IDENTIFICATION OF POTENTIAL TRANSCRIPTOMIC BIOMARKERS FOR VARROA RESISTANCE IN HONEY BEES (APIS MELLIFERA ANATOLIACA)

by HAŞİM HAKANOĞLU

Submitted to the Graduate School of Engineering and Natural sciences in partial fulfilment of the requirements for the degree of Master of Science

Sabancı University
July 2022

ABSTRACT

IDENTIFICATION OF POTENTIAL TRANSCRIPTOMIC BIOMARKERS FOR VARROA RESISTANCE IN HONEY BEES (APIS MELLIFERA ANATOLIACA)

HAŞİM HAKANOĞLU

Molecular Biology, Genetics and Bioengineering, MSc Thesis, July 2022

Thesis Supervisor: Asst. Prof. Dr. Christopher Mayack
Thesis Co-supervisor: Dr. Stuart Lucas

Keywords: Bee health, RNA-seq, DGE analysis, Varroa resistance, biomarkers.

Honey bees are experiencing heavy mortality due to environmental stressors, pesticide exposure, and diseases. Since honey bees contribute significantly to the economy due to their roles as essential pollinators and producers of honey among many other products, increased mortality rates are severely affecting the global and Turkish economy. The lead cause of honey bee mortality and health decline is infestation of honey bee colonies with the ectoparasite Varroa destructor, also known as the Varroa mite. Several honey bee lines have been selectively bred for traits linked to low levels of Varroa mites for generations to generate Varroa-resistant (VR) lines around the globe. Here, we examine disease prevalence, abundance, and gene expression profiles of honey bee lines that have been selectively bred locally for low levels of Varroa mites to identify transcriptomic biomarkers for Varroa resistance. We identified 278 significantly upregulated and 23 downregulated genes in VR colonies. The top differentially expressed gene was the CCAAT/enhancer-binding protein gamma (LOC726550). Next, we queried significantly upregulated genes against STRING to build a protein-protein interaction network. Network topological features revealed its scale-free property, indicating that most genes are connected to a small number of the highly connected genes called hubs. Further inspection of the network revealed that the top two hubs, papilin (LOC413021) and heat shock protein cognate 4 (Hsc70-4), have high influence over the flow of information in the network, confirming their importance. Therefore, we suggest LOC413021, Hsc70-4, and LOC726550 as potential transcriptomic biomarkers for Varroa resistance.

ÖZET

BAL ARILARINDA (APIS MELLIFERA ANATOLIACA) VARROA DİRENCİ İÇİN POTANSİYEL TRANSKRİPTOMİK BİYOBELİRTİÇLERİ TANIMLAMASI

HAŞİM HAKANOĞLU

Moleküler Biyoloji, Genetik ve Biyomühendislik, Yüksek Lisans Tezi, Temmuz 2022

Tez Danışmanı: Dr. Öğr. Üyesi Christopher Mayack
Tez Eş-Danışmanı: Dr. Stuart Lucas

Anahtar Kelimeler: Arı sağlığı, RNA-seq, DGE analizi, Varroa direnci, biyobelirteçler.

Bal arıları, çevresel stres faktörleri, pestisit maruziyeti ve hastalıklar nedeniyle ağır ölüm yaşamaktadır. Bal arıları, temel tozlayıcı ve bal gibi birçok ürünün üreticisi rolleri nedeniyle ekonomiye önemli katkı sağladığından, artan ölüm oranları dünya ve Türkiye ekonomisini ciddi şekilde etkilemektedir. Bal arısı sağlığının azalmasının ve ölümlerinin başlıca nedeni, Varroa akarı olarak da bilinen ektoparazit Varroa yıkıcısı ile bal arısı kolonilerinin istilasıdır. Dünya çapında Varroa dirençli (VR) hatları oluşturmak için nesiller boyunca düşük Varroa akarları seviyelerine bağlı özellikler için bazı bal arısı hattı seçici olarak yetiştirilmiştir. Burada, Varroa direnci için transkriptomik biyobelirteçleri belirlemek için düşük Varroa akarları seviyeleri için yerel olarak seçici olarak yetiştirilen bal arısı hatlarının hastalık prevalansı, bolluğu ve gen ekspresyon profillerini incelemekteyiz. VR kolonilerinde 278 önemli ölçüde yukarı regüle ve 23 aşağı regüle gen tanımlamıştık. Diferansiyel olarak eksprese edilen en önemli gen, CCAAT/arttırıcı-bağlayıcı protein gama (LOC726550) olduğunu anlaşılmıştır. Daha sonra, protein-protein etkileşimleri (PPI) ağı oluşturmak için, STRING'e karşı önemli ölçüde yukarı regüle edilmiş genleri sorgulamıştık. Ağın topolojik özellikleri, ölçeksiz özelliğini ortaya çıkarmıştı; bu, çoğu genin, hub olarak adlandırılan az sayıda yüksek düzeyde bağlantılı gene bağlı olduğunu gösterir. Ağın daha fazla incelenmesi, üst iki merkezlerin, papilin (LOC413021) ve ısı şok protein kognatı 4'ün (Hsc70-4) ağdaki bilgi akışı üzerinde yüksek etkiye sahip olduğunu ortaya çıkarmış ve bunların önemini doğrulamıştır. Bu nedenle, Varroa direnci için potansiyel transkriptomik biyobelirteçler olarak LOC413021, Hsc70-4 ve LOC726550'yi önermekteyiz.

ACKNOWLEDGEMENTS

I would like to first express my gratitude to my family, to whom I owe a great debt. I cannot, and frankly do not even want to, imagine how life would have been without them and their support.

I would like to thank my advisor Dr. Christopher Mayack for constantly supporting and giving me, a student with an entirely different background, a space and allowing me to learn and experiment. Additionally, I would like to thank my co-advisor Dr. Stuart Lucas for his helping me along my yet-to-end journey to become a bioinformatician. I am also grateful for Prof. Dr. İbrahim Çakmak for his constructive feedback and being an integral part of this project.

