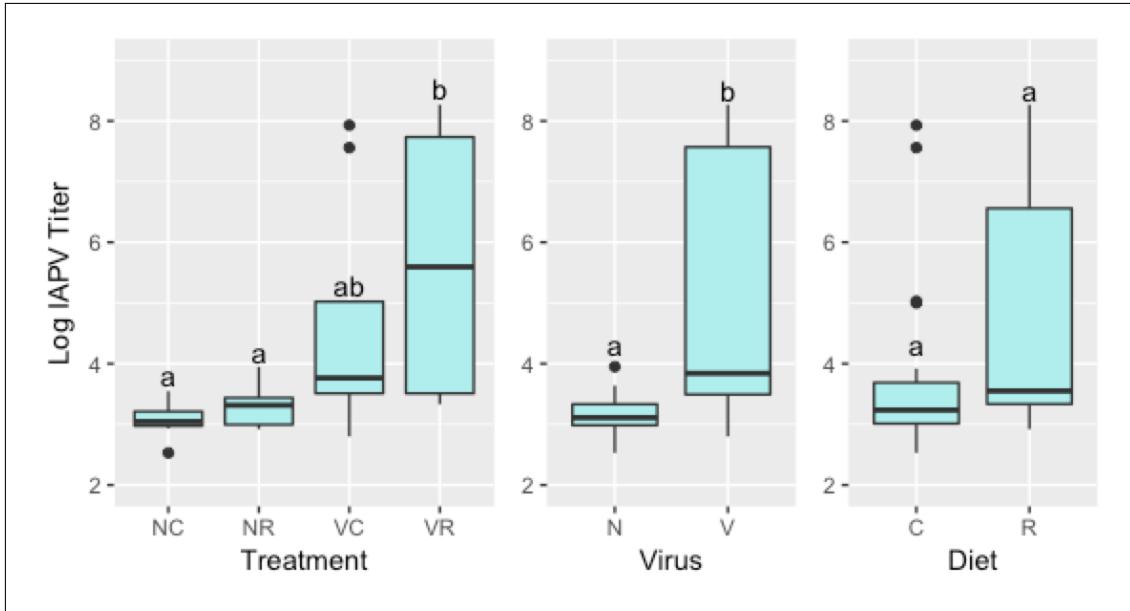


Figure 1.2: Left to right: IAPV titer volumes for the four treatment groups, two virus groups, and two diet groups. “N” represents non-inoculations, “V” represents viral inoculation, “C” represents Chestnut pollen, and “R” represents Rockrose pollen. The IAPV titer data included 38 samples with 10 replicates per treatment group, except for the “NR” group having 8 replicates. ANOVA values and p-values for the statistical tests are listed in the text of the paper. The letters above the bars represent Tukey honest significant differences with a confidence level of 95%.

### 1.3.2 Main effect DEG results

We observed a substantially larger number of DEGs in our diet main effect ( $n = 1914$ ) than in our virus main effect ( $n = 43$ ) (Supplementary table 2.1A and B). In the diet factor, there were more Chestnut-upregulated DEGs ( $n = 1033$ ) than Rockrose-upregulated DEGs ( $n = 881$ ). In the virus factor, there were more virus-upregulated DEGs ( $n = 38$ ) than control-upregulated DEGs ( $n = 5$ ). While these reported DEGs numbers are from the DESeq2 package, we saw similar trends for the edgeR and limma package results (Supplementary table 2.1A and B).

GO analysis of the Chestnut-upregulated DEGs revealed the following enriched categories (Benjamini correction  $< 0.05$ ): Wnt signaling, hippo signaling, and dorso-ventral axis formation, as well as pathways related to circadian rhythm, mRNA surveillance, insulin resistance, inositol phosphate metabolism, FoxO signaling, ECM-receptor interaction, phototransduction, Notch signaling, Jak-STAT signaling, MAPK signaling, and carbon metabolism (Supplementary table 2.2). GO analysis of the Rockrose DEGs revealed pathways related to terpenoid backbone biosynthesis, homologous recombination, SNARE interactions in vesicular transport, aminoacyl-tRNA biosynthesis, Fanconi anemia, and pyrimidine metabolism (Supplementary table 2.3).

With so few DEGs ( $n = 43$ ) in our virus main effect comparison, we focused on individual genes and their known functionalities rather than GO enrichment (Table 1.2). Of the 43 virus-related DEGs, only 10 had GO assignments within the DAVID database. These genes had putative roles in the recognition of pathogen-related lipid products and the cleaving of transcripts from viruses, as well as involvement in ubiquitin and proteosome pathways, transcription pathways, apoptotic pathways, oxidoreductase processes, and several more functions (Table 1.2).

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 D. melanogaster</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 1.2: 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.

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

### 1.3.3 Pairwise comparison of DEG results

The number of DEGs across the six treatment pairings between the diet and virus factor ranged from 0 to 941 (Supplementary table 2.8). Some of the trends observed in the main effect comparisons persisted: The diet level appeared to have greater influence on the number of DEGs than the virus level. Across every pair comparing the Chestnut and Rockrose levels, regardless of the virus level, the number of Chestnut-upregulated DEGs was higher than the number of Rockrose-upregulated DEGs (Supplementary table 2.8 C, D, E, F). For the pairs in which the diet level was controlled, the virus-exposed treatment showed equal to or more DEGs than the control treatment (Supplementary table 2.8 A, B). There were no DEGs between the treatment pair controlling for the control level of the virus effect (Supplementary table 2.8 A). These trends were observed for all three pipelines used (DESeq2, edgeR, and limma).

### 1.3.4 Comparison with Galbraith study

We wished to explore the signal:to:noise ratio between the Galbraith dataset and our dataset. Basic MDS plots were constructed with the DESeq2 analysis pipeline, and we could immediately determine that the Galbraith dataset may better separate the infected and uninfected honey bees better than our dataset (Supplementary figure 2.1). We also noted that the first replicate of both treatment groups in the Galbraith data did not cluster as cleanly in the MDS plots. However, through this automatically-generated plot, we can only visualize information at the sample level. Wanting to learn more about the data at the gene level, we continued with additional visualization techniques.

We used parallel coordinate lines superimposed onto boxplots to visualize the DEGs associated with virus infection in the two studies. The background boxplot represents the distribution of all genes in the data, and each parallel coordinate line represents one DEG. To reduce overlapping of parallel coordinate lines, we often use hierarchical clustering techniques to separate DEGs into common patterns.

We see that the 1,019 DEGs from the Galbraith dataset form relatively clean-looking visual displays (Figure 1.3). We do see that the first replicate of the virus group appears somewhat inconsistent with the other virus replicates in Cluster 2, confirming that this trend in the data that we saw in the MDS plot carried through into the DEG results. In contrast, we see that the 43 virus-related DEGs from our dataset do not look as clean in their visual displays (Figure 1.4). The replicates appear somewhat inconsistent in their esimated expression levels and there is not always such a large difference between treatment groups. We see a similar finding when we also examine a larger subset of 1,914 diet-related DEGs from our study (Supplementary figure 2.2).

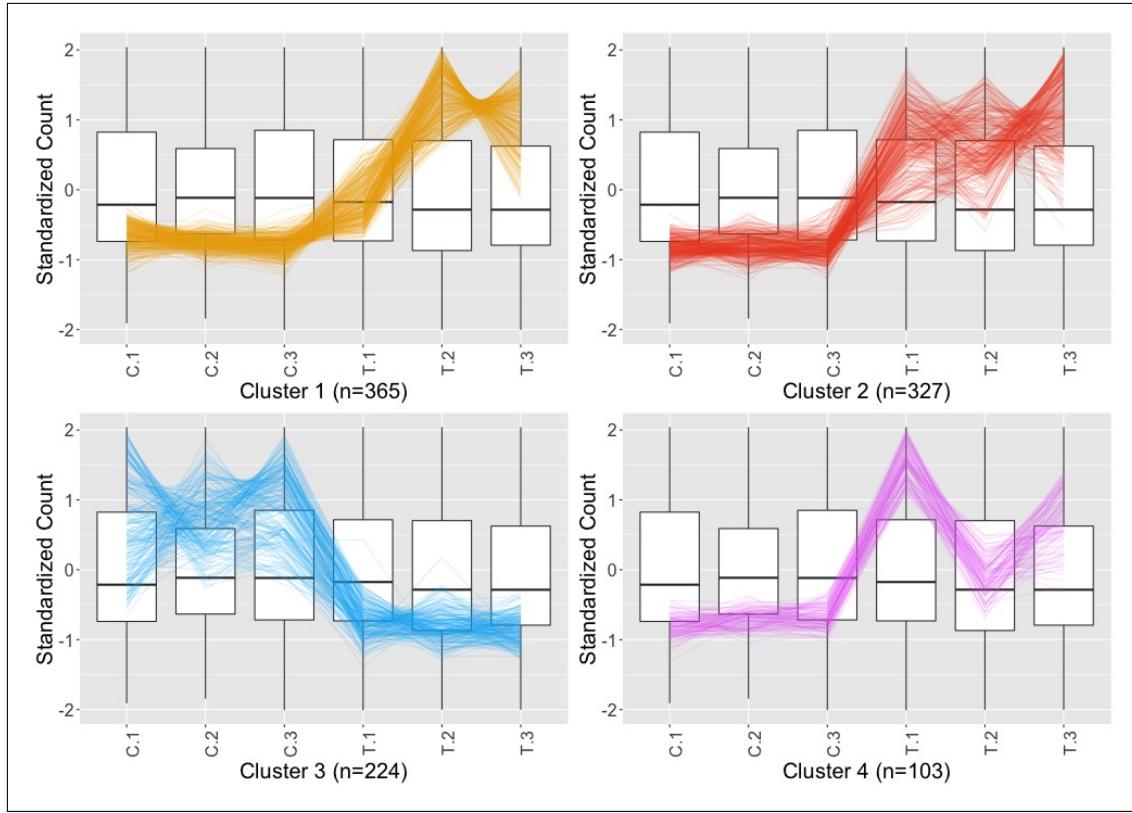


Figure 1.3: Parallel coordinate plots of the 1,019 DEGs after hierarchical clustering of size four between the virus-infected and control groups of the Galbraith study. Here “C” represents control, and “T” represents treatment of virus. Clusters 1, 3, and 4 seem to represent DEGs that were overexpressed in the virus inoculated group, and Cluster 2 seems to represent DEGs that were overexpressed in the control group. In general, the DEGs appeared as expected, but there is rather noticeable deviation of the first replicate from the virus-treated sample (“T.1”) from the other virus-treated replicates in Cluster 2. Cluster 4 also has some inconsistent replicates across the virus-treated replicates.

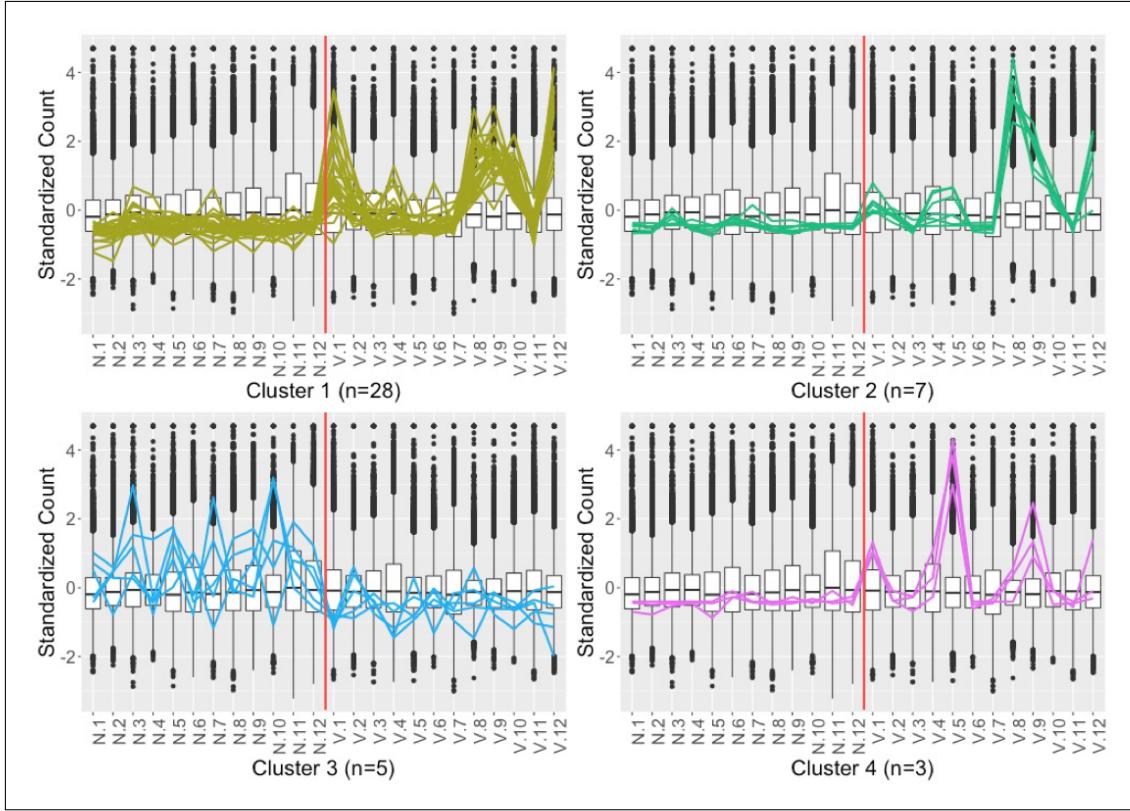


Figure 1.4: Parallel coordinate plots of the 43 DEGs after hierarchical clustering of size four between the virus-infected and control groups of our study. Here “N” represents non-infected control group, and “V” represents treatment of virus. The vertical red line indicates the distinction between treatment groups. We see from this plot that the DEG designations for this dataset do not appear as clean compared to what we saw in the Galbraith dataset in Figure 1.3.

