Chapter 1

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

1.1 Introduction

Commerically managed honeybees 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, vegatables, and nuts) at least benefit from pollination, and yearly insect pollination services are valued wordwide at $175 billion (Gallai et al. 2009). As honeybees are largely considered to be the leading pollinator of numerous crops, their marked loss has considerable implications regarding agricultural sustainability (Klein et al. 2007).

Honeybee 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 honeybees in interactive fashions as the environment changes (Goulson et al. 2015). Nutrition and viral infection are two broad factors that pose heightened dangers to honeybee 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 honeybees (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 honeybees. In particular, honeybees 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 (Engelsdorp et al. 2008). Understanding how undiversified diets affect honeybee health will be crucial to resolve problems that may arise as agriculture continues to intensify throughout the world (Neumann and Carreck 2010, Engelsdorp and Meixner 2010).

Viral infection was a comparatively minor problem in honeybees until the last century when Varroa destructor (an ectoparasitic mite) spread worldwide (Rosenkranz et al. 2010). This mite feeds on honeybee 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 honeybee viruses have been identified (Chen and Siede 2007). One of these viruses that has been linked to honeybee decline is Israeli Acute Paralysis Virus (IAPV). A positive-sense RNA virus of the Dicistroviridae family (Miranda et al. 2010), IAPV causes infected honeybees to display shivering wings, decreased locomotion, muscle spams, and paralysis, and 80% of caged infected adult honeybees die prematurely (Maori et al. 2009). IAPV has demonstrated higher infectious capacities than other honeybee 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 seems likely that IAPV reduces colony strength and survival.

Although there is growing interest in how viruses and diet quality affect the health and sustainability of honeybees, 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 honeybees (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 a significant interaction between 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 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 innoculation in any broad sense.

In this study, we examine how monofloral diets and viral innoculation 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).

1.2.1 Design of two-factor experiment

There are several reasons why 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 honeybees as they rarely face a total lack of pollen (Pasquale et al. 2013). Second, in studies that compared honeybee 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 honeybee health as a step toward mitigating the negative impact of human activity on the honeybee population.

Consequently, for our nutrition factor, we examined two monofloral pollen diets, Cistus (Rockrose) and Castanea (Chestnut). Cistus pollen is generally considered less nutritious than Castanea 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 (Cistus pollen without IAPV exposure, Castanea pollen without IAPV exposure, Cistus pollen with IAPV exposure, and Castanea pollen with IAPV exposure) that allowed us to assess main effects and interactive effects between diet quality and IAPV infection in honeybees.

1.2.2 RNA extraction

Fifteen cages per treatment were originally sampled. Six live honeybees from each cage were randomly selected 36 hours post inoculation and placed into tubes. 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 honeybees per cage were randomly selected from the original six live honeybees per cage. Whole body RNA from each pool of two honeybees were extracted using Qiagen RNeasy MiniKit followed by Qiagen DNase treatment. Samples were suspended in water to 200-400 ng/μl. 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). 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 honeybees (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).

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

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). While our study examines honeybees from polyandrous colonies, the Galbraith study examined honeybees from single-drone colonies. As a consequence, our honeybees will have an average of about 75% genetic variance, and the honeybees from the Galbraith study will have an average of about 25% genetic variance (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 honeybees 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 honeybees are naturally very polyandrous (Brodschneider et al. 2012), our honeybees 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 innoculation 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 honeybees.

1.2.5 Visualization

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

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 identified using the BEEBASE_ID identifier. To fine-tune the GO term list, only significant terms (FDR < 0.05) and those 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. @@@ 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).

Mortality rates of honeybees 72 hour post-inoculation significantly differed among the treatment groups (mixed model ANOVA across all treatment groups, df=3, 55; F=10.07; p<2.18e-05). The effect of virus treatment (mixed model ANOVA, df=1, 55; F=24.343; p<7.84e-06) and diet treatment (mixed model ANOVA, df=1, 55; F=5.796; p<0.0194) were significant, but the interaction between the two factors (mixed model ANOVA, df=1, 55; F=0.062, p=0.8039) was not significant. The virus treatment was significant: For a given diet, honeybees exposed to the virus showed significantly higher mortality rate than

honeybees not exposed to the virus (Tukey HSD, $p<0.05$). Without virus exposure, there was only a numerical reduction in mortality rate for bees fed Castanea pollen compared to Rockrose pollen (Tukey HSD, $p>0.05$). Similarly, with virus exposure, there was only a numerical reduction in mortality rate for beeds fed Castanea pollen compared to Rockrose pollen (Tukey HSD, $p>0.05$). Overall, we discovered that the higher-quality Castanea diet had the ability to numerically reduce mortality in the presence of IAPV infection compared to the lower-quality Rockrose diet (Figure 1.1 A).