I would like to thank my previous and current (soon-to-be previous) lab mates Saleh Ghanem, İrem Akülkü, Mert Özle, and Elif Kıvrak for providing an entertaining and supportive environment. Lastly, I would like to thank the undergraduate students who worked with me on this project: Burak Onbaşı, Dalya Koprulu, Rahaf Elshanawany, and Rukiye-ayshe Egeli.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

°C	Degree Celsius
A230	Absorbance at 230 nm
A260	Absorbance at 260 nm
A280	Absorbance at 280 nm
ABPV	Acute bee paralysis virus
AFB	American foulbrood
ALPV	Aphid lethal paralysis virus
BAM	Binary alignment and map
BeeMLV	Bee Macula-like virus
BH	Benjamini-Hochberg
BQCV	Black queen cell virus
BSRV	Big Sioux River virus
C/EBP-γ	CCAAT-enhancer-binding protein gamma
CB	Chalkbrood
CB	betweenness centrality
CBPV	Chronic bee paralysis virus
CD	Degree centrality
cDNA	Complementary DNA
DEG	Differentially expressed gene
DGE	Differential gene expression
DNA	Deoxyribonucleic acid
dT	Deoxythymidine
dUTP	Deoxyuridine triphosphate
DWV	Deformed wing virus
EDTA	Ethylenediaminetetraacetic acid
EFB	European foulbrood
FDR	False discovery rate
GLM	Generalized linear models
GLMM	Generalized mixed linear models
GO	Gene Ontology
Hsc70-4	Heat shock protein cognate 4

IAPV	Israeli acute paralysis virus
IQR	interquartile range
ITS	Internal transcribed spacer
KBV	Kashmir bee virus
KEGG	Kyoto Encyclopedia of Genes and Genomes
KV	Kakugo virus
LFC	Log2 fold change
LRT	Likelihood ratio test
LSV	Lake Sinai virus
mL	Milliliter
mM	Millimolar
mRNA	Messenger ribonucleic acid
NaCl	Sodium chloride
ng	Nanogram
NO	Nitric oxide
NOS	Nitric oxide synthase
p	P-value
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
PE150	Paired-end 150 bp reads
Poly(A)	Repetitive adenine nucleotides sequence
PPI	Protein-protein interaction
QC	Quality control
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
RNA-seq	Ribonucleic acid sequencing
RpS5	Ribosomal protein S5
rRNA	Ribosomal ribonucleic acid
SAM	Sequence alignment and map
SB	Stonebrood
SBPV	Slow bee paralysis virus
SBV	Sacbrood virus

SDS	Sodium lauryl sulfate
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
Tris-HCl	Trisaminomethane hydrochloride
U	Unit
VDV1	Varroa destructor virus-1
VR	Varroa resistant
VRH	Varroa resistant with high load
VRL	Varroa resistant with low load
VS	Varroa sensitive
VSH	Varroa sensitive with high load
VSL	Varroa sensitive with low load
W	Wilcoxon rank-sum test statistic
µL	Microliter
µM	Micromolar
χ^2	Chi-square test statistic

1. INTRODUCTION

1.1. Economic Value of Honey Bees

Honey bees provide an indispensable service to the ecosystem as pollinators. Pollination is critical for crop production, maintenance of the ecosystem, wild plant reproduction, and food security. Honey bees are also a major source of honey, as well as beeswax, propolis, and royal jelly. The first domesticated bee species is the western honey bee, *Apis mellifera*. It is the most commonly managed and economically valuable pollinator in the world (Klein et al., 2007; Potts et al., 2016). *Apis mellifera* is native to Europe, Asia (including the Middle East), and Africa, and was later introduced to other continents. Turkey, a bridge between Europe and Asia, encompasses a diverse ecosystem with diverse organisms including, but not limited to, honey bees. Examples representing the diverse pool of honey bee species found in Turkey are *A. m. anatoliaca*, *A. m. caucasica*, *A. m. meda*, and *A. m. syriaca* (Kandemir et al., 2000, 2006). It is estimated that beekeeping generates a contribution of $542 million per year to the Turkish economy (Turkish Statistical Institute [TUIK]). Moreover, the contribution of pollination by bees is estimated to be 10-15-fold the amount generated from beekeeping, or about $5.4 billion (TUIK, 2015). Turkey reportedly has more than 8 million beehives (Food and Agriculture Organization of the United Nations [FAO], 2020; Figure 1A), ranking third in the world in terms of the number of beehives after India and Mainland China (FAO, 2020). In addition, Turkey ranks second globally in honey production, producing approximately 100,000 tons (FAO, 2020; Figure 1B). However, beekeeping in Turkey has yet to reach its full potential as its honey production per hive is substantially lower than countries such as Canada, Mainland China, Brazil, and the US (FAO, 2020; Figure 1C). This is linked to a decline in bee health in Turkey. (Seven Çakmak & Çakmak, 2016).

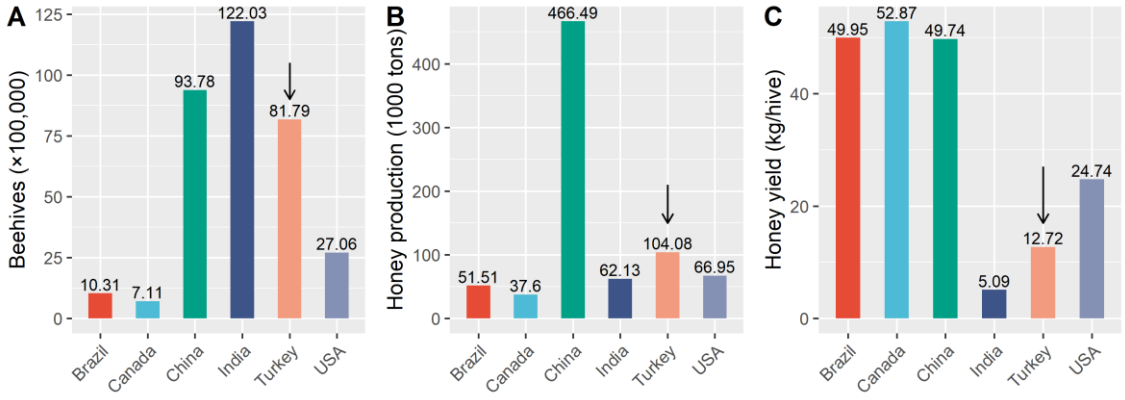

Figure 1. Number of beehives (A), honey production in thousand metric tons (B), and honey yield in kilograms per hive (C) of representative countries in 2020. Data was retrieved from the Food and Agriculture Organization of the United Nations (FAO).