We also used litre plots to examine the structure of individual DEGs: We see that indeed the individual virus DEGs from our data (Supplementary figure 2.3) show less consistent replications and less differences between the treatment groups compared to the individual virus DEGs from the Galbraith data (Supplementary figures 2.4 and 2.5). For the Galbraith data, we examined individual DEGs from the first cluster (Supplementary figure 2.4) and second cluster (Supplementary figure 2.5) because the second cluster had previously shown less consistency in the first replicate of the treatment group (figure 1.3). We verify this trend again in the litre plots with the DEG points in the second cluster showing less tight cluster patterns (Supplementary figures 2.4 and 2.5).

Finally, we looked at scatterplot matrices to assess the DEGs. We created standardized scatterplot matrices for each of the four clusters (Figure 1.3) of the Galbraith data (Supplementary figures 2.6, 2.7, 2.8, and 2.9). We also created standardized scatterplot matrices for our data. However, as our dataset contained 24 samples, we would need to include

276 scatterplots in our matrix, which would be too numerous to allow for efficient visual assessment of the data. As a result, we created four scatterplot matrices of our data, each with subsets of 6 samples to be more comparable to the Galbraith data (Supplementary figures 2.10, 2.11, 2.12, and 2.13). We can again confirm through these plots that the DEGs from the Galbraith data appeared more as expected: Deviating more from the  $x=y$  line in the treatment scatterplots while staying close to the  $x=y$  line in replicate scatterplots.

Despite the virus-related DEGs ( $n = 1,019$ ) from the Galbraith dataset displaying the expected patterns more than those from our dataset ( $n = 43$ ), there was significant overlap ( $p\text{-value} < 2.2\text{e-}16$ ) in the DEGs between the two studies (Supplementary figure 2.14).

### 1.3.5 Tolerance versus resistance

Using the contrasts specified in Table 1.1, we discovered 122 “tolerance” candidate genes and 125 “resistance” candidate genes. We again used parallel coordinate lines superimposed onto boxplots to visualize these candidate genes. To reduce overlapping of parallel coordinate lines, we again used hierarchical clustering techniques to separate DEGs into common patterns. Perhaps unsurprisingly, we still see a substantial amount of noise (inconsistency between replicates) in our resulting candidate genes (Figures 2.15 and 2.16). However, the broad patterns we expect to see still emerge: For example, based on the contrasts we created to obtain the ‘tolerance’ candidate genes, we expect them to display larger count values in the “NC” group compared to the “VC” group and larger count values in the “NR” group compared to the “VR” group. Indeed, we see this pattern in the associated parallel coordinate plots (Figure 2.15). Likewise, based on the contrasts we created to obtain the ‘resistance’ candidate genes, we still expect them to display larger count values in the “VC” group compared to the “VR” group, but we no longer expect to see a difference between the “NC” and “NR” groups. We do generally see these expected patterns in the associated parallel coordinate plots: While there are large outliers in the “NC” group, the “NR” replicates are no longer typically below a standardized count of zero (Figure 2.16). The genes in Cluster 3 in particular may follow the expected pattern the most distinctively (Figure 2.16).

Within our 122 “tolerance” gene ontologies, we found functions related to metabolism (such as carbohydrate metabolism, fructose metabolism, and chitin metabolism). However, we also discovered gene ontologies related to RNA polymerase II transcription, immune response, and regulation of response to reactive oxygen species (Figure 1.5A). Within our 125 “resistance” gene ontologies, we found functions related to metabolism (such as carbohydrate metabolism, chitin metabolism, oligosaccharide biosynthesis, and general metabolism) (Figure 1.5B).

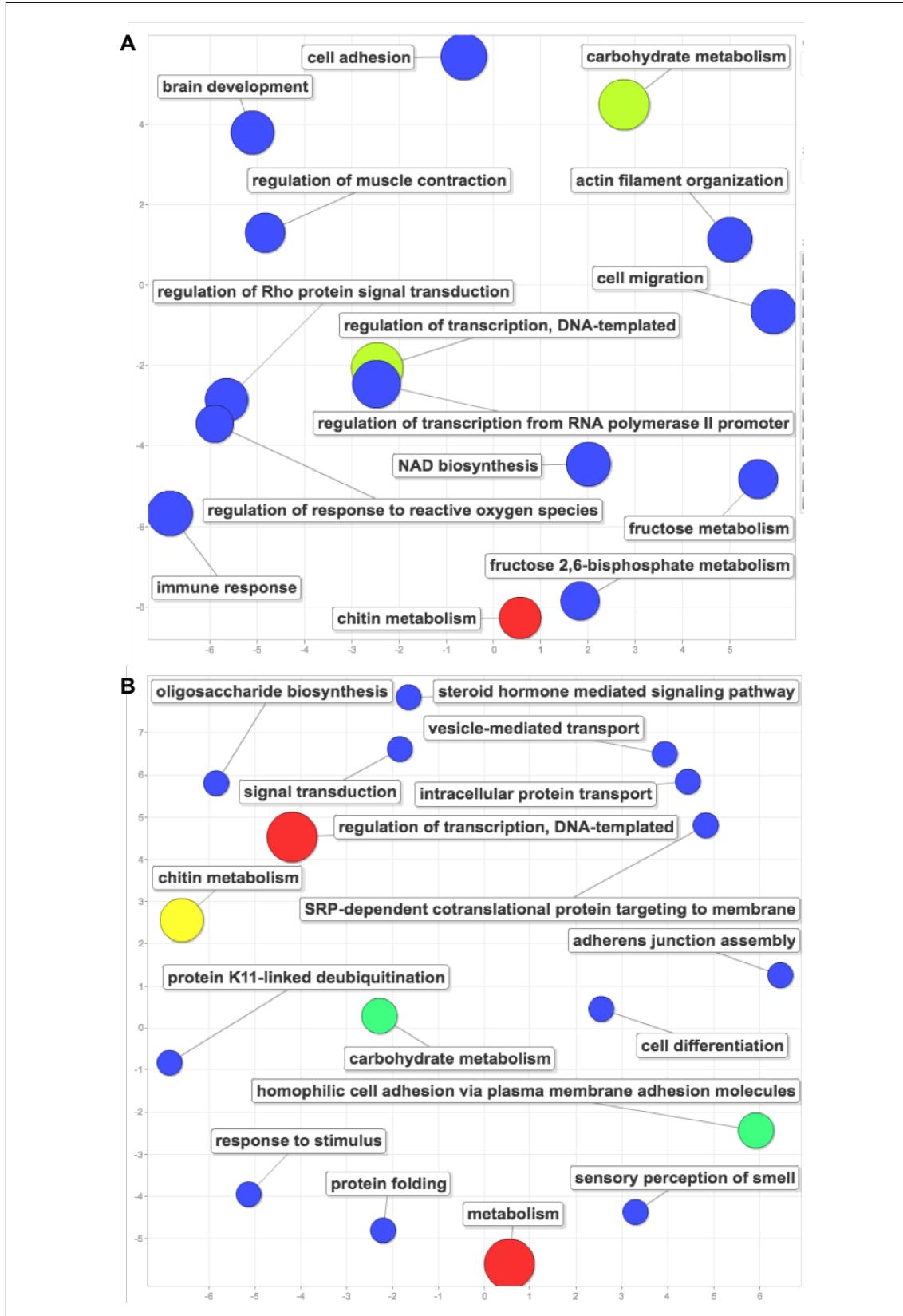


Figure 1.5: GO analysis results for the 122 DEGs related to our “tolerance” hypothesis (A) and for the 125 DEGs related to our “resistance” hypothesis (B).

## 1.4 Discussion

Challenges to honey bee health are a growing concern, in particular the combined, interactive effects of nutritional stress and pathogens (Dolezal and Toth 2018). In this study, we used RNA-sequencing to probe mechanisms underlying honey bee responses to two effects, diet quality and infection with the major virus of concern, IAPV. In general, we found a major nutritional transcriptomic response, with nearly 2,000 transcripts changing in response to diet quality (rockrose/poor diet versus chestnut/good diet). The majority of these genes were upregulated in response to high quality diet, and these genes were enriched for functions (Supplementary table 2.2) such as nutrient signaling (insulin resistance) metabolism, and immune response (Notch signaling and Jak-STAT pathways). These data suggest high quality nutrition may allow bees to alter their metabolism, favoring investment of energy into innate immune responses.

Somewhat surprisingly, the transcriptomic response to virus infection in our experiment was fairly limited. We found only 43 transcripts to be differentially expressed, some with known immune functions (Table 1.2) such as argonaute-2 and a gene with similarity to MD-2 lipid recognition protein, as well as additional genes related to transcriptional regulation, and muscle contraction. The small number of DEGs in this study may be partly explained by the large amount of noise in the data (Figure 1.4 and Supplementary figures 2.1B, 2.3, 2.10, 2.11, 2.12, and 2.13).

Given the noisy nature of our data, and our desire to hone in on genes with real expression differences, we compared our data to the Galbraith study (Galbraith et al. 2015), which also examined bees response to viral 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 1.3 and Supplementary figures 2.1A, 2.4, 2.5, 2.6, 2.7, 2.8, and 2.9). To identify the most reliable virus-responsive genes from our study, we looked for overlap in the DEGs associated with virus infection on both experiments. We found a large, statistically significant ( $p\text{-value} < 2.2\text{e-}16$ ) overlap, with 26/38 (68%) of virus-responsive DEGs from our study also showing response to virus infection in Galbraith et al. (Supplementary figure 2.14). This result gives us confidence that, although noisy, we were able to uncover consistent, replicable gene expression responses to virus infection with our data.

Data visualization is a useful method to identify noise and robustness in RNA-seq data. In this study, we used extensive data visualization to improve the interpretation of our RNA-seq results. For example, the DESeq2 package comes with certain visualization options that are popular in RNA-sequencing analysis. One of these visualization is the multidimensional scaling (MDS) 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 (Supplementary figure 2.1). However, the MDS 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, litre plots, and scatterplot matrices. To prevent overlapping issues in our plots, 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 the Galbraith dataset to make effective use of the scatterplot matrices. After these tailorizations, we determined that the same patterns we saw in the MDS 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 2 of Figure 1.3 showed an inconsistent first replicate in the treatment group (“T.1”), which was something we observed in the MDS plot. This indicates that this feature also extended down the analysis pipeline into DEG calls. We believe these visualization applications can be useful for future researchers analyzing RNA-sequencing data to quickly and effectively ensure that the DEG calls look reliable or at least overlap with DEG calls from similar studies that look reliable. We also expect this type of visualization exploration can be especially crucial when studying complex organisms that do not have genetic identicalness or similarity between replicates and/or when using experiments that may lack rigid design control.

One of the goals of this study was to use our RNA-seq data to assess whether transcriptomic responses to diet quality and virus infection provide insight into whether high quality diet can buffer bees from pathogen stress via mechanisms of “resistance” or “tolerance”. Recent evidence has suggested that overall immunity is determined by more than just “resistance” (the reduction of pathogen fitness within the host by mechanisms of avoidance and control) (Carval and Ferriere 2010). Instead, overall immunity is related to “resistance” in conjunction with “tolerance” (the reduction of adverse effects and disease resulting from pathogens by mechanisms of healing) (Miller and Cotter 2017, Carval and Ferriere 2010). Immune-mediated resistance and diet-driven tolerance mechanisms are costly and may compete with each other (Miller and Cotter 2017, Moret 2006). Data and models have suggested that selection can favor an optimum combination of both resistance and tolerance (Mauricio et al. 1997, Fornoni et al. 2004, Restif and Koella 2003, Chambers and Schneider 2012). We attempted to address this topic through specific gene expression contrasts (Table 1.1), accompanied by GO analysis of the associated gene lists. We found an approximately equal number of resistance ( $n = 125$ ) and tolerance ( $n = 122$ ) related candidate genes, suggesting both processes may be playing significant roles in dietary buffering from pathogen induced mortality. Resistance candidate genes had functions related

to several forms of metabolism (chitin and carbohydrate), regulation of transcription, and cell adhesion. Tolerance candidate genes had functions related to carbohydrate metabolism and chitin metabolism. However, they also showed functions related to immune response, including RNA polymerase II transcription and regulation of response to reactive oxygen species (Figure 1.5A). 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 (Xu et al. 2012). Moreover, circulating haemocytes in insects encapsulate and nodulate pathogens by forming a barrier between the pathogen and the host tissues. This barrier undergoes apoptosis and melanization through the phenoloxidase enzyme cascade, which produces reactive oxygen species (Miller and Cotter 2017, Cerenius and Söderhäll 2004, Sadd and Siva-Jothy 2006). These possible immunological defense mechanisms within our “tolerance” candidate genes and metabolic processes within our “resistance” candidate genes may provide additional evidence of feedbacks between diet and disease in honeybees (Dolezal and Toth 2018).