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

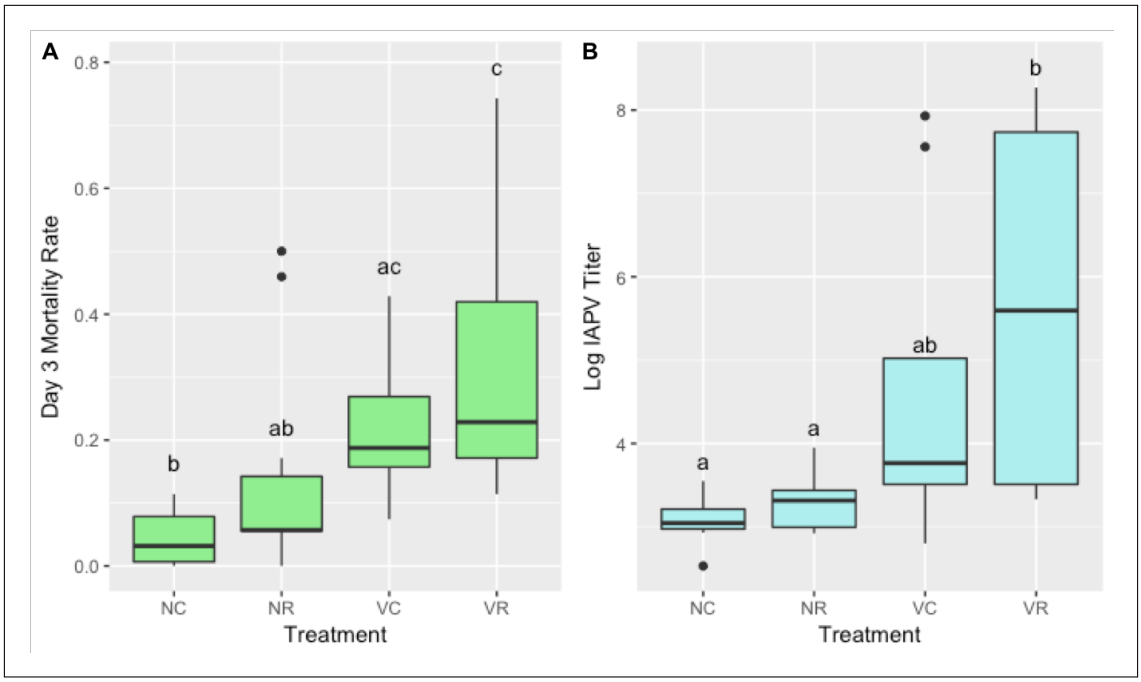

Figure 1.1: Mortality rates (A) and IAPV titers (B) for the four treatment groups.

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) (Table 1.1A and B). In the diet factor, there were more Castanea DEGs (n = 1033) than Rockrose DEGs (n = 881). In the virus factor, there were more virus-exposed DEGs (n = 38) than control DEGs (n = 5). While these reported DEGs numbers are from the DESeq2 package, we saw similar trends for the edgeR and limma package results (Table 1.1A and B).

Pathway analysis of the Castanea DEGs revealed enriched (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 (Table 1.2). Pathway 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 (Table 1.3).

It was difficult to perform a pathway analysis with so few DEGs (n = 43) in our virus main effect study. As a result, we focused on individual genes and their known functionalities (Table 1.5). Of the 43 virus-related DEGs, only 10 of them successfully mapped with DAVID software.

1.3.3 Interaction DEG results

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

1.3.4 Pairwise comparison DEG results

The number of DEGs across the six treatment pairings between the diet and virus factor ranged from 0 to 941 (Table 1.4). 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 Castanea and Rockrose levels, regardless of the virus level, the number of Castanea-related DEGs was higher than the number of Rockrose-related DEGs (Table 1.4 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 (Table 1.4 A, B). There were no DEGs between the treatment pair controlling for the control level of the virus effect (Table 1.4 A). These trends were observed for all three pipelines used (DESeq2, edgeR, and limma).

1.3.5 Comparison with Galbraith study

Differences between signal to noise ratio

Figure 1.2 Figure 1.3 Figure 1.4

Overlap between studies Figure 1.5.

27 genes overlapped ("GB40006" "GB41739" "GB42046" "GB42048" "GB42313" "GB43783" "GB43784" "GB45462" "GB45704" "GB47214" "GB47381" "GB47407" "GB48747" "GB48755" "GB49920" "GB50178" "GB50401" "GB50550" "GB50813" "GB50955" "GB51305" "GB52281" "GB52441" "GB52449" "GB53500" "GB53833" "GB55188")

26 genes overlapped virus ("GB40006" "GB41739" "GB42046" "GB42048" "GB42313" "GB43783" "GB43784" "GB45462" "GB45704" "GB47214" "GB47381" "GB47407" "GB48747" "GB48755" "GB49920" "GB50178" "GB50401" "GB50550" "GB50813" "GB50955" "GB52281" "GB52441" "GB52449" "GB53500" "GB53833" "GB55188")

1.4 Discussion

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.