1.2. Global Decline in Honey Bee Health: A Multifactorial Complex Problem

With the rise of industrial agriculture to meet the demands of the growing world population, the reliance on pesticides is ever-increasing. A recent study shed light on a major consequence of intensive use of pesticides: the contamination of honey samples from 27 countries, including Turkey, with 92 different pesticides (El-Nahhal, 2020). Pesticides target pests, such as insects, that are considered harmful for crops. However, pesticides also have lethal and sublethal effects on non-target groups such as honey bees, thus contributing to the decline in honey bee health (Potts et al., 2010; Vanbergen & the Insect Pollinators Initiative, 2013). Examples of sublethal effects on honey bees are reduction in sperm viability/quality and queen failure, which in turn are linked to colony losses (Pettis et al., 2016). High sperm mortality and number of defective sperm cells were shown to be a result of chronic exposure to a common insecticide and different bee medications commonly used for beekeeping in Turkey at field-realistic levels (Ben Abdelkader et al., 2021). In addition to pesticide exposure, massive mortalities in the honey bee populations worldwide are due to loss of foraging habitat, and disease prevalence (Potts et al., 2010; Vanbergen & the Insect Pollinators Initiative, 2013). In the US, the number of beehives has declined from 4.6 million in 1970 to 2.8 million in 2019, a 40% decline (FAO, 2019). In Europe, the number of beehives declined from 2.1 million in 1970 to 1.6 million in 2019, a 24% decline (FAO, 2019). In contrast, Turkey has witnessed a constant increase in the number of beehives, going from 1.8 million in 1970

to 8.1 million in 2019, a 450% increase (FAO, 2019). However, as mentioned previously, Turkey has yet to reach its full potential due to poor management and the prevalence of pathogens such as the parasitic mite *Varroa destructor*, deformed wing virus, the microsporidian fungi *Nosema* spp. (causative agent of Nosema disease), the fungi *Aspergillus* spp. (causative agents of stonebrood disease) and *Ascosphaera apis* (causative agent of chalkbrood disease), and the bacteria *Paenibacillus larvae* (causative agent of American foulbrood disease) and *Melissococcus plutonius* (causative agent of European foulbrood disease) foulbrood diseases causing major honey bee colony losses in some regions (Seven Çakmak & Çakmak, 2016).

1.3. Parasites and Pathogens Affecting Honey Bee Health

1.3.1. Varroa destructor

The most prominent parasite that infests honey bee colonies and is considered the leading cause to honey bee colony losses worldwide is the ectoparasite *Varroa destructor*, also known as the *Varroa* mite (Rosenkranz et al., 2010). *Varroa* mites reproduce only in honey bee colonies, and feed on the fat body tissues of both immature and mature honey bees (Ramsey et al., 2019). This impairs immunity and pupal development and reduces tolerance to pesticides (Ramsey et al., 2019). Impairment in immune function of honey bees complements the fact that *Varroa* mites are vectors for several viruses that infect honey bees, which ultimately results in drastic reduction in honey production and lifespan of honeybees. Nation-wide qualitative assessments on the prevalence of honey bee diseases and pathogens in Turkey revealed that *Varroa* mites and/or damage caused by *Varroa* mites are highly prevalent in the Aegean Region (Beyazıt et al., 2012; Çağirgan et al., 2022; Kalayci et al., 2020; Özbilgin et al., 1999), Black Sea Region (Kekeçoğlu et al., 2013; Okur Gumusova et al., 2010; Parlakay & Esengün, 2005; Yaşar et al., 2002; Yilmaz et al., 2018), Central Anatolia Region (Bayrakal et al., 2020; Tunca & Çimrin, 2012), Eastern Anatolia Region (Aydin, 2012; Aydın, 1998; Balkaya, 2016; Erkan, 1998; Kaftanoğlu et al., 1995; Karapınar et al., 2018; Kesik et al., 2022; Önk & Kılıç, 2014; Şimşek, 2005), Marmara Region (Borum & Ülgen, 2010; Çakmak, Aydın, et al., 2003;

D. Muz & Muz, 2017; Sıralı & Doğaroğlu, 2005), and Mediterranean Region (D. Muz & Muz, 2009; M. N. Muz et al., 2012; Sahinler & Gul, 2005; Yalçınkaya & Keskin, 2010).

1.3.2. Stonebrood

Stonebrood (SB) is a honey bee disease caused by *Aspergillus*, a genus consisting of more than 200 fungal species that are ubiquitously present in soil, of which 40 are opportunistic pathogens that can infect both larvae and adults (K. Foley et al., 2014; Varga & Samson, 2008), as well as humans (Dagenais & Keller, 2009). When a colony is weak, SB infection occurs when larvae ingest fungal spore-contaminated food that is transmitted into and within colonies by forager and nurse bees, respectively. Once ingested, those *Aspergillus* spp. kill and transform larvae into hard mummies. Two of the *Aspergillus* spp., *A. flavus* and *A. fumigatus*, are the primary pathogens causing SB, with *A. flavus* being the most virulent and prolific strain (K. Foley et al., 2014; Nasri et al., 2015). Three *Aspergillus* species—*A. flavus*, *A. fumigatus*, and *A. niger*—are present in adult bees from colonies showing clinical symptoms of SB, with *A. niger* being the most abundant, followed by *A. flavus* then *A. fumigatus* (Shoreit & Bagy, 1995). On the other hand, in adult bees from colonies not showing clinical symptoms of SB, only *A. niger* and *A. flavus* are present, with *A. niger* being significantly more abundant than *A. flavus* (Shoreit & Bagy, 1995). Spores from *A. flavus*, *A. fumigatus*, and *A. niger* can also spread through the air inside colonies, with *A. niger* and *A. flavus* spores being the most abundant airborne spores inside colonies not showing clinical symptoms of SB and ones that are, respectively (Shoreit & Bagy, 1995). SB is a rare disease and it has only been reported in three regions of Turkey: Central Anatolia (Bayrakal et al., 2020), Eastern Anatolia (Balkaya, 2016), and Marmara region (Dümen et al., 2013; Sıralı & Doğaroğlu, 2005).

1.3.3. Chalkbrood

The fungus *Ascosphaera apis* is the causative agent of chalkbrood (CB), a disease that exclusively kills honey bee larvae and leaves the brood with a mummified chalky appearance (Spiltoir, 1955; Spiltoir & Olive, 1955). CB, analogous to SB, is a stress-related disease because its outbreak and severity depend on multiple stressors interacting

together to weaken colonies (Evison, 2015). Similar to SB, *A. apis* infection occurs in weak colonies when larvae ingest fungal spore-contaminated food that is transmitted into and within colonies by forager and nurse bees, respectively (Aronstein & Murray, 2010). *A. apis* spores can accumulate in all honey bee products and parts of the colony and remain viable for at least 15 years (Aronstein & Murray, 2010; Evison, 2015). Therefore, CB infections can spread between colonies when re-using contaminated materials. CB infections do not necessarily lead to colony losses, they however lead to reduction in the workforce and hence a reduced honey production (Evison, 2015). Nation-wide qualitative assessments on the prevalence of CB in Turkey revealed that the disease is present in the Aegean (Beyazıt et al., 2012; Özbilgin et al., 1999), Black Sea (Yaşar et al., 2002), Central Anatolia (Bayrakal et al., 2020; Tunca & Çimrin, 2012), Eastern Anatolia (Balkaya, 2016; Kaftanoğlu et al., 1995), and Marmara region (Borum & Ülgen, 2010; Çakmak, Aydın, et al., 2003; Dümen et al., 2013; Sıralı & Doğaroğlu, 2005).