Overall, these data suggest complex transcriptomic responses to multiple stressors in honey bees. Diet has the potential for large and profound effects on transcriptional responses in honey bees, and differences in diet may set up the potential for both resistance and tolerance to virus infection. Moreover, this study in general also demonstrated the possible 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 but in a more controlled manner, we were able to place much higher confidence in the differential gene expression results from our otherwise noisy data. We hope these results underline the need for researchers to use data visualization techniques to understand and interpret RNA-sequencing datasets.

## *Chapter 2*

---

# Supplementary material

---

This section contains supplementary analyses that the reader might find useful, and that extend upon several of the main concepts discussed in the paper.

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 2.1: 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. For the diet effects, “C” represents Chestnut diet and “R” represents Rockrose diet. For the virus effects, “V” represents virus-innocalated and “C” represents control non-innocalated. Green cells represent the level that showed a larger number of 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 2.2: Pathways related to diet main effect Chestnut-upregulated 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 2.3: Pathways related to diet main effect Rockrose-upregulated DEGs.

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

Table 2.4: GO analysis results for the 601 DEGs that were upregulated in the NC treatment in the NC versus NR treatment pair analysis. These DEGs represent genes that are upregulated when non-infected honey bees are given high quality Chestnut pollen compared to being given low quality Rockrose pollen.

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

Table 2.5: GO analysis results for the 340 DEGs that were upregulated in the NR treatment in the NC versus NR treatment pair analysis. These DEGs represent genes that are upregulated when non-infected honey bees are given low quality Rockrose pollen compared to being given high quality Chestnut pollen.

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

Table 2.6: GO analysis results for the 247 DEGs that were upregulated in the VC treatment in the VC versus VR treatment pair analysis. These DEGs represent genes that are upregulated when infected honey bees are given high quality Chestnut pollen compared to being given low quality Rockrose pollen.

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

Table 2.7: GO analysis results for the 129 DEGs that were upregulated in the VR treatment in the VC versus VR treatment pair analysis. These DEGs represent genes that are upregulated when infected honey bees are given low quality Rockrose pollen compared to being given high quality Chestnut pollen.

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 2.8: 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-innocalated, and “N” represents control non-innocalated. Green cells represent the level that showed a larger number of DEGs.

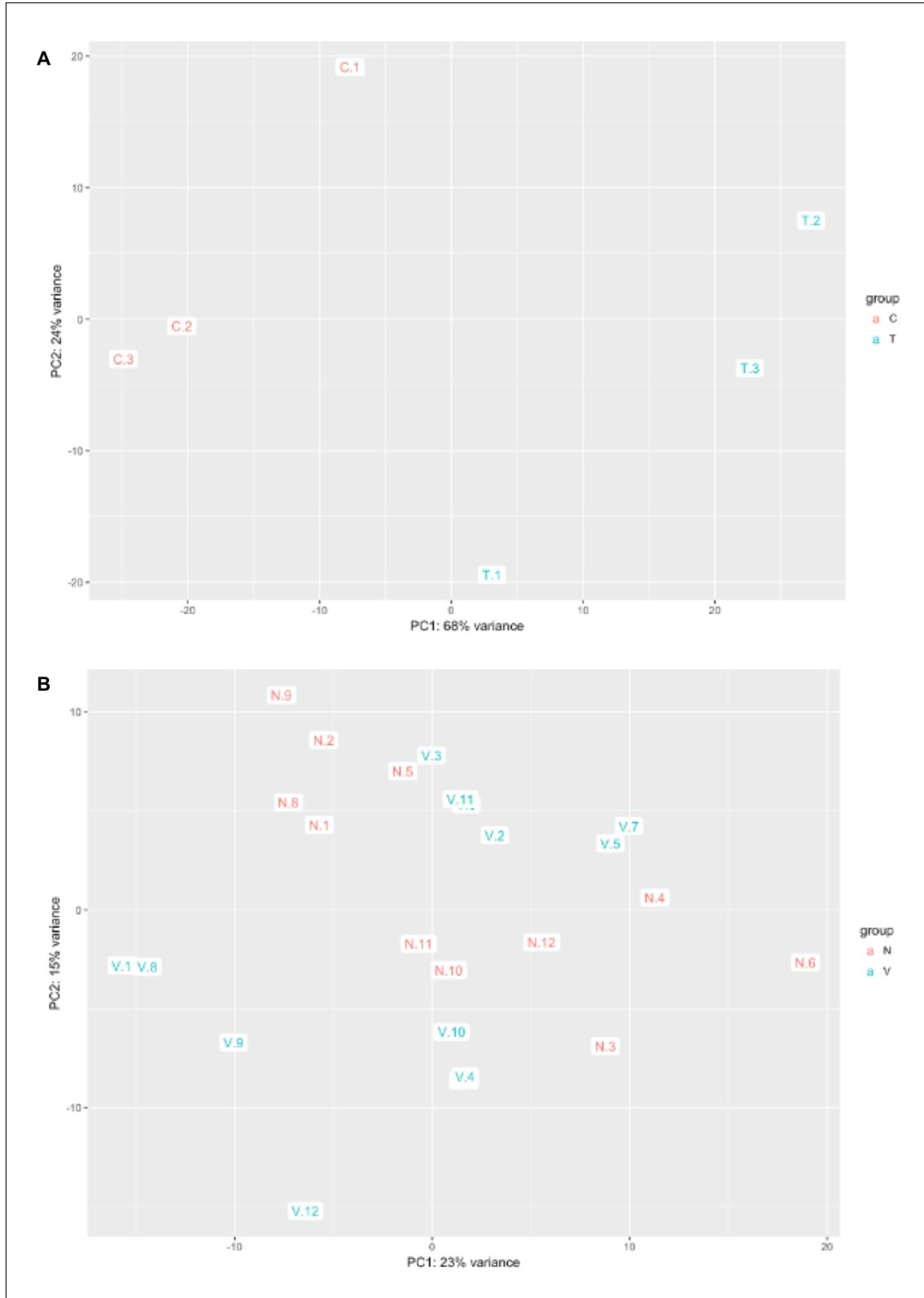


Figure 2.1: MDS plots constructed from DESeq2 package for the Galbraith dataset for non-infected control “C” and virus treated “T” samples (A) and our dataset for the non-infected control “N” and virus treated “V” samples (B). the x-axis represents the principal component with the most variation and the y-axis repesents the principal component with the second-most variation.

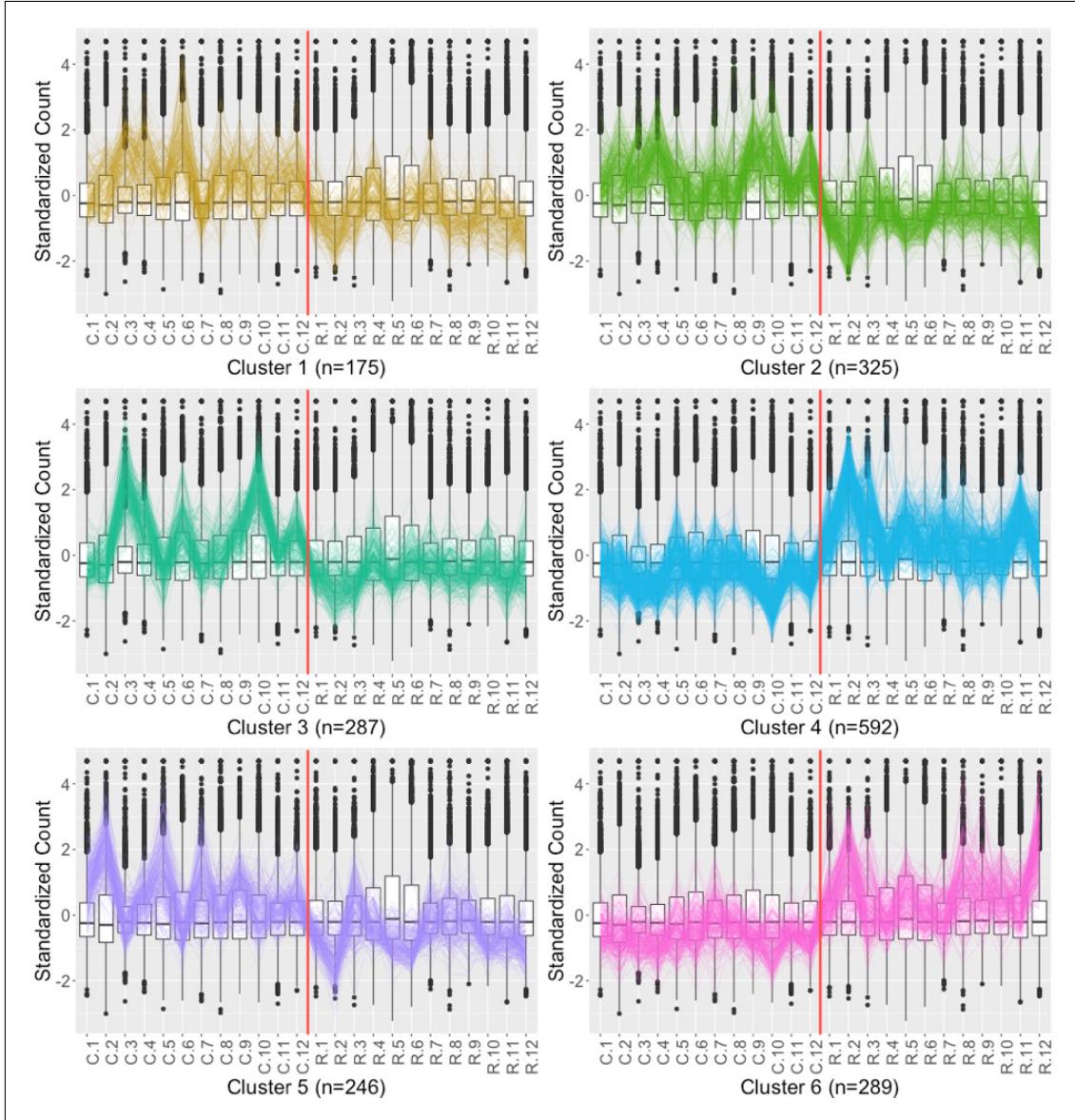


Figure 2.2: Parallel coordinate plots of the 1,914 DEGs after hierarchical clustering of size six between the Chestnut and Rockrose groups of our study. Here “N” represents non-infected control group, and “V” represents treatment of virus. The vertical red line indicates the distinction between treatment groups. We see from this plot that the DEG designations for this dataset do not appear as clean compared to what we saw in the Galbraith dataset in Figure 1.3.

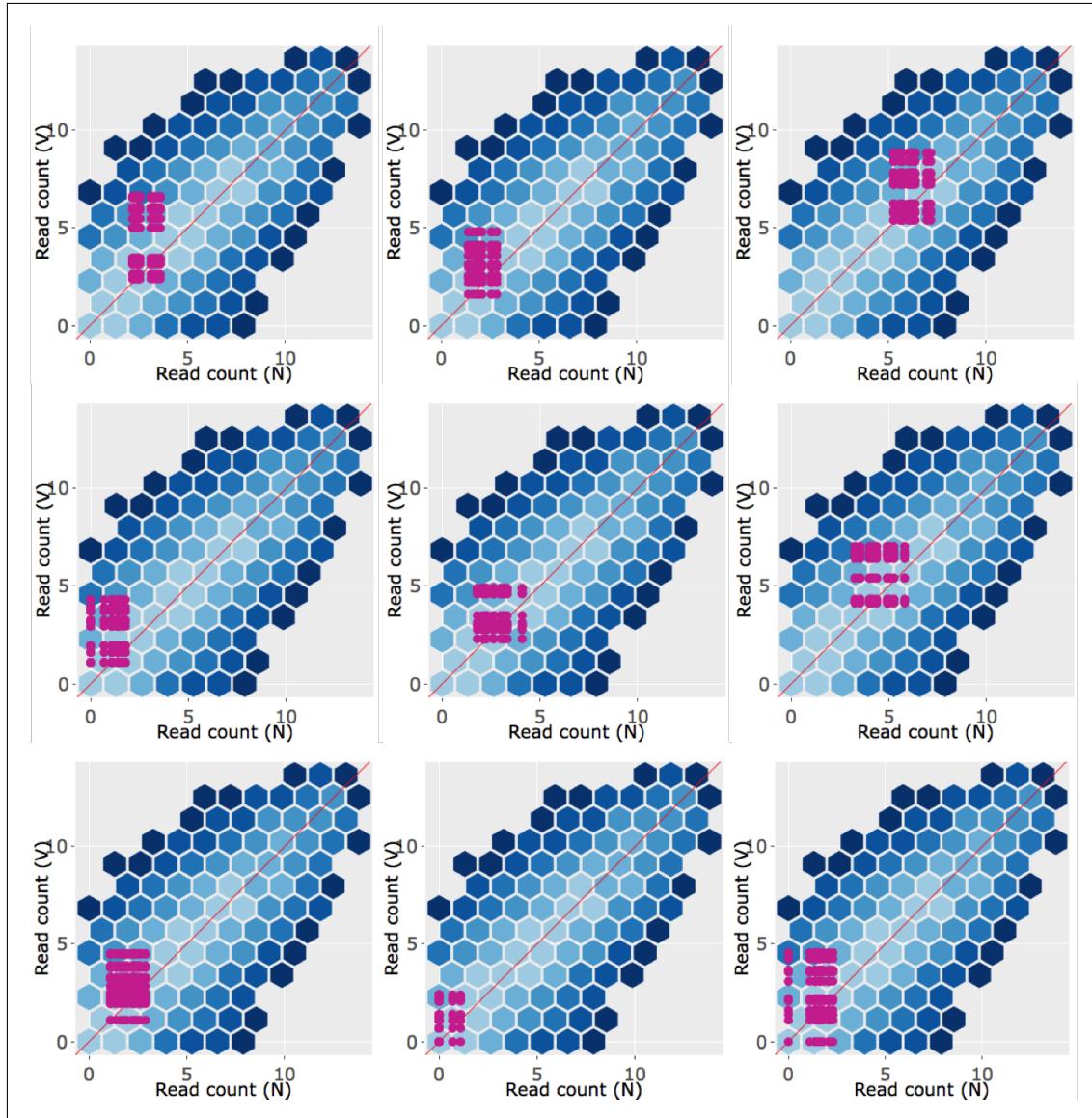


Figure 2.3: Example litre plots of the nine DEGs with the lowest FDR values from the 43 DEGs of our dataset. “N” represents non-infected control samples and “V” represents virus-treated samples. Most of the magenta points (representing the 144 combinations of samples between treatment groups for a given DEG) do not reflect the expected pattern as clearly compared to what we saw in the litre plots of the Galbraith data. They are not as clustered together (representing replicate inconsistency) and they sometimes overlap 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 1.3.