Pathway Term	# of Genes	Benjamini	Example Genes
Wnt signaling pathway	15	2.20E-03	*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/threonin-protein kinase NLK, stress-activated protein kinase JNK*
Dorso-ventral axis formation	8	2.80E-02	*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*
Hippo signaling pathway	12	3.00E-02	*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*
Circadian rhythm	4	2.40E-01	*casein kinase I-like, protein cycle, protein kinase shaggy, thyrotroph embryonic factor*
mRNA surveillance pathway	10	2.60E-01	*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*
Insulin resistance	8	2.80E-01	*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*
Inositol phosphate metabolism	8	2.90E-01	*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*
FoxO signaling pathway	9	3.00E-01	*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*
ECM-receptor interaction	5	3.20E-01	*agrin-like, collagen alpha-1 (IV) chain, collagen alpha-5 (IV) chain, dystroglycan, integrin beta-PS-like*
Phototransduction	6	3.30E-01	*1-phosphatidylinositol 4,5-biphosphate phosphodiesterase, actin muscle-like, calcium/calmodulin-dependent protein kinase II, G protein-coupled receptor kinase 1, protein kinase*
Notch signaling pathway	5	3.80E-01	*C-terminal-binding protein, histone acetyltransferase p300-like, neurogenic locus Notch protein, protein jagged-1, protein numb*
Jak-STAT signaling pathway	4	3.90E-01	*histone acetyltransferase p300-like, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein son of sevenless*
MAPK signaling pathway	4	4.40E-01	*epidermal growth factor receptor-like, ETS-like protein pointed, protein son of sevenless, proto-oncogene tyrosine-protein kinase ROS*
Carbon metabolism	12	4.50E-01	*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*

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	*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/threonin-protein kinase NLK, stress-activated protein kinase JNK*
Dorso-ventral axis formation	8	2.80E-02	*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*
Hippo signaling pathway	12	3.00E-02	*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*
Circadian rhythm	4	2.40E-01	*casein kinase I-like, protein cycle, protein kinase shaggy, thyrotroph embryonic factor*
mRNA surveillance pathway	10	2.60E-01	*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*
Insulin resistance	8	2.80E-01	*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*
Inositol phosphate metabolism	8	2.90E-01	*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*
FoxO signaling pathway	9	3.00E-01	*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*
ECM-receptor interaction	5	3.20E-01	*agrin-like, collagen alpha-1 (IV) chain, collagen alpha-5 (IV) chain, dystroglycan, integrin beta-PS-like*
Phototransduction	6	3.30E-01	*1-phosphatidylinositol 4,5-biphosphate phosphodiesterase, actin muscle-like, calcium/calmodulin-dependent protein kinase II, G protein-coupled receptor kinase 1, protein kinase*
Notch signaling pathway	5	3.80E-01	*C-terminal-binding protein, histone acetyltransferase p300-like, neurogenic locus Notch protein, protein jagged-1, protein numb*
Jak-STAT signaling pathway	4	3.90E-01	*histone acetyltransferase p300-like, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform, protein son of sevenless*
MAPK signaling pathway	4	4.40E-01	*epidermal growth factor receptor-like, ETS-like protein pointed, protein son of sevenless, proto-oncogene tyrosine-protein kinase ROS*
Carbon metabolism	12	4.50E-01	*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*

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.