1.3.4. Nosema

Nosema is an adult bee disease caused by two spore-forming microsporidian intracellular parasites, *Nosema apis* and *Nosema ceranae*. *N. ceranae* is more virulent and has displaced *N. apis* in most of the world (Chen et al., 2008; Higes et al., 2013; Klee et al., 2007; Paxton et al., 2007). *N. ceranae* can cause bee mortality and it has been implicated with colony losses globally (Higes et al., 2008, 2009). It is also associated with suppressed immune function, foraging behavior, dysregulated metabolism and oxidative stress (Burnham, 2019). Nation-wide qualitative assessments on the prevalence of *Nosema* spp. in Turkey revealed that the disease is present in all regions (Ütük et al., 2011).

1.3.5. American and European foulbrood

American foulbrood (AFB) and European foulbrood (EFB) are caused by the Gram-positive and spore-forming bacterium *Paenibacillus larvae*, and the Gram-positive and lanceolate coccus (sometimes pleomorphic and rod-like) *Melissocccus plutonius*, respectively (Forsgren, 2010; Genersch, 2010). As their names suggest, they are both exclusively brood diseases, affecting larvae only when *P. larvae* spores (in the case of

AFB) or *M. plutonius* cells are ingested. Adult bees are not susceptible to the two pathogens (Forsgren, 2010; Genersch, 2010). AFB is a highly virulent disease that can spread widely (Genersch, 2010). Its clinical symptoms include patchy brood with darkened, sunken, greasy cell caps (Genersch, 2010). On the other hand, EFB is not as virulent as AFB. Its clinical symptoms include uncapped cells containing discolored dead larvae (Forsgren, 2010). Nation-wide qualitative assessments on the prevalence of foulbrood diseases in Turkey revealed that the AFB is present in the Aegean (Beyazıt et al., 2012), Black Sea (Kekeçoğlu et al., 2013; Parlakay & Esengün, 2005; Yaşar et al., 2002), Central Anatolia (Tunca & Çimrin, 2012), Eastern Anatolia (Balkaya, 2016; Kaftanoğlu et al., 1995), Marmara (Dümen et al., 2013; Sıralı & Doğaroğlu, 2005), and Mediterranean region (M. N. Muz et al., 2012; Sahinler & Gul, 2005; Yalçınkaya & Keskin, 2010).

1.3.5. Viruses

Varroa mites act as active and passive vectors of multiple viruses that infect honey bees with and without *Varroa* mite-mediated transmission, all of which have RNA as their genetic material. Viruses that can be vectored in *Varroa* mites are Deformed wing virus (DWV) (Di Prisco et al., 2016; Gisder et al., 2009; Nazzi et al., 2012; Wilfert et al., 2016; Yue & Genersch, 2005), Kashmir bee virus (KBV) (Chen et al., 2004; de Miranda et al., 2010), Israeli acute paralysis virus (IAPV) (de Miranda et al., 2010; Di Prisco et al., 2011), Chronic bee paralysis virus (CBPV) (Celle et al., 2008), Acute bee paralysis virus (ABPV) (de Miranda et al., 2010), Bee Macula-like virus (BeeMLV) (De Miranda et al., 2015), Lake Sinai virus (LSV) (Daughenbaugh et al., 2015), Varroa destructor virus-1 (VDV-1) (Ongus et al., 2004), Black queen cell virus (BQCV) (Chantawannakul et al., 2006; Mondet et al., 2014), Sacbrood virus (SBV) (Chantawannakul et al., 2006; Shen et al., 2005), Kakugo virus (KV) (Fujiyuki et al., 2006), and Slow bee paralysis virus (SBPV) (Carreck et al., 2010; Santillán-Galicia et al., 2014). Recently, two additional RNA viruses, aphid lethal paralysis virus (ALPV) and Big Sioux River virus (BSRV) (Runckel et al., 2011), were shown to infect honey bees. All of these viruses may be maintained as covert (lack of symptoms with minimal or no effect on performance and lifespan) or overt infections (cause symptomatic infections or death) (McMenamin & Flenniken, 2018).

1.4. Breeding Efforts for Resistance Against *Varroa* Mites

In efforts to combat the *Varroa* mite infestation and its negative impact on colony losses, several honey bee lines have been selectively bred for their low levels or absence of varroa mite infestation for generations to generate *Varroa*-resistant (VR) honey bee colonies around the globe (Bienefeld, 2016; Büchler, 1994; Harbo & Hoopingarner, 1997). The earliest observations of true resistance against *Varroa* mites emerging from natural selection were observed in Avignon, France (Conte et al., 2007), and on the island of Gotland, Sweden (Fries et al., 2006). Since then, generating VR honey bee lines became the primary research goal. To that end, several programs were established around the globe to produce resistance against varroa mites and uncover its mechanism (Rinderer et al., 2010). Reduction in *Varroa* mite infestation was linked to infertility of mites on the brood (Locke et al., 2014), unattractive brood to the mite (Nazzi & Le Conte, 2016), rapid brood development time (Calderón et al., 2010), removal of mite-infested brood (Villa et al., 2017), and grooming behavior (Pritchard, 2016). However, multiple factors, such as environmental conditions, compromise the link between such brood/adult bee characteristics and *Varroa* resistance (Cakmak & Fuchs, 2013). Recent breeding programs are targeting a specific type of hygienic behavior called *Varroa*-sensitive hygiene, a phenotypical trait that can be passed down from generation to generation whereby in-hive honey bees detect *Varroa* mites on larvae and remove them (Rinderer et al., 2010). However, *Varroa*-sensitive hygiene may not be sufficient for developing resistance against *Varroa* mites (Çakmak, 2010), indicating that resistance against *Varroa* mites remains largely unknown. Molecularly, global changes in gene expression and functional pathways in honey bees associated resistance to *Varroa* mites have not been explored thoroughly, especially the ones that are not associated with the traits linked to reduced *Varroa* mite infestation. Identification of *Varroa* resistance genes, regardless of brood/adult bee characteristics, will contribute immensely to the identification of biomarkers that can be used to select VR honey bee lines.