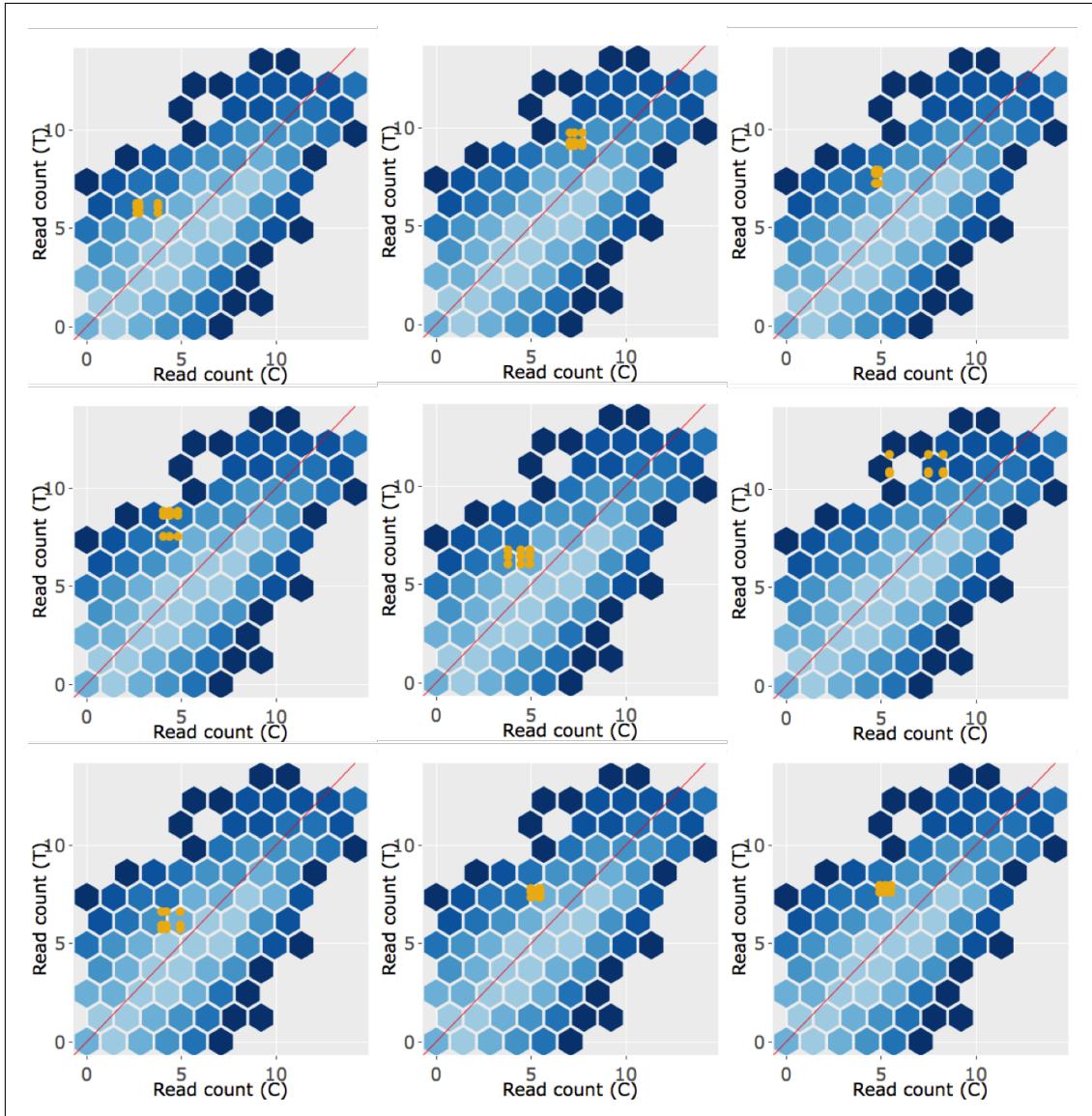


Figure 2.4: Example litre plots of the nine DEGs with the lowest FDR values from the 365 DEGs in Cluster 1 (originally shown in Figure 1.3) of the Galbraith dataset. “C” represents non-infected control samples and “T” represents virus-treated samples. Most of the light orange points (representing the nine combinations of samples between treatment groups for a given DEG) deviate from the  $x=y$  line in a cluster as expected.

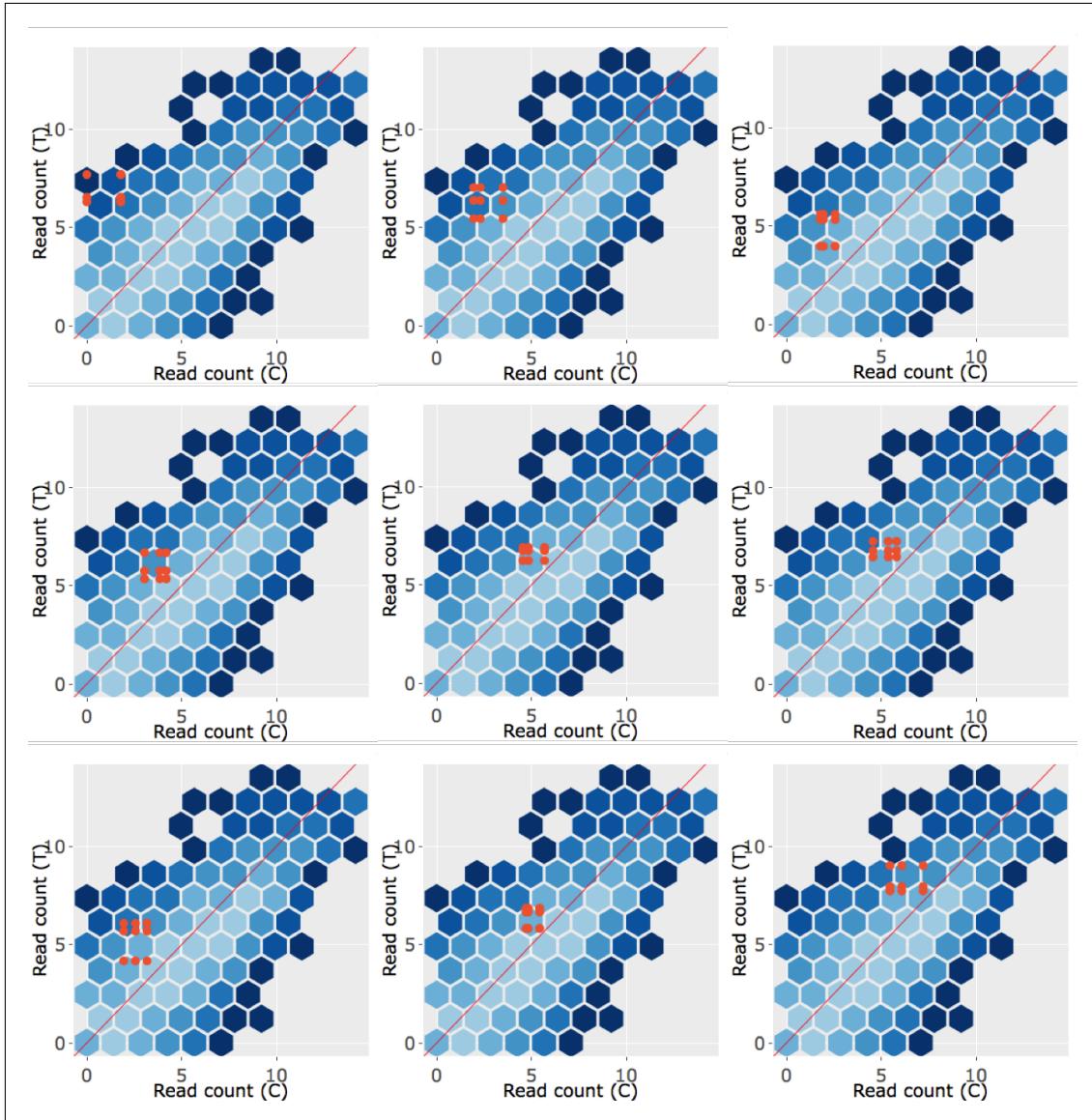


Figure 2.5: Example litre plots of the nine DEGs with the lowest FDR values from the 327 DEGs in Cluster 2 (originally shown in Figure 1.3) of the Galbraith dataset. “C” represents non-infected control samples and “T” represents virus-treated samples. Most of the dark orange points (representing all combinations of samples between treatment groups for a given DEG) deviate from the  $x=y$  line in a cluster as expected. However, they are not as tightly clustered together compared to what we saw in the example litre plots of Cluster 1 (shown in Supplementary figure 2.4). As a result, what we see in these litre plots reflects what we saw in the parallel coordinate lines of Figure 1.3: The replicate consistency in the Cluster 2 DEGs is not as clean as that in the Cluster 1 DEGs, but is still relatively clean.

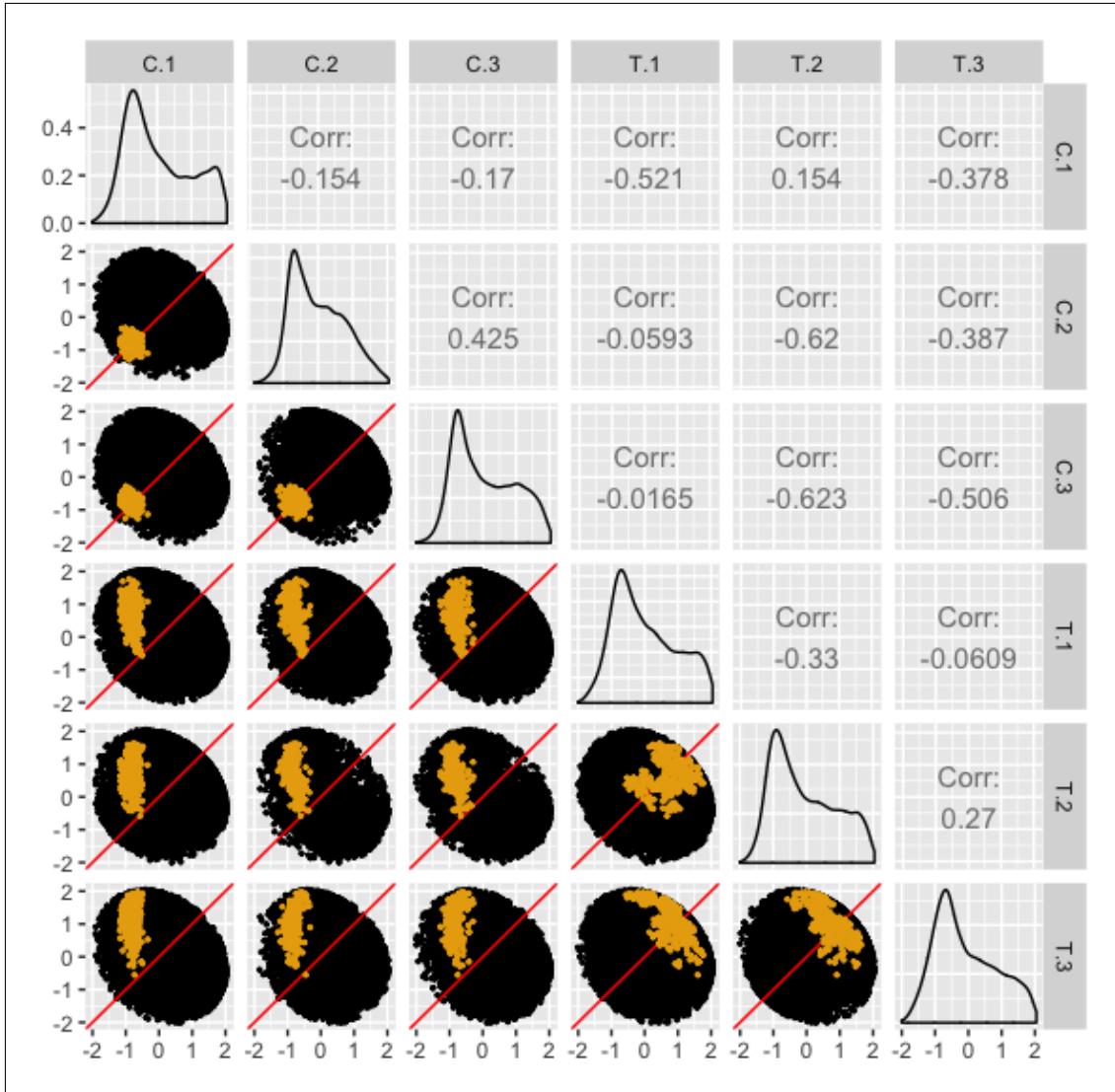


Figure 2.6: The 365 DEGs from the first cluster of the Galbraith dataset (shown in Figure 1.3) superimposed as light orange dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. “C” represents non-infected control samples and “T” represents virus-treated samples. We confirm that the DEGs mostly follow the expected structure, with their placement deviating from the  $x=y$  line in the treatment scatterplots, but adhering to the  $x=y$  line in the replicate scatterplots.