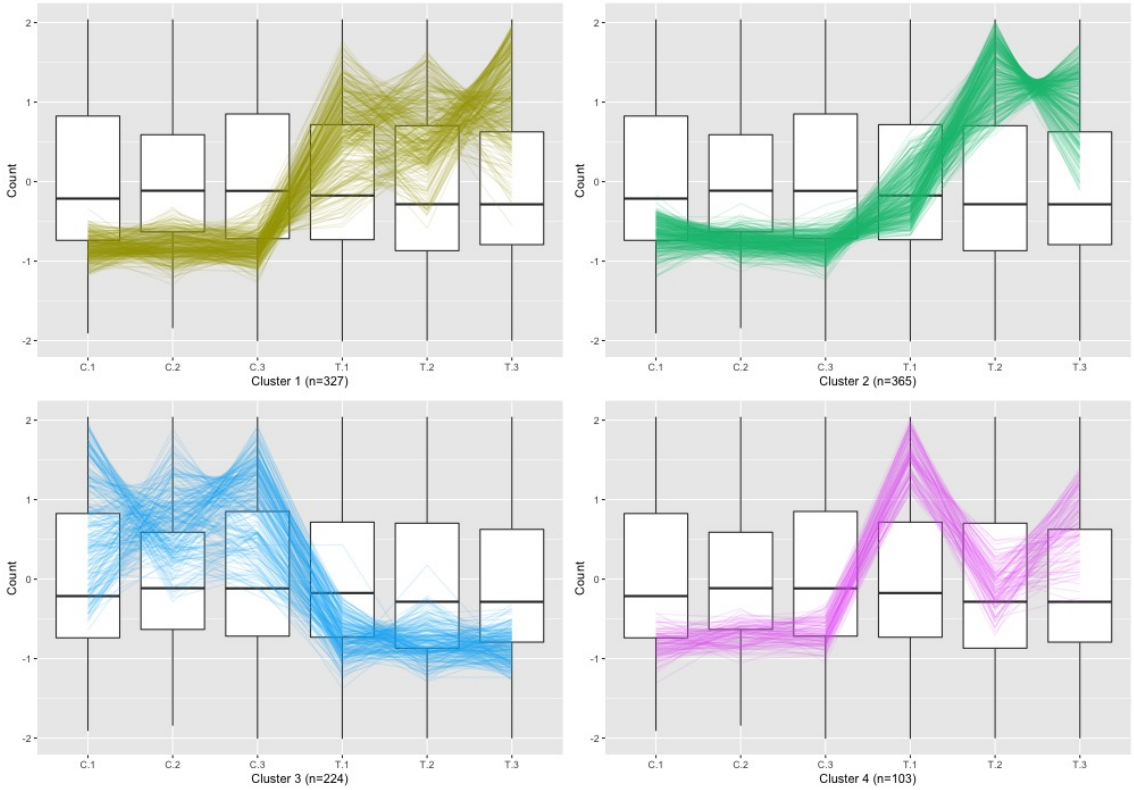

Figure 1.2: Parallel coordinate plots of DEGs between the virus-infected and control groups of the Galbraith study.

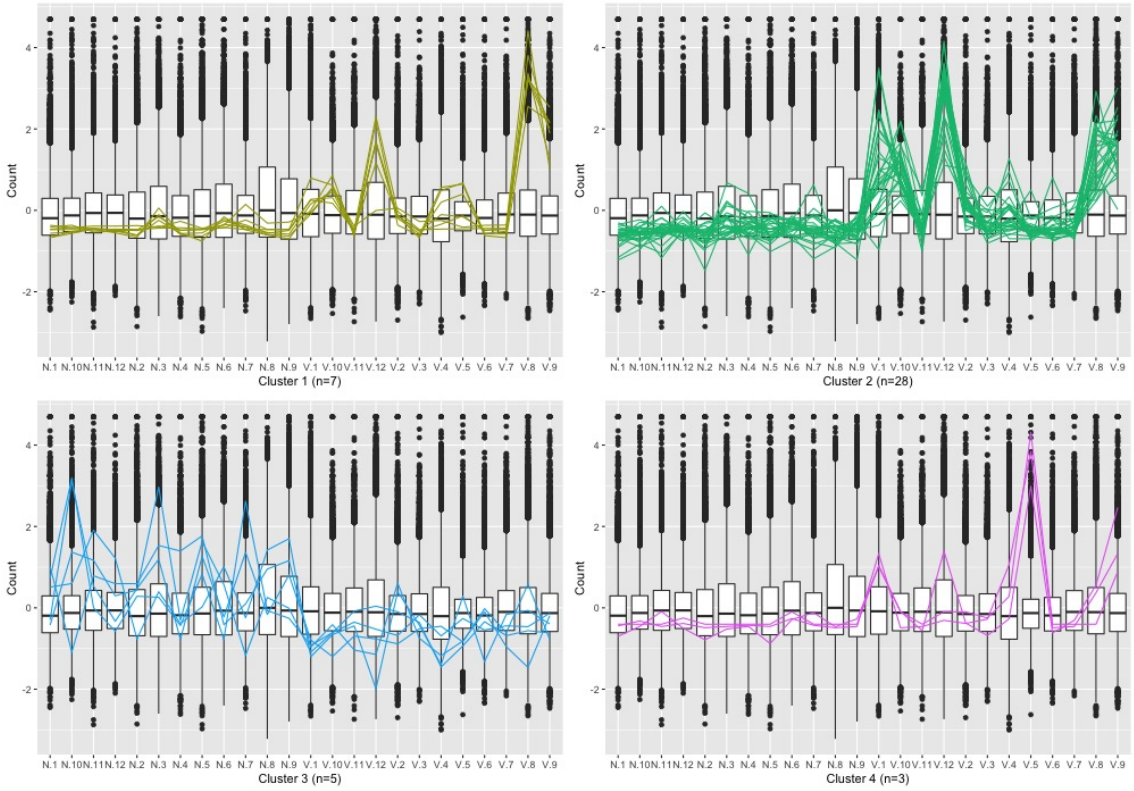

Figure 1.3: Parallel coordinate plots of DEGs between the virus-infected and control groups of our study.