1.5. Molecular Mechanisms Conferring Resistance Against *Varroa* Mites

A wealth of studies have investigated the molecular mechanisms behind phenotypic traits that are linked to resistance against *Varroa* mites, particularly *Varroa*-sensitive hygiene (reviewed by Kaskinova et al., 2020). Several of these suggest that olfaction, the sense of smell, plays a role in detecting *Varroa* mites and induce shifts in the behavioral states of in-hive honey bees to remove the mites. (Le Conte et al., 2011; Mondet et al., 2015; Oxley et al., 2010; Spivak et al., 2003). Therefore, most of the studies that identified genes influencing such behaviors that are linked to resistance against *Varroa* mites investigated them on the level of brain or antenna (Le Conte et al., 2011; Mondet et al., 2015; Spivak et al., 2003). However, the changes in gene expression in VR honey bee lines that are not selected for any of the traits linked to *Varroa* mite-resistance have not been investigated. In addition, none of the studies examined transcriptomic differences in whole bees instead of specific tissues. This is crucial because differences in gene expression could arise in response to various pathogenic covert and overt infections that cannot be detected in brain or antennae. In addition, *Varroa* resistance biomarkers that can be measured from whole-bee homogenates are of greater utility because there is no need to dissect specific parts of the bee out before RNA extraction. This study is the first to explore the difference in gene expression levels between *Varroa* mite-sensitive (VS) and *Varroa* mite-resistant Anatolian honey bees (*Apis mellifera anatoliaca*) that were only selected for low levels of mites. Lastly, to the best of our knowledge, this work is the first to integrate comprehensive disease screenings to disentangle their effect from the effect of *Varroa* mite resistance and utilize a network-based approach with differential gene expression analysis for biomarker discovery.

2. AIMS

This study has five major aims, the first is to determine disease and pathogen prevalence and levels in colonies located in the southern Marmara region, the second is to determine whether diseases and pathogens are less prevalent and/or abundant in *Varroa*-resistant (VR) colonies, the third is to elucidate relationships between diseases and pathogens, the fourth is to identify genes, pathways and gene sets associated with *Varroa* resistance only, and the fifth aim is to identify biomarkers for *Varroa* resistance using a network-based approach coupled with differential gene expression analysis.

3. MATERIALS AND METHODS

3.1. Origin, Sample Collection, and Quantification of Varroa Mite Levels

Varroa-resistant (VR) honey bee colonies were established from an isolated Anatolian honey bee (*Apis mellifera anatoliaca*) population on Marmara Island, Turkey, that have been selected for the past 15 years for low levels of *Varroa* mite infestation by the Beekeeping Development-Application and Research Center (AGAM), Bursa Uludağ University, Turkey, under the supervision of Prof. Dr. İbrahim Çakmak. Prior to each selection round, varroa mite infestation level was determined using the sugar roll method each summer (Çakmak et al., 2011). The method involves placing around 300 collected nurse bees in a mason jar where they are coated with powdered sugar for 2 minutes. The jar is then shaken vigorously for 3 minutes, causing the mites to dislodge and fall through a mesh screen for collection on a white paper plate. The mites are then counted to determine the varroa mite load with a 94% sensitivity, that is, the method will be able to identify 94% of *Varroa* mite infested colonies (Çakmak et al., 2011). In addition, this method recovers 94% of *Varroa* mites in infestation colonies, that is, the counts are 94% accurate (Çakmak et al., 2011). The selection is maintained via natural mating. *Varroa*-sensitive (VS) honey bee colonies originated from an unselected Anatolian honey bee (*A. m. anatoliaca*) population local to the Marmara region of Turkey.

Nurse bees were collected, in two batches, from 42 VR colonies located on the Marmara Island, and 73 VS colonies from located on Marmara Island, Mustafakemalpaşa, Çınarcık, Yalova, and Karacabey from the Marmara region. VR colonies are maintained without any treatments, while VS colonies were periodically (at different seasons) treated with miticides and organic acids for *Varroa* mite infestation, and antibiotics for bacterial and fungal diseases. The first batch of colonies was sampled in December 2020 and

consisted of bees from 10 VR and 9 VS colonies located on the island. The second batch of colonies was sampled in July 2021 and consisted of the rest of the colonies. Each batch of collected bees were placed in wooden cages, where each cage was comprised of 150-300 bees, and transported to Sabancı University. Upon arrival, they were flash-frozen in liquid nitrogen and then stored at -80 °C. Prior to sampling, the varroa mite infestation level was determined for each colony using the sugar roll method. Samples were then stratified based on *Varroa* mite counts. VR colonies were classified as *Varroa*-resistant with low load (VRL) if their *Varroa* mite counts ≤ 3 and *Varroa*-resistant with high load (VRH) otherwise. Similarly, VS colonies were classified as *Varroa*-sensitive with low load (VSL) if their *Varroa*-mite counts < 10 and *Varroa*-sensitive with high load (VSH) otherwise.

3.2. Sample Processing

Flash-frozen honey bees were then transferred into one or more 50 mL falcon tubes (each tube can house up to approximately 150 honey bees) and were either homogenized immediately or stored at -80 °C prior to homogenization. Whole-bee homogenates were made by macerating 100-150 frozen honey bees per sampled colony in 15 mL of DEPC-treated water using the Covidien Precision™ Disposable Tissue Grinder System (Medtronic, Ireland). The homogenization was done in three rounds for each 50 mL falcon tube; each round consisted of macerating the content of one third of a tube in 5 mL of DEPC-treated water. 50 mL Falcon tubes were used to collect 15 mL of liquid from macerated honey bees. 6 Aliquots of 150 µL were made from each 15 mL of honey bee homogenate. Each aliquot was used to screen for one of the 5 pathogens that infect honey bees, and the remaining aliquot was allocated for RNA extraction.

3.3. Disease Screening

3.3.1. Stonebrood (SB) Screening

Genomic DNA for SB screening was extracted from whole-bee homogenate aliquots using a custom-made lysis buffer (300 µL of buffer containing 200 mM Tris-HCl (pH 7.5), 25 mM EDTA, 0.5% w/v SDS, and 250 mM NaCl per 150 µL of homogenate) followed by a standard phenol:chloroform (1:1) extraction (Nasri et al., 2015). Assessment of DNA yield and purity was done using the NanoDrop 1000c spectrophotometer (Thermo Fisher Scientific, US). Amplification of the β-tubulin gene of the *Aspergillus* species that cause SB, and the ribosomal protein S5 (RpS5) gene of *Apis mellifera* was performed using PCR. RpS5 is a honey bee house-keeping gene in which the expression is stable across different tissues and seasons (Jeon et al., 2020), thus allowing for monitoring of extraction failures or PCR amplification inhibition. All PCR amplifications were performed using 2X Taq PCR MasterMix (abm, Canada) in 25 µL reactions containing 1 µL of each primer (10 µM) targeting either *Aspergillus* β-tubulin or *A. mellifera* RpS5 and at least 100 ng of DNA per reaction. Reaction conditions were set as described by Nasri, et al. (2015). PCR products, a positive control, and a 100bp Opti-DNA Marker (abm, Canada) were mixed with Gel Loading Dye, Purple (6X), no SDS (NEB, US) and separated by electrophoresis (100 V, 30 min), on 1.5% agarose gels, stained with GelRed Nucleic Acid Stain (10000x in water) (Biotium, US). Gel visualization was obtained using the Bio-Rad Gel Doc EZ Gel Documentation System (Bio-Rad, US).