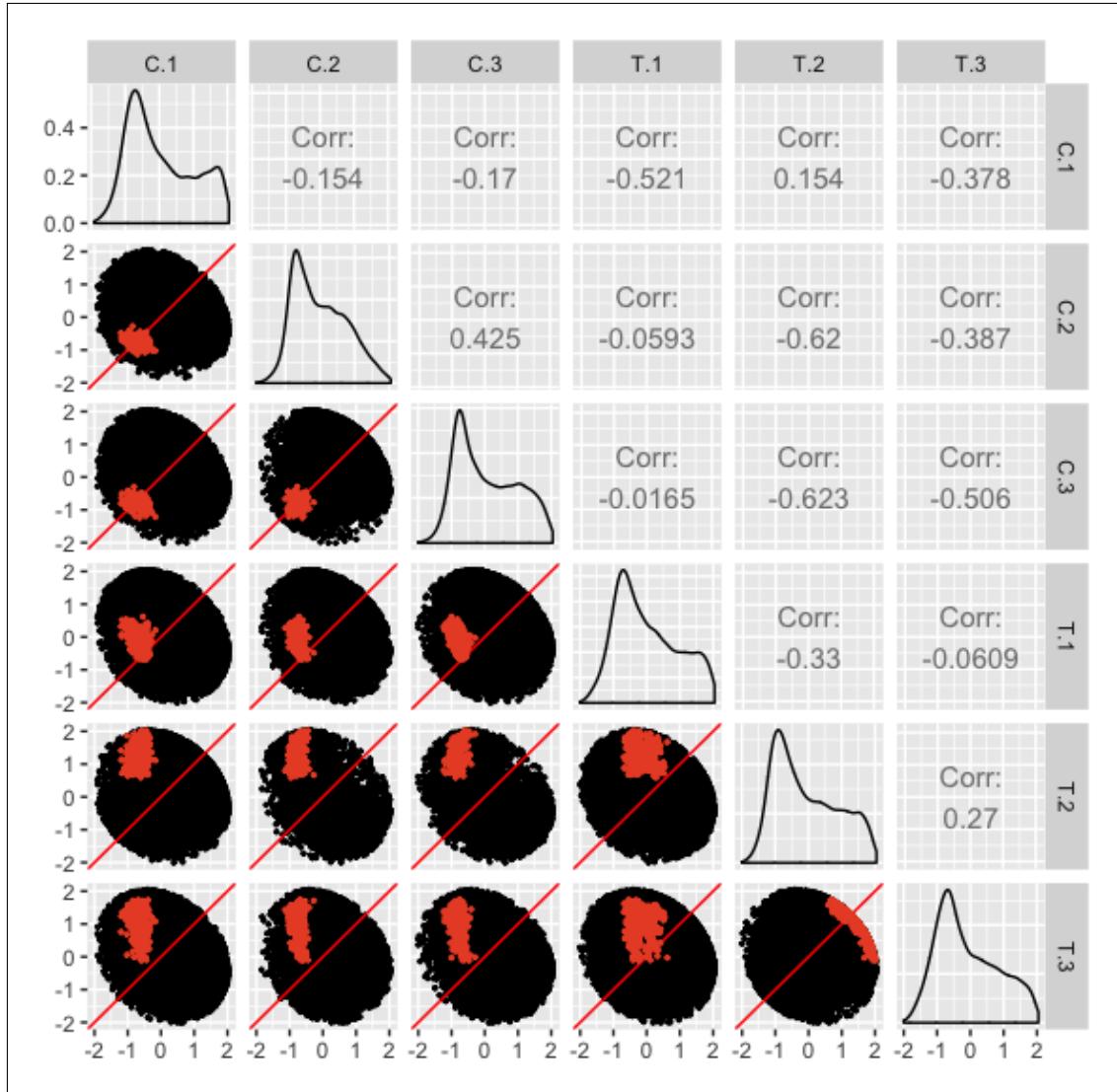


Figure 2.7: The 327 DEGs from the second cluster of the Galbraith dataset (shown in Figure 1.3) superimposed as dark orange dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. “C” represents non-infected control samples and “T” represents virus-treated samples. We confirm that the DEGs mostly follow the expected structure, with their placement deviating from the  $x=y$  line in the treatment scatterplots, but adhering to the  $x=y$  line in the replicate scatterplots. We also see again that the first replicate from the virus-treated sample (“T.1”) may be somewhat inconsistent in these DEGs, as its presence in the replicate scatterplots results in the DEGs unexpectedly deviating from the  $x=y$  line and its presence in the treatment scatterplots results in the DEGs unexpectedly adhering to the  $x=y$  line.

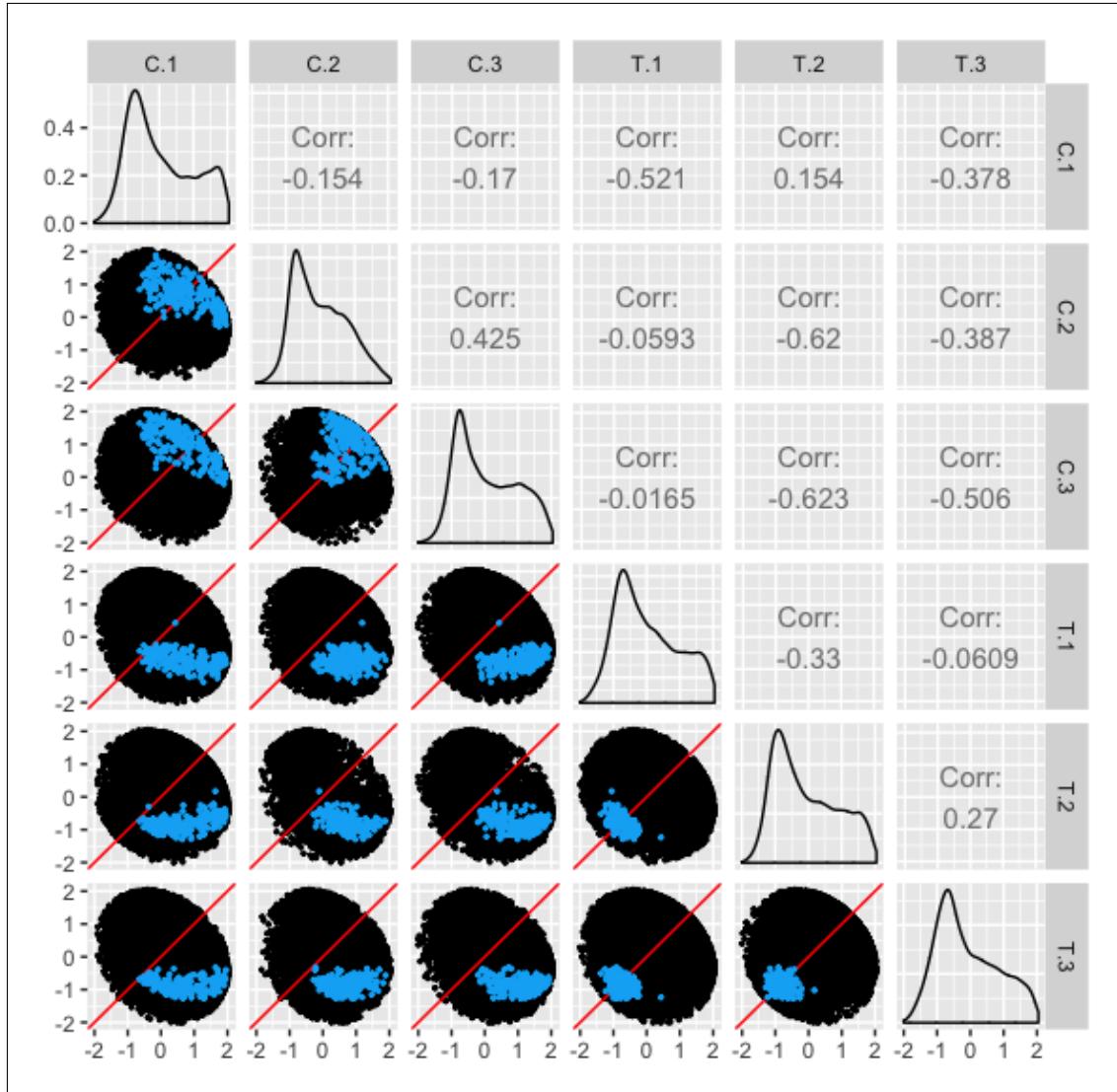


Figure 2.8: The 224 DEGs from the third cluster of the Galbraith dataset (shown in Figure 1.3) superimposed as turquoise dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. “C” represents non-infected control samples and “T” represents virus-treated samples. We confirm that the DEGs mostly follow the expected structure, with their placement deviating from the  $x=y$  line in the treatment scatterplots, but adhering to the  $x=y$  line in the replicate scatterplots.

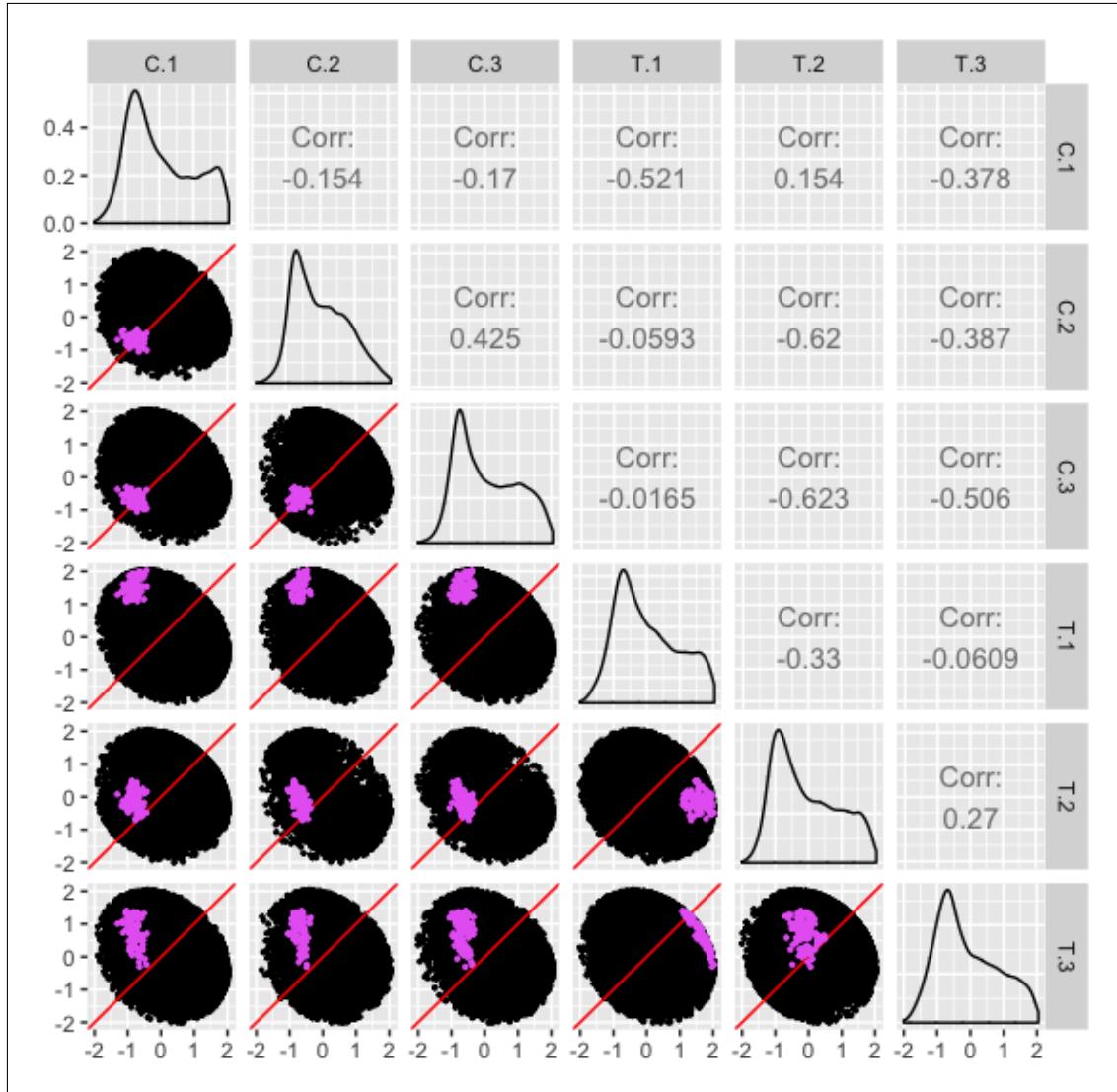


Figure 2.9: The 103 DEGs from the fourth cluster of the Galbraith dataset (shown in Figure 1.3) superimposed as pink dots onto all genes as black dots in the form of a scatterplot matrix. The data has been standardized. “C” represents non-infected control samples and “T” represents virus-treated samples. We confirm that the DEGs mostly follow the expected structure, with their placement deviating from the  $x=y$  line in the treatment scatterplots, but adhering to the  $x=y$  line in the replicate scatterplots. We also see that the second replicate from the virus-treated sample (“T.2”) may be somewhat inconsistent in these DEGs, as its presence in the replicate scatterplots results in the DEGs unexpectedly deviating from the  $x=y$  line and its presence in the treatment scatterplots results in the DEGs unexpectedly adhering to the  $x=y$  line.