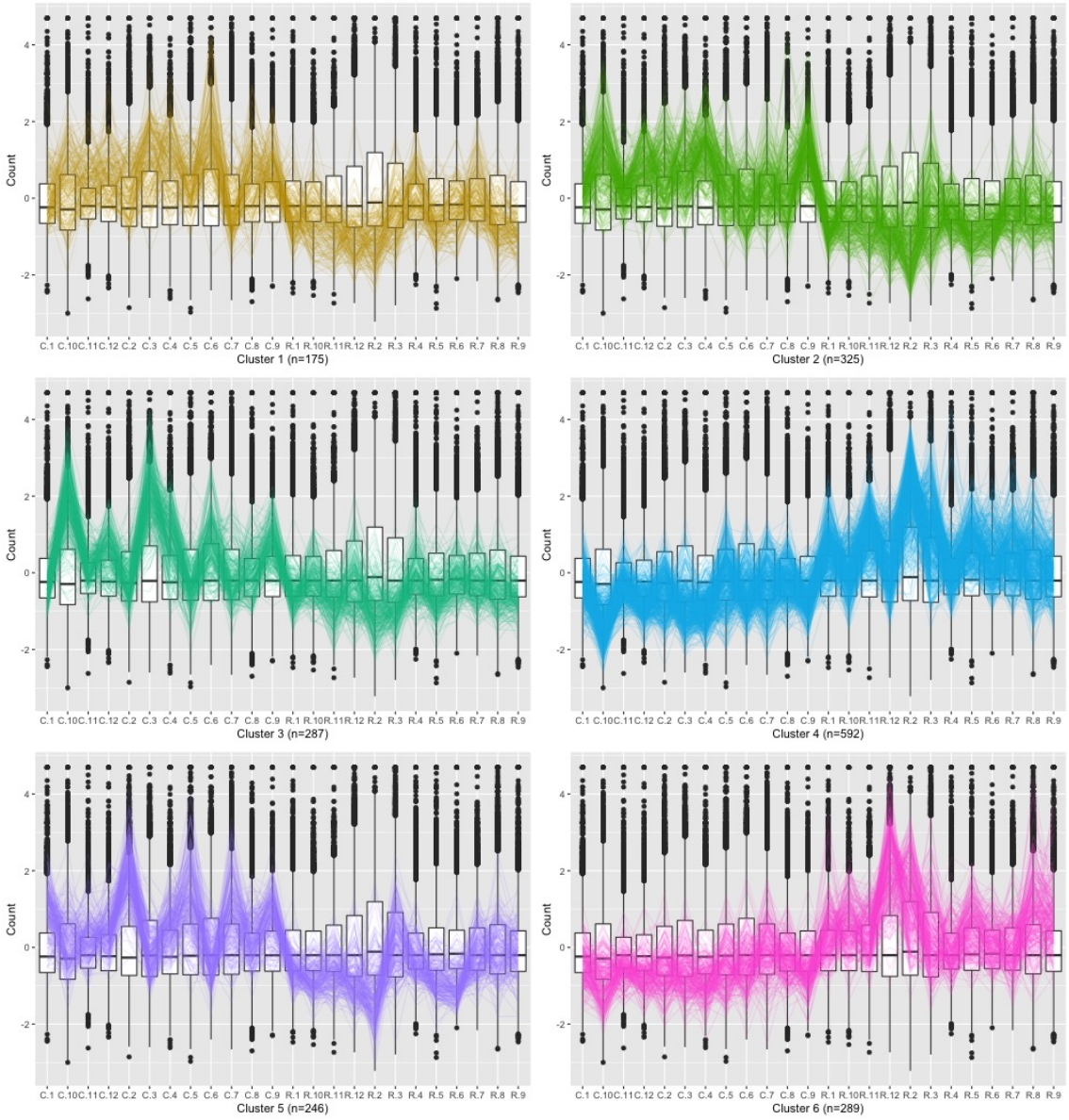

Figure 1.4: Parallel coordinate plots of DEGs between the Castanea and Rockrose groups of our study.