Table 1. List of primers used for the amplification of *Aspergillus* β-tubulin and *A. mellifera* RpS5 genes.

Primer	**Primer sequence (5'-3')**	**Product length (bp)**		**Ref.**
SB-F	GGTAACCAAATCGGTGCTGCTTTC	*A. fumigatus*	549	Nasri et al., 2015
SB-R	ACCCTCAGTGTAGTGACCCTTGGC	*A. flavus*	550	
		A. niger	531	
		A. terreus	564	
		A. clavatus	562	
		A. nidulans	475	
RpS5-F	AATTATTTGGTCGCTGGAATTG	115		

RpS5-R	TAACGTCCAGCAGAATGTGGTA	Thompson et al., 2007

3.3.2. Chalkbrood (CB) Screening

Genomic DNA for CB screening was extracted from whole-bee homogenate aliquots using the DNeasy Plant Mini Kit (Qiagen, Germany) according to the manufacturer's protocol (Jensen et al., 2012). Assessment of DNA yield and purity was performed using the NanoDrop 1000c spectrophotometer (Thermo Fisher Scientific, US). Amplification of the internal transcribed spacer (ITS) region of the nuclear ribosomal repeat unit of the fungus *Ascosphaera apis* that causes CB, and the ribosomal protein S5 (RpS5) gene of *Apis mellifera* was performed using PCR. All PCR amplifications contained 2X Taq PCR MasterMix (abm, Canada) in 25 µL reactions consisting of 1 µL of each primer (10 µM) targeting either *A. apis* ITS or *A. mellifera* RpS5 (Table 2) and at least 100 ng of DNA per reaction. Reaction conditions were set as described in the COLOSS BEEBOOK: fungal brood diseases (Jensen et al., 2013). PCR products were separated by electrophoresis and visualized as described in section 1.2.

Table 2. List of primers used for the amplification of *A. apis* ITS region of the nuclear ribosomal repeat unit and *A. mellifera* RpS5 genes.

Primer	**Primer sequence (5'-3')**	**Product length (bp)**	**Ref.**
CB -F	TGTCTGTGCGGCTAGGTG	648	James & Skinner, 2005
CB-R	CCACTAGAAGTAAATGATGGTTAGA		
RpS5-F	AATTATTTGGTCGCTGGAATTG	115	Thompson et al., 2007
RpS5-R	TAACGTCCAGCAGAATGTGGTA		

3.3.3. American Foulbrood (AFB) Screening

Genomic DNA for AFB screening was extracted by heating the honey bee homogenate (modified from Govan et al., 1999). Briefly, 800 µL of DEPC-treated water was added to a whole-bee homogenate aliquot and centrifuged at 800 g for 10 min. 200 µL of suspension from each aliquot were incubated at 95 °C for 15 min with lids open, then centrifuged at 5000 g for 5 min. For 51 samples, AFB screening was done using the BactoReal Kit American Foulbrood 1.1 (Ingenetix GmbH, Austria), a probe-based qPCR assay that detects the 16S rRNA gene of *P. larvae*, according to the manufacturer's protocol. Reactions were carried out using TaqProbe 2X qPCR MasterMix (abm, Canada) in 20 µL reactions containing 5 µL of supernatant from heated honey bee homogenate and ran in the LightCycler® 480 System (Roche Diagnostics, Roche, Switzerland). For the rest of the samples, the screening was done using a SYBR Green-based qPCR for the amplification and detection of target region in the 16S rRNA gene of *P. larvae* (Rossi et al., 2018). Reactions were carried out using BrightGreen 2X qPCR MasterMix (abm, Canada) in 20 µL reactions containing 5 µL of supernatant from heated honey bee homogenate and 1 µL of each primer (5 µM) (Table 3). qPCR amplifications and detections were done in the LightCycler® 480 System (Roche Diagnostics, Roche, Switzerland) with a program consisting of initial denaturation at 94 °C for 4 min, followed by 45 cycles of denaturation at 95 °C for 15 s and annealing and signal acquisition at 56 °C for 10 s. All screenings involved no template negative controls and positive controls.

Table 3. List of primers used for the amplification and detection of *P. larvae* 16S rRNA gene.

Primer	Primer sequence (5'-3')	Product length (bp)	Ref.
AFB-F	TTCGGGAGACGCCAGGTTA	131	Rossi et al., 2018
AFB-F	CTTTCATGACTTCTTCATGCGAAG		

3.3.4. European Foulbrood (EFB) Screening

Genomic DNA for EFB screening was extracted as described in the COLOSS BEEBOOK: European foulbrood (Forsgren et al., 2013) using a custom-made grinding buffer (500 µL of grinding buffer containing 0.25 g guanidine thiocyanate, 26.5 µL 1M Tris-Cl (pH 7.6), and 26.5 µL 0.2 M EDTA per 150 µL of homogenate) followed by the

DNeasy Plant Mini Kit (Qiagen, Germany) according to the manufacturer's protocol. Assessment of DNA yield and purity was done using the NanoDrop 1000c spectrophotometer (Thermo Fisher Scientific, US). DNA samples with A260/230 < 1.7 were subjected to PCR amplification of *A. mellifera RpS5* in duplicates and products were visualized after being separated by electrophoresis on 2% agarose gels as described in section 1.2. Samples that showed the *RpS5* amplicon in both of their PCR duplicates or passed the QC using the NanoDrop 1000c spectrophotometer were subjected to EFB screening. Samples were subjected again to DNA extraction, quantity and quality assessment, and A. mellifera RpS5 amplification using PCR if one or none of their PCR duplicates produced RpS5 amplicons. EFB screening was done using BactoReal European Foulbrood Kit (Ingenetix GmbH, Austria), a probe-based qPCR assay that detects the 16S rRNA gene of *M. plutonius*. According to the manufacturer's protocol, reactions were carried out using TaqProbe 2X qPCR MasterMix (abm, Canada) in 20 µL, containing 5 µL of template. All screenings involved no template negative controls and positive controls supplied by the kits.