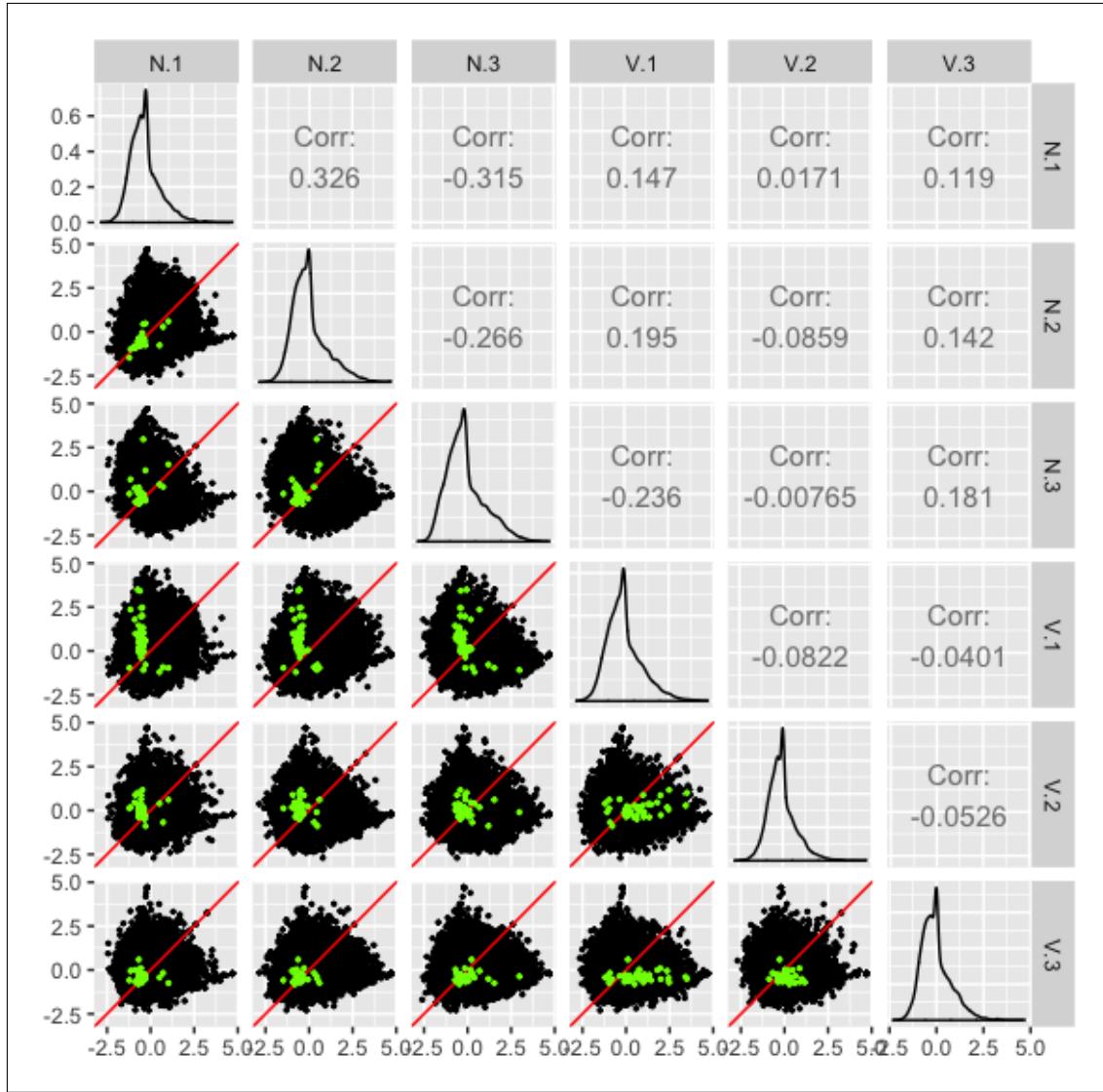


Figure 2.10: The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only replicates 1, 2, and 3 are shown from both treatment groups. The data has been standardized. “N” represents non-infected control samples and “V” represents virus-treated samples. We see that, compared to the scatterplot matrices from 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.

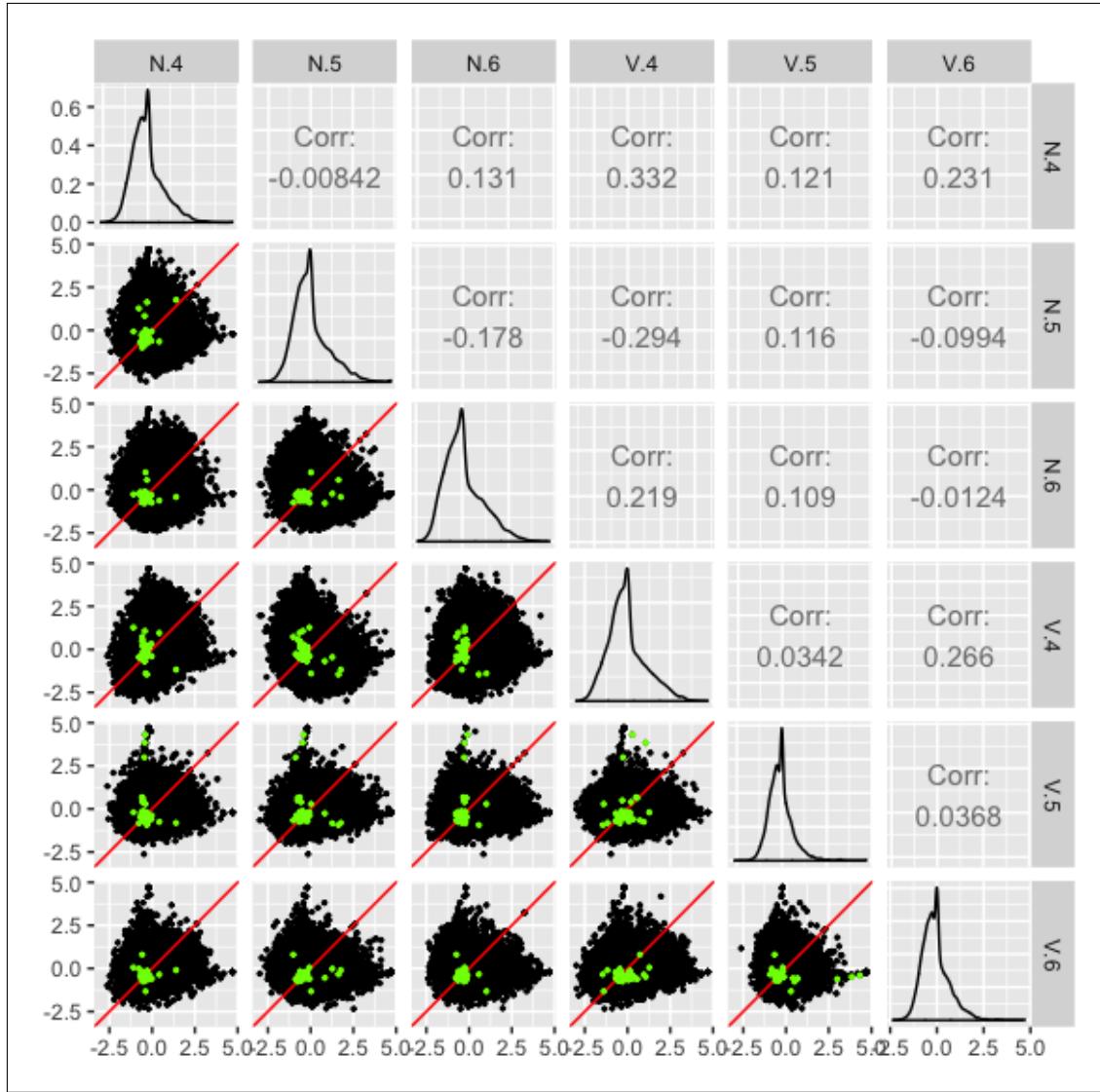


Figure 2.11: The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only replicates 4, 5, and 6 are shown from both treatment groups. The data has been standardized. “N” represents non-infected control samples and “V” represents virus-treated samples. We see that, compared to the scatterplot matrices from 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.

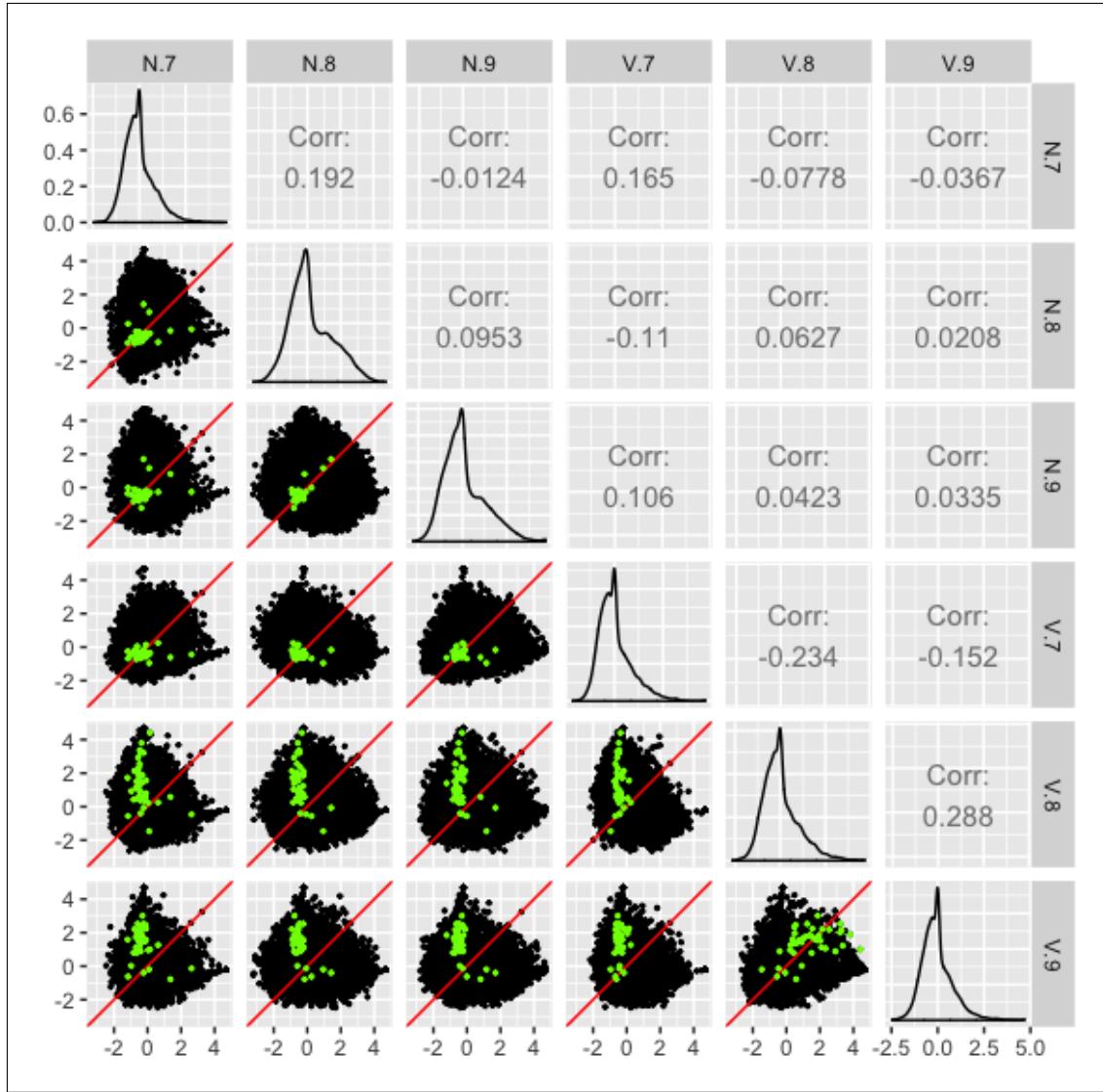


Figure 2.12: The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only replicates 7, 8, and 9 are shown from both treatment groups. The data has been standardized. “N” represents non-infected control samples and “V” represents virus-treated samples. We see that, compared to the scatterplot matrices from 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.