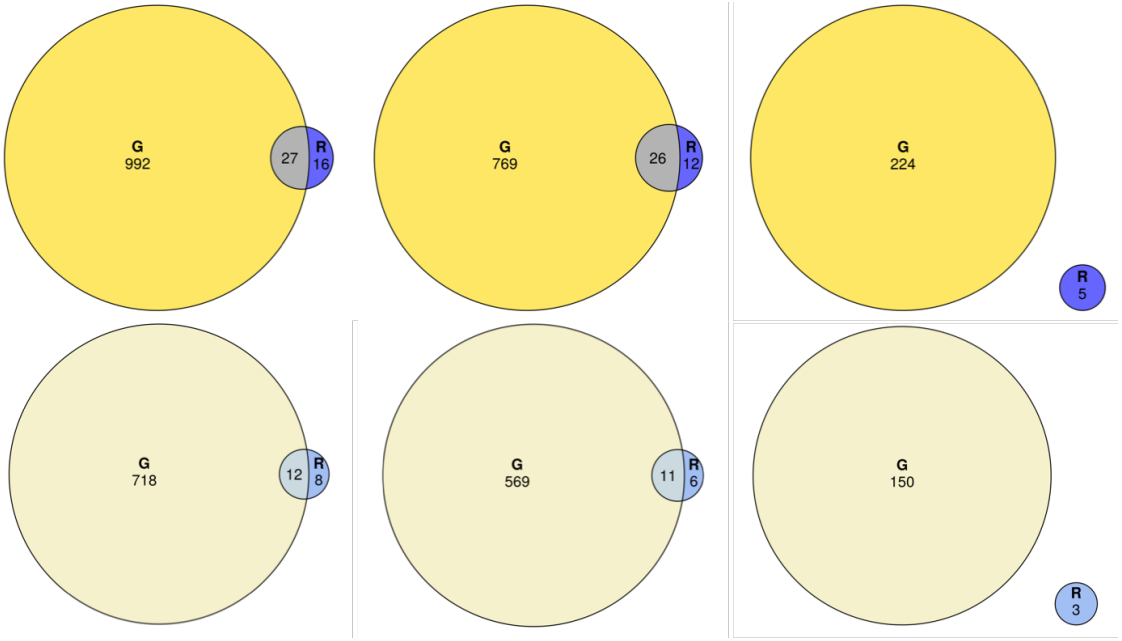

Figure 1.5: Venn diagrams comparing the DEG overlaps of the diet and treatment groups between the Galbraith study and our study. Top row left to right: DESeq results for all DEGs, virus DEGs, and control DEGs. Bottom row left to right: EdgeR results for all DEGs, virus DEGs, and control DEGs.

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

Bibliography

C. Alaux, F. Ducloz, and D. Crauser Y. Le Conte. Diet effects on honeybee immunocompetence. *Biol. Lett.*, 6:562–565, 2010.

C. Alaux, C. Dantec, H. Parrinello, and Y. Le Conte. Nutrigenomics in honey bees: digital gene expression analysis of pollen's nutritive effects on healthy and varroa-parasitized bees. *BMC Genomics*, 12:496, 2011.

Y. Benjamini and Y. Hochberg. 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.

J. Bond, K. Plattner, and K. Hunt. *Fruit and Tree Nuts Outlook: Economic Insight U.S. Pollination- Services Market*. USDA. Economic Research Service Situation and Outlook FTS-357SA, 2014.

R. Brodschneider and K. Crailsheim. Nutrition and health in honey bees. *Apidologie*, 41: 278–294, 2010.

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

D. Caron and R. Sagili. Honey bee colony mortality in the Pacific Northwest: Winter 2009/2010. *Am Bee J*, 151:73–76, 2011.

J. Carrillo-Tripp, A.G. Dolezal, M.J. Goblirsch, W.A. Miller, A.L. Toth, and B.C. Bonning. In vivo and in vitro infection dynamics of honey bee viruses. *Sci Rep*, 6:22265, 2016.

Y.P. Chen and R. Siede. Honey bee viruses. *Adv Virus Res*, 70:33–80, 2007.

Y.P. Chen, J.S. Pettis, M. Corona, W.P. Chen, C.J. Li, M. Spivak, P.K. Visscher, G. DeGrandi-Hoffman, H. Boncristiani, Y. Zhao, D. van Engelsdorp, K. Delaplane, L. Solter, F. Drummond, M. Kramer, W.I. Lipkin, G. Palacios, M.C. Hamilton, B. Smith,

S.K. Huang, H.Q. Zheng, J.L. Li, X. Zhang, X.F. Zhou, L.Y. Wu, J.Z. Zhou, M-L. Lee, E.W. Teixeira, Z.G. Li, and J.D. Evans. Israeli acute paralysis virus: Epidemiology, pathogenesis and implications for honey bee health. *Plos Pathog*, 10:e1004261, 2014.

Honey Bee Genome Sequencing Consortium. Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics*, 15:86, 2014.

Y. Le Conte, J-L. Brunet, C. McDonnell, and C. Alaux. *Interactions between risk factors in honey bees*. CRC Press, 2011.

R.S. Cornman, D.R. Tarpy, Y. Chen, L. Jeffreys, D. Lopez, and J.S. Pettis. Pathogen webs in collapsing honey bee colonies. *Plos One*, 7:e43562, 2012.