3.3.5. Nosema screening and semi-quantification of Nosema infection

Genomic DNA extraction for Nosema screening was done using the HBRC method that relies on a custom-made buffer (300 µL of 3 mM hexadecyltrimethylammonium bromide (CTAB), 5 mM Tris-HCl, 1 mM EDTA, and 1.1 M NaCl per 150 µL of homogenate) and proteinase K, followed by a standard phenol:chloroform (1:1) extraction (Hamiduzzaman et al., 2010). Assessment of DNA yield and purity was performed using the NanoDrop 1000c spectrophotometer (Thermo Fisher Scientific, US). Co-amplification of 16S rRNA gene of *N. apis* and *N. ceranae,* and RpS5 of *A. mellifera* was obtained using PCR (Hamiduzzaman et al., 2010). All PCR amplifications were performed using 2X Taq PCR MasterMix (abm, Canada) in 25 µL reactions containing 1 µL of each primer (10 µM) (Table 4) and at least 100 ng of DNA per reaction. The PCR products were separated by electrophoresis on 1% agarose gels as described in section 1.2. The pixel intensity of amplified bands was measured using ImageJ v1.53k (Schneider et al., 2012). The ratio of *N. apis* 16S rRNA band intensity to the *A. mellifera* RpS5 band intensity was calculated to semi-quantify the relative abundance of *N. apis* for each sample. Similarly, the ratio of

N. ceranae 16S rRNA band intensity to the *A. mellifera* RpS5 band intensity was calculated to semi-quantify the relative abundance of *N. ceranae* for each sample.

Table 4. List of primers used for the amplification and detection of *N. ceranae* 16S rRNA, *N. apis* 16S rRNA gene, and *A. mellifera* RpS5 genes.

Primer	Primer sequence (5'-3')	Product length (bp)	Ref.
CER-F	CGGCGACGATGTGATATGAAAATATTAA	218	Hamiduz zaman et al., 2010
CER-R	CCCGGTCATTCTCAAACAAAAAACCG		
APIS-F	GGGGGCATGTCTTTGACGTACTATGTA	321	Hamiduz zaman et al., 2010
APIS-R	GGGGGGCGTTTAAAATGTGAAACAACTATG		
RpS5-F	AATTATTTGGTCGCTGGAATTG	115	Thompson et al., 2007
RpS5-R	TAACGTCCAGCAGAATGTGGTA		

3.3.6. Statistical Analysis and Visualization

Differences in pathogen levels between VR and VS colonies were assessed using Wilcoxon rank-sum test. Differences between proportions of disease-positive in VR and VS colonies were assessed using Fisher's exact test. Similarities between VR (VRL/VRH) and VS (VSL/VSH) colonies based on their pathogen loads, and correlations between pathogen loads, were examined using principal component analysis (PCA).

The effect of selective breeding for low *Varroa* mite levels on diseases, as well as relationships between diseases across all colonies and within VR and VS colonies were examined while accounting for variation in location/batches using generalized mixed linear models (GLMMs). For *Varroa* mites counts, a negative binomial GLMM was constructed with all other diseases/pathogens and group (VR/VS) as fixed predictors. For SB and AFB, binomial GLMMs were built with all other diseases/pathogens and group (VR/VS) as fixed predictors. For Nosema, zero-inflated Gamma GLMMs were constructed with all other diseases/pathogens and group (VR/VS) as fixed predictors. For viral infections, either negative binomial GLMMs or generalized linear models (GLMs)

were built with viral read counts as the response variable and and the other diseases/pathogens, group (VR/VS), and offset for total number of RNA-seq reads to account for different sequencing depths as fixed predictors. Variations associated with geographical locations and batches were accounted for by including them as random effects in each model, excepts on GLMs where they were included as fixed effects. GLMMs with and without interactions were tested for differences in residual deviance using a likelihood ratio test (LRT). Insignificant differences in residual deviance indicates that the two models fit the data similarly. Model residuals, over-/under-dispersion, outliers, and zero-inflation were checked using a simulation-based approach via the package DAHRMa v0.4.5 (Hartig, 2022). Predictors were standardized prior to model fitting. All GLMMs were fitted by maximum likelihood with Laplace approximation using the 'glmer' function for binomial GLMMs and 'glmer.nb' for negative binomial GLMMs of the R package lme4 v1-1.27.1 (Bates et al., 2015). Zero-inflated Gamma models using the function 'glmmTMB' of the R package glmmTMB v1.1.3 (Brooks et al., 2017). Negative binomial GLMs were fitted by maximum likelihood using the 'glm.nb' function of the R package MASS v7.3-54 (Venables & Ripley, 2002). Coefficient estimates, significance and 95% confidence intervals were retrieved from models using the 'get_model_data' function of the package sjPlot v2.8.9 (Lüdecke, 2021). All analyses and visualizations were performed in R 4.1.0 (R Core Team, 2021). Boxplots and PCA biplots were created using the packages ggpubr v0.4 (Kassambara, 2020), and ggplot2 v3.3.5 (Wickham, 2016) and ggrepel v0.9.1 (Slowikowski, 2021), respectively. Model coefficient estimate plots were created using the package ggplot2 v3.3.5 (Wickham, 2016). The Nature Publishing Group (NPG) color palette used in plots was retrieved from ggsci v2.9 (Xiao, 2018).

3.4. RNA Extraction, Quality Control (QC), and Sequencing

RNA extraction was performed using the EcoPURE Total RNA Kit (ECOTECH Biotechnology, Turkey) according to the manufacturer's protocol. The lysis buffer from the kit was mixed with β-mercaptoethanol in a volume ratio of 100:1. An on-column DNase I treatment step was applied during RNA extraction. Briefly, a 10 µL reaction containing 1 U of DNase I and 10X Reaction Buffer I (EURx, Poland) was added to each

column and were then incubated at 37 °C for 15 min for complete digestion of DNA. Extractions were carried out in batches of 12 samples. Preliminary assessment of total RNA yield and purity, and integrity was done using NanoDrop 1000c spectrophotometer (Thermo Fisher Scientific, US). RNA integrity was assessed using the 'bleach gel' method (Aranda et al., 2012). Briefly, RNA was separated by electrophoresis for 35 min on 1% agarose gels mixed with 0.5% household bleach (6% sodium hypochlorite) prior to melting. Representative samples with different RNA concentrations, levels of degradation and intensities of 28S and 18S rRNA bands were assessed further for integrity and amount using the Agilent 2100 Bioanalyzer with the RNA 6000 Nano Kit (Agilent Technologies, US), according to the manufacturer's protocol, without the heat denaturation step, and quantified using the QuantiFluor RNA System (Promega, US).. Heat denaturation, a standard step in integrity assessment of RNA to destroy secondary structures, results in the fragmentation of 28S rRNA into two similarly sized fragments that migrate closely with 18S rRNA in honey bees (Winnebeck et al., 2010). Arthropoda, along with most protostome animals, and other groups have a 'gap deletion' (also referred to as a 'hidden break') in their 28S rRNA sequences. 28S rRNA is excised at this site when it is subjected to heat denaturation (McCarthy et al., 2015). Therefore, this fragmentation does not translate to RNA degradation.