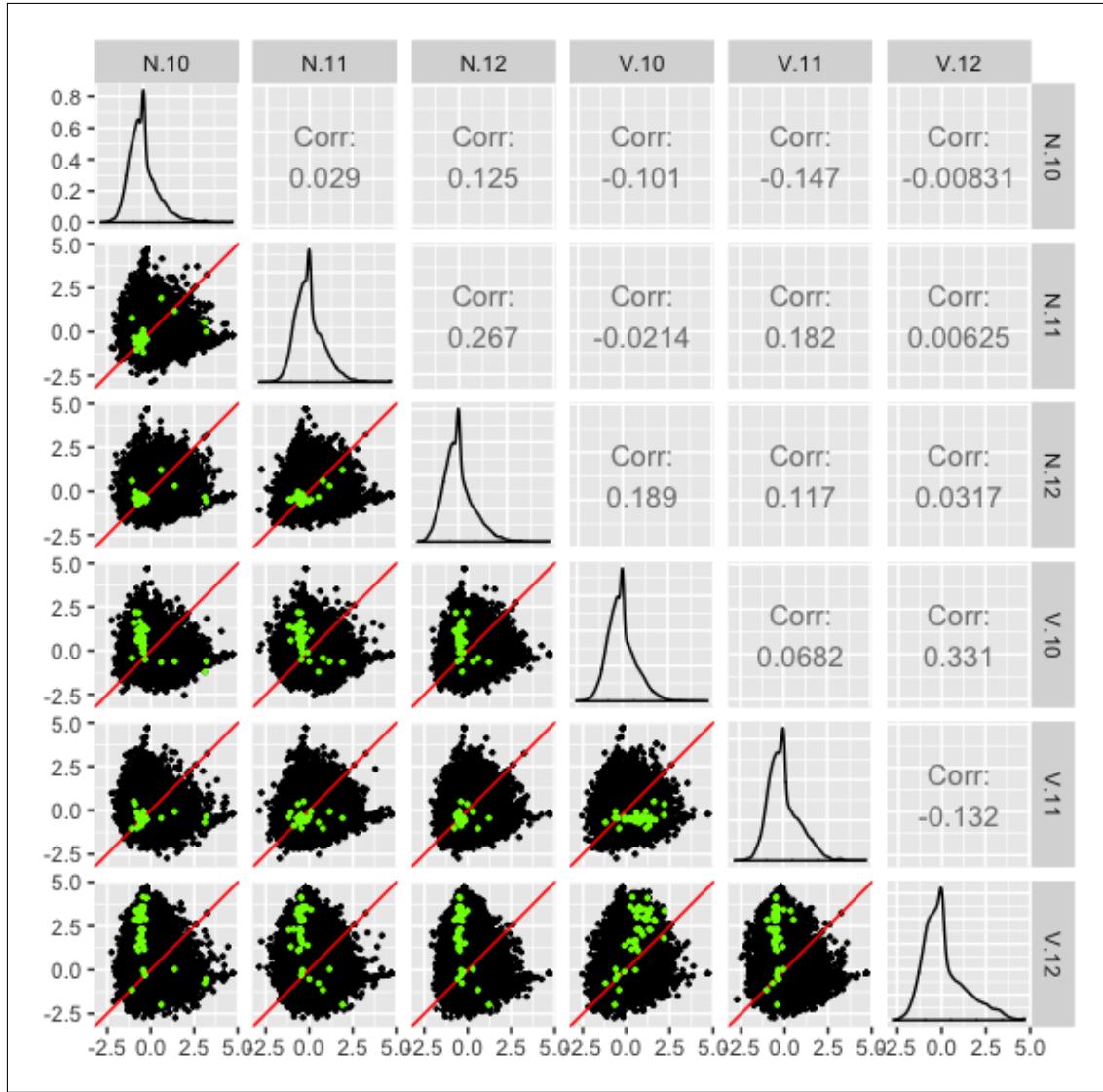


Figure 2.13: The 43 virus-related DEGs from our dataset superimposed onto all genes in the form of a scatterplot matrix. Only replicates 10, 11, and 12 are shown from both treatment groups. The data has been standardized. “N” represents non-infected control samples and “V” represents virus-treated samples. We see that, compared to the scatterplot matrices from 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.

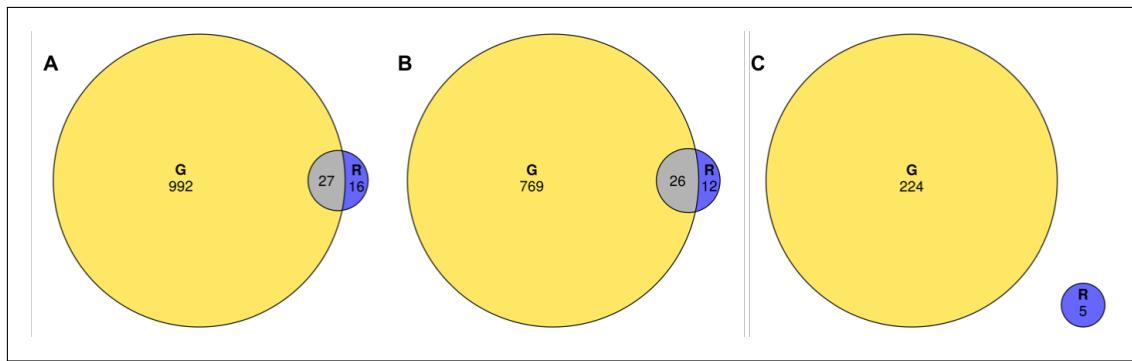


Figure 2.14: Venn diagrams comparing the virus-related DEG overlaps between the Galbraith study (labeled as “G”) and our study (labeled as “R”). From left to right: Total virus-related DEGs (subplot A), virus-upregulated DEGs (subplot B), control-upregulated DEGs (subplot C). Both the total virus-related and virus-upregulated DEGs showed significant overlap between the studies ( $p\text{-value} < 2.2\text{e-}16$ ) as per Fisher’s Exact Test for Count Data. There was one gene that was virus-upregulated in the Galbraith study but control-upregulated in our study.

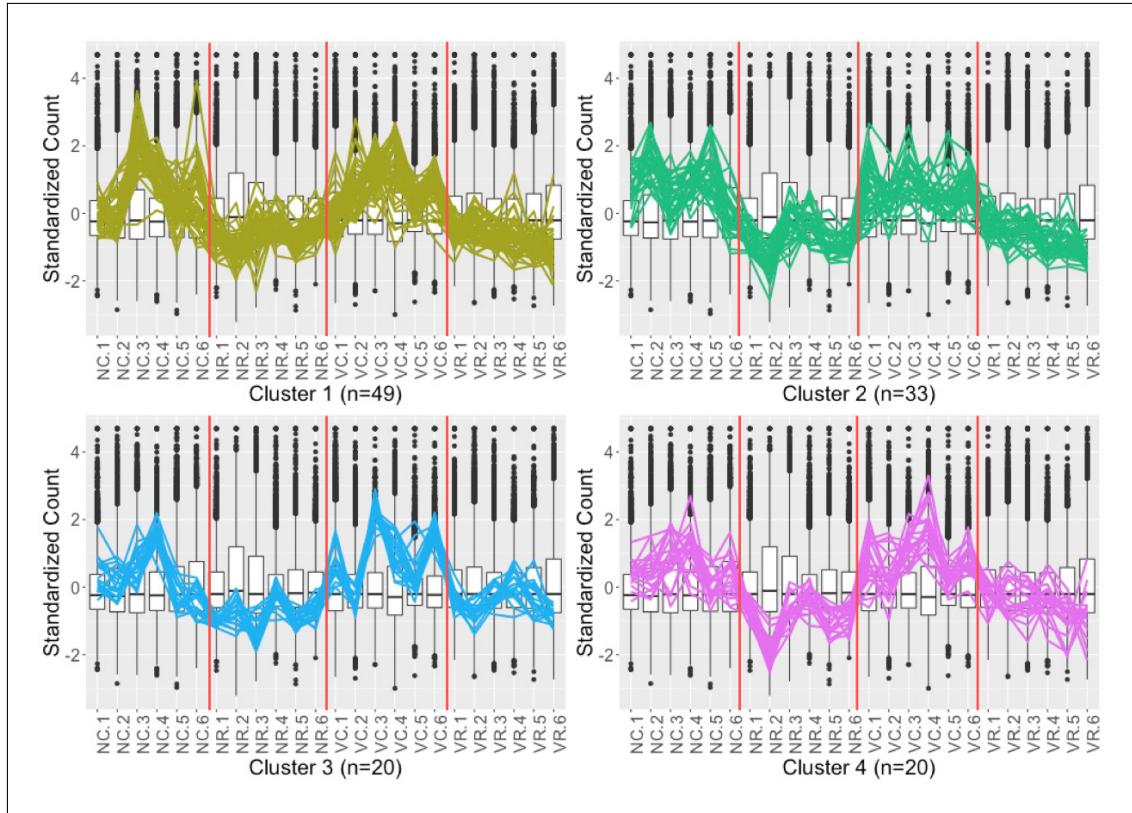


Figure 2.15: Parallel coordinate plots of the 122 DEGs after hierarchical clustering of size four between the “tolerance” candidate genes. Here “N” represents non-infected control group, “V” represents treatment of virus, “C” represents high-quality Chestnut diet, and “R” represents low-quality Rockrose diet. The vertical red line indicates the distinction between treatment groups. We see there is considerable noise in the data (non-level replicate values), but that the general patterns of the DEGs follow what we expect based on our “tolerance” contrast.

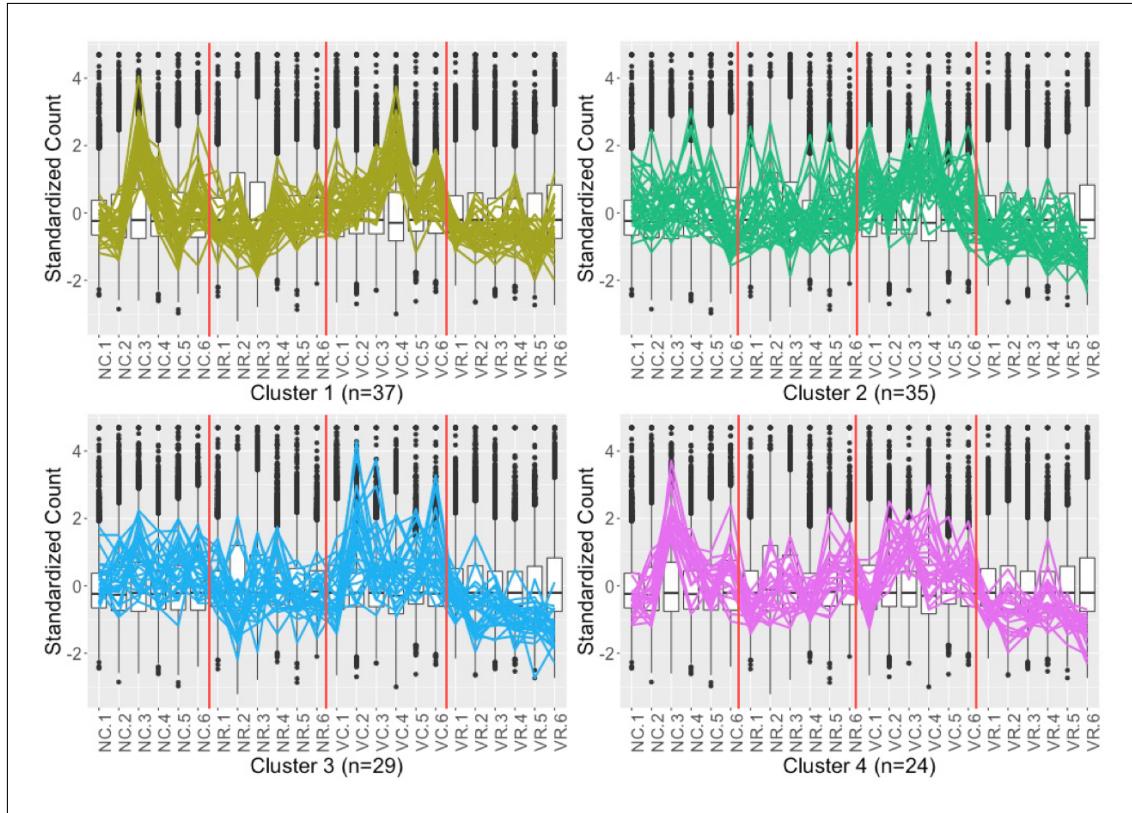


Figure 2.16: Parallel coordinate plots of the 125 DEGs after hierarchical clustering of size four between the “resistance” candidate genes. Here “N” represents non-infected control group, “V” represents treatment of virus, “C” represents high-quality Chestnut diet, and “R” represents low-quality Rockrose diet. The vertical red line indicates the distinction between treatment groups. We see there is considerable noise in the data (non-level replicate values), but that the general patterns of the DEGs follow what we expect based on our “resistance” contrasts.