D.L. Cox-Foster, S. Conlan, E.C. Holmes, G. Palacios, J.D. Evans, N.A. Moran, P-L. Quan, T. Briese, M. Hornig, D.M. Geiser, V. Martinson, D. vanEngelsdorp, A.L. Kalkstein, A. Drysdale, J. Hui, J. Zhai, L. Cui, S.K. Hutchison, J.F. Simons, M. Egholm, J.S. Pettis, and W.I. Lipkin. A metagenomic survey of microbes in honey bee colony collapse disorder. *Science*, 318:283–287, 2007.

K. Crailsheim. The flow of jelly within a honeybee colony. *J Comp Physiol B*, 162:681–689, 1992.

K. Crailsheim, L.H.W Schneider, N. Hrassnigg, G. Bühlmann, U. Brosch, R. Gmeinbauer, and B. Schöffmann. Pollen consumption and utilization in worker honeybees (Apis mellifera carnica): dependence on individual age and function. *J Insect Physiol*, 38: 409–419, 1992.

R.H. Crozier and R.E. Page. On being the right size: Male contributions and multiple mating in social Hymenoptera. *Behav. Ecol. Sociobiol.*, 18:105–115, 1985.

A. Decourtye, E. Mader, and N. Desneux. Landscape enhancement of floral resources for honey bees in agro-ecosystems. *Apidologie*, 41:264–277, 2010.

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

G. DeGrandi-Hoffman, Y. Chen, E. Huang, and M.H Huang. 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.

A.G. Dolezal and A.L. Toth. Feedbacks between nutrition and disease in honey bee health. *Current Opinion in Insect Science*, 26:114–119, 2018.

A.G. Dolezal, J. Carrillo-Tripp, T. Judd, A. Miller, B. Bonning, and A. Toth. Interacting stressors matter: Diet quality and virus infection in honey bee health. *In prep*, 2018.

C.G. Elsik, A. Tayal, C.M. Diesh, D.R. Unni, M.L. Emery, H.N. Nguyen, and D.E. Hagen. Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine. *Nucleic Acids Research*, 4:D793–800, 2016.

D. Van Engelsdorp and M.D. Meixner. 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.

D. Van Engelsdorp, J. Jr. Hayes, R.M Underwood, and J. Pettis. A survey of honey bee colony losses in the U.S., fall 2007 to spring 2008. *Plos One*, 3:e4071, 2008.

D.A. Galbraith, X. Yang, E.L. Niño, S. Yi, and C. Grozinger. Parallel epigenomic and transcriptomic responses to viral infection in honey bees (Apis mellifera). *Plos Pathogens*, 11:e1004713, 2015.

N. Gallai, J-M. Salles, J. Settele, and B.B. Vaissière. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecol. Econ.*, 68:810–821, 2009.

D. Goulson, E. Nicholls, C. Botías, and E.L. Rotheray. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science*, 347:1255957, 2015.

M.H. Haydak. Honey bee nutrition. *Annu Rev Entomol*, 15:143–156, 1970.

C. Hou, H. Rivkin, Y. Slabezki, and N. Chejanovsky. Dynamics of the presence of israeli acute paralysis virus in honey bee colonies with colony collapse disorder. *Viruses*, 6: 2012–2027, 2014.

D.W. Huang, B.T. Sherman, and R. Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4:44–57, 2009a.

D.W. Huang, B.T. Sherman, and R.A. Lempicki. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37: 1–13, 2009b.

A-M. Klein, B.E. Vaissière, J.H. Cane, I. Steffan-Dewenter, S.A. Cunningham, C. Kremen, and T. Tscharntke. Importance of pollinators in changing landscapes for world crops. *Proc Biol Sci*, 274:303–313, 2007.

K. Kulhanek, N. Steinhauer, K. Rennich, D.M. Caron, R.R. Sagili, J.S. Pettis, J.D. Ellis, M.E. Wilson, J.T. Wilkes, D.R. Tarpy, R. Rose, K. Lee, J. Rangel, and D. vanEngelsdorp. A national survey of managed honey bee 2014–2015 annual colony losses in the USA. *Journal of Apicultural Research*, 56:328–340, 2017.

Johan Larsson. *eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses*, 2018. URL https://cran.r-project.org/package=eulerr. R package version 4.0.0.

363 M. Laurent, P. Hendrikx, M. Ribiere-Chabert, and M-P. Chauzat. A pan-European
364 epidemiological study on honeybee colony losses 2012–2014. *Epilobee*, 2013:44, 2016.

365 M.I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion
366 for RNA-seq data with DESeq2. *Genome Biology*, 15:550, 2014.