---

# Bibliography

---

- Alaux, C., Ducloz, F., and Conte, D. C. Y. L. Diet effects on honeybee immunocompetence. *Biol. Lett.*, 6:562–565, 2010.
- Alaux, C., Dantec, C., Parrinello, H., and 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.
- Benjamini, Y. and Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57:289–300, 1995.
- Bond, J., Plattner, K., and Hunt, K. *Fruit and Tree Nuts Outlook: Economic Insight U.S. Pollination- Services Market*. USDA. Economic Research Service Situation and Outlook FTS-357SA, 2014.
- Brodschneider, R. and Crailsheim, K. Nutrition and health in honey bees. *Apidologie*, 41: 278–294, 2010.
- Brodschneider, R., Arnold, G., Hrassnigg, N., and Crailsheim, K. Does patriline composition change over a honey bee queen’s lifetime? *Insects*, 3:857–869, 2012.
- Caron, D. and Sagili, R. Honey bee colony mortality in the Pacific Northwest: Winter 2009/2010. *Am Bee J*, 151:73–76, 2011.
- Carrillo-Tripp, J., Dolezal, A., Goblirsch, M., Miller, W., Toth, A., and Bonning, B. In vivo and in vitro infection dynamics of honey bee viruses. *Sci Rep*, 6:22265, 2016.
- Carval, D. and Ferriere, R. A unified model for the coevolution of resistance, tolerance, and virulence. *Evolution*, 64:2988–3009, 2010.
- Cerenius, L. and Söderhäll, K. The prophenoloxidase-activating system in invertebrates. *Immunological Reviews*, 198:116–126, 2004.

- Chambers, M. and Schneider, D. Balancing resistance and infection tolerance through metabolic means. *PNAS*, 109:13886–13887, 2012.
- Chen, Y. and Siede, R. Honey bee viruses. *Adv Virus Res*, 70:33–80, 2007.
- Chen, Y., Pettis, J., Corona, M., Chen, W., Li, C., Spivak, M., Visscher, P., DeGrandi-Hoffman, G., Boncristiani, H., Zhao, Y., van Engelsdorp, D., Delaplane, K., Solter, L., Drummond, F., Kramer, M., Lipkin, W., Palacios, G., Hamilton, M., Smith, B., Huang, S., Zheng, H., Li, J., Zhang, X., Zhou, X., Wu, L., Zhou, J., Lee, M.-L., Teixeira, E., Li, Z., and Evans, J. Israeli acute paralysis virus: Epidemiology, pathogenesis and implications for honey bee health. *Plos Pathog*, 10:e1004261, 2014.
- Consortium, H. B. G. S. Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics*, 15:86, 2014.
- Conte, Y. L., Brunet, J.-L., McDonnell, C., and Alaux, C. *Interactions between risk factors in honey bees*. CRC Press, 2011.
- Cornman, R., Tarpy, D., Chen, Y., Jeffreys, L., Lopez, D., and Pettis, J. Pathogen webs in collapsing honey bee colonies. *Plos One*, 7:e43562, 2012.
- Cox-Foster, D., Conlan, S., Holmes, E., Palacios, G., Evans, J., Moran, N., Quan, P.-L., Briese, T., Hornig, M., Geiser, D., Martinson, V., vanEngelsdorp, D., Kalkstein, A., Drysdale, A., Hui, J., Zhai, J., Cui, L., Hutchison, S., Simons, J., Egholm, M., Pettis, J., and Lipkin, W. A metagenomic survey of microbes in honey bee colony collapse disorder. *Science*, 318:283–287, 2007.
- Crailsheim, K. The flow of jelly within a honeybee colony. *J Comp Physiol B*, 162:681–689, 1992.
- Crailsheim, K., Schneider, L., Hrassnigg, N., Bühlmann, G., Brosch, U., Gmeinbauer, R., and Schöffmann, B. Pollen consumption and utilization in worker honeybees (*Apis mellifera carnica*): dependence on individual age and function. *J Insect Physiol*, 38: 409–419, 1992.
- Crozier, R. and Page, R. On being the right size: Male contributions and multiple mating in social Hymenoptera. *Behav. Ecol. Sociobiol.*, 18:105–115, 1985.
- Decourtey, A., Mader, E., and Desneux, N. Landscape enhancement of floral resources for honey bees in agro-ecosystems. *Apidologie*, 41:264–277, 2010.
- DeGrandi-Hoffman, G. and Chen, Y. Nutrition, immunity and viral infections in honey bees. *Current Opinion in Insect Science*, 10:170–176, 2015.

- DeGrandi-Hoffman, G., Chen, Y., Huang, E., and Huang, M. 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.
- Dolezal, A. and Toth, A. Feedbacks between nutrition and disease in honey bee health. *Current Opinion in Insect Science*, 26:114–119, 2018.
- Dolezal, A., Carrillo-Tripp, J., Judd, T., Miller, A., Bonning, B., and Toth, A. Interacting stressors matter: Diet quality and virus infection in honey bee health. *In prep*, 2018.
- Elsik, C., Tayal, A., Diesh, C., Unni, D., Emery, M., Nguyen, H., and Hagen, D. Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine. *Nucleic Acids Research*, 4:D793–800, 2016.
- Engelsdorf, D. V. and Meixner, M. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *J Invertebr Pathol*, 103:S80–S95, 2010.
- Engelsdorf, D. V., Hayes, J. J., Underwood, R., and Pettis, J. A survey of honey bee colony losses in the U.S., fall 2007 to spring 2008. *Plos One*, 3:e4071, 2008.
- Fornoni, J., Nunez-Farfán, J., Valverde, P., and Rausher, M. Evolution of mixed plant defense allocation against natural enemies. *Evolution*, 58:1685–1695, 2004.
- Galbraith, D., Yang, X., Niño, E., Yi, S., and Grozinger, C. Parallel epigenomic and transcriptomic responses to viral infection in honey bees (*Apis mellifera*). *Plos Pathogens*, 11:e1004713, 2015.
- Gallai, N., Salles, J.-M., Settele, J., and Vaissière, B. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecol. Econ.*, 68:810–821, 2009.
- Goulson, D., Nicholls, E., Botías, C., and Rotheray, E. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science*, 347:1255957, 2015.
- Haydak, M. Honey bee nutrition. *Annu Rev Entomol*, 15:143–156, 1970.
- Hou, C., Rivkin, H., Slabezki, Y., and Chejanovsky, N. Dynamics of the presence of israeli acute paralysis virus in honey bee colonies with colony collapse disorder. *Viruses*, 6: 2012–2027, 2014.
- Huang, D., Sherman, B., and Lempicki, R. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4:44–57, 2009a.
- Huang, D., Sherman, B., and Lempicki, R. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37:1–13, 2009b.

- Klein, A.-M., Vaissière, B., Cane, J., Steffan-Dewenter, I., Cunningham, S., Kremen, C., and Tscharntke, T. Importance of pollinators in changing landscapes for world crops. *Proc Biol Sci*, 274:303–313, 2007.
- Kulhanek, K., Steinhauer, N., Rennich, K., Caron, D., Sagili, R., Pettis, J., Ellis, J., Wilson, M., Wilkes, J., Tarpy, D., Rose, R., Lee, K., Rangel, J., and 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.
- Larsson, J. *eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses*, 2018. URL <https://cran.r-project.org/package=eulerr>. R package version 4.0.0.
- Laurent, M., Hendrikx, P., Ribiere-Chabert, M., and Chauzat, M.-P. A pan-European epidemiological study on honeybee colony losses 2012–2014. *Epilobee*, 2013:44, 2016.
- Love, M., Huber, W., and Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15:550, 2014.
- Maori, E., Paldi, N., Shafir, S., Kalev, H., Tsur, E., Glick, E., and Sela, I. IAPV, a bee-affecting virus associated with Colony Collapse Disorder can be silenced by dsRNA ingestion. *Insect Mol Biol*, 18:55–60, 2009.
- Mattila, H. and Seeley, T. Genetic diversity in honey bee colonies enhances productivity and fitness. *Science*, 317:362–364, 2007.
- Mauricio, R., Rausher, M., and Burdick, D. Variation in the defense strategies of plants: are resistance and tolerance mutually exclusive? *Ecology*, 78:1301–1310, 1997.
- Miller, C. and Cotter, S. Resistance and tolerance: The role of nutrients on pathogen dynamics and infection outcomes in an insect host. *Journal of Animal Ecology*, 87: 500–510, 2017.
- Miranda, J. D., Cordoni, G., and Budge, G. The acute bee paralysis virus-Kashmir bee virus-Israeli acute paralysis virus complex. *J Invertebr Pathol*, 103:S30–47, 2010.
- Moret, Y. Trans-generational immune priming: Specific enhancement of the antimicrobial immune response in the mealworm beetle, *Tenebrio molitor*. *Proceedings of the Royal Society B: Biological Sciences*, 273:1399–1405, 2006.
- Naug, D. Nutritional stress due to habitat loss may explain recent honeybee colony collapses. *Biol Conserv*, 142:2369–2372, 2009.
- Neumann, P. and Carreck, N. Honey bee colony losses. *J Apicult Res*, 49:1–6, 2010.
- Page, R. and Laidlaw, H. Full sisters and supersisters: A terminological paradigm. *Anim. Behav.*, 36:944–945, 1988.

- Pasquale, G., Salignon, M., Conte, Y., Belzunces, L., Decourtye, A., Kretzschmar, A., Suchail, S., Brunet, J.-L., and Alaux, C. Influence of pollen nutrition on honey bee health: Do pollen quality and diversity matter? *Plos One*, 8:e72016, 2013.
- Potts, S., Biesmeijer, J., Kremen, C., Neumann, P., Schweiger, O., and Kunin, W. . *Global pollinator declines: trends, impacts and drivers*, 25:345–353, 2010.
- Restif, O. and Koella, J. Shared control of epidemiological traits in a coevolutionary model of host-parasite interactions. *The American Naturalist*, 161:827– 836, 2003.
- Ritchie, M., Phipson, B., Wu, D., Hu, Y., Law, C., Shi, W., and Smyth, G. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47, 2015.
- Robinson, M., McCarthy, D., and Smyth, G. edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26:139–140, 2010.
- Rosenkranz, P., Aumeier, P., and Ziegelmann, B. Biology and control of Varroa destructor. *J Invertebr Pathol*, 103:S96–S119, 2010.
- Roulston, T. and Buchmann, S. A phylogenetic reconsideration of the pollen starch-pollination correlation. *Evol Ecol Res*, 2:627–643, 2000.
- Sadd, B. and Siva-Jothy, M. Self-harm caused by an insect's innate immunity. *Proceedings of the Royal Society B: Biological Sciences*, 273:2571–2574, 2006.
- Schmidt, J. Feeding preference of *Apis mellifera* L. (Hymenoptera: Apidae): Individual versus mixed pollen species. *J. Kans. Entomol. Soc.*, 57:323–327, 1984.
- Schmidt, J., Thoenes, S., and Levin, M. Survival of honey bees, *Apis mellifera* (Hymenoptera: Apidae), fed various pollen sources. *J. Econ. Entomol.*, 80:176–183, 1987.
- Shen, M., Cui, L., Ostiguy, N., and Cox-Foster, D. Intricate transmission routes and interactions between picorna-like viruses (Kashmir bee virus and sacbrood virus) with the honeybee host and the parasitic varroa mite. *J Gen Virol*, 86:2281–2289, 2005.
- Sherman, P., Seeley, T., and Reeve, H. Parasites, pathogens, and polyandry in social Hymenoptera. *Am. Nat*, 131:602–610, 1988.
- Spivak, M., Mader, E., Vaughan, M., and Euliss, N. The Plight of the Bees. *Environ Sci Technol*, 45:34–38, 2011.
- Stanley, R. and Linskens, H. *Pollen: Biology, biochemistry, management*. Springer Verlag, 1974.
- Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. REVIGO summarizes and visualizes long lists of Gene Ontology terms. *Plos ONE*, 6:e21800, 2011.

- Tarpy, D. Genetic diversity within honeybee colonies prevents severe infections and promotes colony growth. *Proc. R. Soc. Lond. B*, 270:99–103, 2003.
- van Engelsdorp, D., Evans, J., Saegerman, C., Mullin, C., Haubrige, E., Nguyen, B., Frazier, M., Frazier, J., Cox-Foster, D., Chen, Y., Underwood, R., Tarpy, D., and Pettis, J. Colony collapse disorder: A descriptive study. *PLoS One*, 4:e6481, 2009.
- Weinberg, K. and Madel, G. The influence of the mite Varroa Jacobsoni Oud. on the protein concentration and the haemolymph volume of the brood of worker bees and drones of the honey bee Apis Mellifera L. *Apidologie*, 16:421–436, 1985.
- Wu, T., Reeder, J., Lawrence, M., Becker, G., and Brauer, M. GMAP and GSNAP for genomic sequence alignment: Enhancements to speed, accuracy, and functionality. *Methods Mol Biol*, 1418:283–334, 2016.
- Xu, J., Grant, G., Sabin, L., Gordesky-Gold, B., Yasunaga, A., Tudor, M., and Cherry, S. Transcriptional pausing controls a rapid antiviral innate immune response in Drosophila. *Cell Host Microbe*, 12:531–543, 2012.
- Yang, X. and Cox-Foster, D. Effects of parasitization by Varroa destructor on survivorship and physiological traits of Apis mellifera in correlation with viral incidence and microbial challenge. *Parasitology*, 134:405–412, 2007.
- Yang, X. and Cox-Foster, D. Impact of an ectoparasite on the immunity and pathology of an invertebrate: Evidence for host immunosuppression and viral amplification. *P Natl Acad Sci USA*, 102:7470–7475, 2005.