367 E. Maori, N. Paldi, S. Shafir, H. Kalev, E. Tsur, E. Glick, and I. Sela. IAPV, a bee-affecting
368 virus associated with Colony Collapse Disorder can be silenced by dsRNA ingestion.
369 *Insect Mol Biol*, 18:55–60, 2009.

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

372 J.R. De Miranda, G. Cordoni, and G. Budge. The acute bee paralysis virus-Kashmir bee
373 virus-Israeli acute paralysis virus complex. *J Invertebr Pathol*, 103:S30–47, 2010.

374 D. Naug. Nutritional stress due to habitat loss may explain recent honeybee colony collapses.
375 *Biol Conserv*, 142:2369–2372, 2009.

376 P. Neumann and N.L. Carreck. Honey bee colony losses. *J Apicult Res*, 49:1–6, 2010.

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

379 G.D. Pasquale, M. Salignon, Y.L. Conte, L.P. Belzunces, A. Decourtye, A. Kretzschmar,
380 S. Suchail, J-L. Brunet, and C. Alaux. Influence of pollen nutrition on honey bee health:
381 Do pollen quality and diversity matter? *Plos One*, 8:e72016, 2013.

382 S.G. Potts, J.C. Biesmeijer, C. Kremen, P. Neumann, O. Schweiger, and W.E. Kunin. .
383 *Global pollinator declines: trends, impacts and drivers*, 25:345–353, 2010.

384 M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, and G.K. Smyth. limma
385 powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic*
386 *Acids Research*, 43(7):e47, 2015.

387 M.D. Robinson, D.J. McCarthy, and G.K. Smyth. edger: a bioconductor package for
388 differential expression analysis of digital gene expression data. *Bioinformatics*, 26:139–
389 140, 2010.

390 P. Rosenkranz, P. Aumeier, and B. Ziegelmann. Biology and control of Varroa destructor.
391 *J Invertebr Pathol*, 103:S96–S119, 2010.

392 T.H. Roulston and S.L. Buchmann. A phylogenetic reconsideration of the pollen starch-
393 pollination correlation. *Evol Ecol Res*, 2:627–643, 2000.

394 J.O. Schmidt. Feeding preference of Apis mellifera L. (Hymenoptera: Apidae): Individual
395 versus mixed pollen species. *J. Kans. Entomol. Soc.*, 57:323–327, 1984.

396 J.O. Schmidt, S.C. Thoenes, and M.D. Levin. Survival of honey bees, Apis mellifera
397 (Hymenoptera: Apidae), fed various pollen sources. *J. Econ. Entomol.*, 80:176–183, 1987.

398 M.Q. Shen, L.W. Cui, N. Ostiguy, and D. Cox-Foster. Intricate transmission routes and
399 interactions between picorna-like viruses (Kashmir bee virus and sacbrood virus) with
400 the honeybee host and the parasitic varroa mite. *J Gen Virol*, 86:2281–2289, 2005.

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

403 M. Spivak, E. Mader, M. Vaughan, and N.H. Euliss. The Plight of the Bees. *Environ Sci*
404 *Technol*, 45:34–38, 2011.

405 R.G. Stanley and H.F. Linskens. *Pollen: Biology, biochemistry, management*. Springer
406 Verlag, 1974.

407 F. Supek, M. Bošnjak, N. Škunca, and T. Šmuc. REVIGO summarizes and visualizes long
408 lists of Gene Ontology terms. *Plos ONE*, 6:e21800, 2011.

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

411 D. van Engelsdorp, J.D. Evans, C. Saegerman, C. Mullin, E. Haubruge, B.K. Nguyen,
412 M. Frazier, J. Frazier, D. Cox-Foster, Y. Chen, R. Underwood, D.R. Tarpy, and J.S.
413 Pettis. Colony collapse disorder: A descriptive study. *PLos One*, 4:e6481, 2009.

414 K.P. Weinberg and G. Madel. The influence of the mite Varroa Jacobsoni Oud. on the
415 protein concentration and the haemolymph volume of the brood of worker bees and
416 drones of the honey bee Apis Mellifera L. *Apidologie*, 16:421–436, 1985.

417 T.D. Wu, J. Reeder, M. Lawrence, G. Becker, and M.J. Brauer. GMAP and GSNAP
418 for genomic sequence alignment: Enhancements to speed, accuracy, and functionality.
419 *Methods Mol Biol*, 1418:283–334, 2016.

420 X. Yang and D. Cox-Foster. Effects of parasitization by Varroa destructor on survivorship
421 and physiological traits of Apis mellifera in correlation with viral incidence and microbial
422 challenge. *Parasitology*, 134:405–412, 2007.

423 X.L. Yang and D.L. Cox-Foster. Impact of an ectoparasite on the immunity and pathology
424 of an invertebrate: Evidence for host immunosuppression and viral amplification. *P Natl*
425 *Acad Sci USA*, 102:7470–7475, 2005.