RNA concentrations ranged between 60-389 ng/µL based on results from the NanoDrop 1000c spectrophotometer. Concentrations of representative samples (n = 26) determined using the QuantiFluor RNA System were higher than the ones determined using the NanoDrop 1000c spectrophotometer for the same samples. RNA A260/280 and A260/230 values ranged between 2.0-2.2, and 2.0-2.3, respectively. Migration profiles of RNA samples in bleach gels showed a variety of degradation levels. Samples with highly degraded RNA were re-extracted. Pure RNA with low to medium degradation, and intact 28S and 18S rRNA bands (regardless of intensity) were sent to Novogene Corporation Inc. (Cambridge, UK) for sequencing. There, another round of QC was performed on the samples using NanoDrop and the Agilent 2100 Bioanalyzer. An unsmooth baseline was observed in the electropherograms, generated from the Bioanalyzer for 76 samples, indicating high risk of library construction and compromised sequencing quality. Nevertheless, all samples were subjected to first-strand library preparation, cluster generation, and sequencing. Briefly, Poly(A)+ selection to purify mRNA from total RNA was done using oligo(dT) beads. Fragmentation of the purified mRNA was done using a

fragmentation buffer containing divalent cations. First-strand cDNA synthesis was performed using reverse transcriptase in the presence of random hexamer primers and dNTPs. RNA removal was achieved via RNase H treatment. Strand-specificity was achieved by incorporating dUTP during the second-strand cDNA synthesis. End repair (i.e., generating 5' blunt ends) and A-tailing (i.e., addition of a single adenine nucleotide to 3'-ends of cDNA fragments) was performed to prevent fragments from ligating to each other and to provide an overhang for adapters (Table 5). Adapter ligation to the ends of double-stranded cDNA fragments was implemented to prepare them for hybridization. cDNA fragments ranging between 250 - 300 bp in length were selected using a double-sided bead clean-up. cDNA strands containing dUTPs were degraded using USER enzyme digestions, thus preserving information from the first strand only. Library enrichment was achieved using PCR, by generating cDNA libraries. Preliminary assessment of library concentration was assessed using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, US), assessment of insert length was made using the Agilent 2100 Bioanalyzer, and accurate library quantification was from qPCR assessment. Libraries that pass the quality control were sequenced on an Illumina NovaSeq 6000, generating ≥ 30 million paired-end 150 bp reads (PE150) for 113 samples because library construction failed for 2 samples.

Table 5. Adapter sequences ligated to ends of cDNA fragments. Index adapter sequences are eight bases as underlined.

End	Sequence (5'-3')
5'	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT
3'	GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGATGACTATCTCGTATGCCGTCTTCTGCTTG

3.5. Bioinformatic Analyses

3.5.1. Quality Control and Pre-processing of RNA-seq Reads

Quality Control (QC) of raw reads was completed using FastQC v0.11.7, and QC reports were summarized using MultiQC v1.12 (Ewels et al., 2016). Adapter and N base trimming of raw reads was accomplished using Cutadapt v2.5 (Martin, 2011) with the

following options: -a agatcggaagagcacacgtctgaactccagtca -A agatcggaagagcgtcgtgtagggaaagagtgt --trim-n --max-n 0.1 -m 25. The trimmed reads were aligned with the *Apis mellifera* genome assembly (Amel_HAv3.1) containing the genomes of 14 different RNA viruses that are known to infect honey bees. These viruses are Acute Bee Paralysis Virus (ABPV), NC_002548.1; Aphid Lethal Paralysis Virus (ALPV), NC_004365.1; Bee macula-like virus, NC_027631.1; Big Sioux River Virus (BSRV), NC_035184.1; Black Queen Cell Virus (BQCV), NC_003784.1; Chronic Bee Paralysis Virus (CBPV), NC_010711.1, NC_010712.1; Deformed Wing Virus (DWV), NC_004830.2; Israeli Acute Paralysis Virus (IAPV), NC_009025.1; Kakugo virus (KV), AB070959.1; Kashmir Bee Virus (KBV), NC_004807.1; Lake Sinai Virus (LSV), NC_032433.1; Sacbrood Virus (SBV), NC_002066.1; Slow Bee Paralysis Virus (SBPV), NC_014137.1; Varroa Destructor Virus-1 (VDV1), NC_006494.1. The alignment was done using HISAT2 v2.2.1 (Kim et al., 2019) using the following options: --rna-strandness RF.

The produced SAM files containing the alignments were sorted based on coordination, converted to BAM format, and indexed. BAM entries of reads aligned to viral genomes were extracted and reads mapped to each viral genome were counted. Sorting, conversion, indexing, extraction, and counting reads aligning only to viral genomes were done using Samtools v1.14 (Li et al., 2009). The ratio of number of reads aligned to the viral genome to the total number of reads for each virus was calculated as means to quantify viral loads in samples. BAM entries of reads aligned only to the *A. mellifera* genome were used for differential gene expression analysis. Reads aligned to exons annotated in the GFF annotation file of AmelHAv3.1 were counted using HTSeq v0.11.2 (Anders et al., 2015) in intersection-strict as the overlap resolution mode (htseq-count -f bam -s reverse -m intersection-strict --additional-attr=product). By default, htseq-count uses gene_id as the feature ID (GFF attribute). Thus, counts of reads that align to exons are summarized at the gene level.

3.5.2. Exploratory data and differential gene expression analyses

Exploratory analysis of count data and identification of DEGs was done in the R environment v4.1.0 (R Core Team, 2021), using the DESeq2 package v1.32.0 (Love et

al., 2014). Briefly, an n × m (12,319 × 113) count matrix K, where each entry K_{ij} is the number of reads aligned unambiguously to a gene i where $i \in \{1, \ldots, n\}$ in sample j where $j \in \{1, \ldots, n\}$, was generated by merging tab-delimited count files from htseq-counts that were imported into R. DESeq2 assumes K_{ij} follows a negative binomial (NB) distribution with a fitter mean μ_{ij} and a gene-specific dispersion α_i:

$$K_{ij} \sim NB(\mu_{ij}, \alpha_i)$$

A DESeqDataSet object was constructed with the count matrix, a data frame containing the variables group, subgroup, varroa mite counts, and disease screening results, and a design matrix x_{jr} (where r is the total number of coefficients), specifying the model to be fitted using the 'DESeqDataSetFromMatrix' function. Removal of low abundance genes was accomplished by retaining genes with 10 counts or higher in at least 7 samples. Prior to performing PCA to examine similarities between samples, transformation of count data is required to decrease the dependence of the variance on the mean as genes with low read counts tend to have high variance (Love et al., 2014). First, DESeq2 estimates size factors sample-specific s_j that accounts for sequencing depth or library size using the following formula:

$$s_j = median_i \frac{K_{ij}}{(\prod_{j=1}^{m} K_{ij})^{1/m}}$$

Then, it divides raw read counts for each gene in each sample by the sample-specific size factor. Afterwards, among two options that DESeq2 offers to flatten the experiment-wide trend of variance over mean, VST was selected. VST transforms normalized read counts using the dispersion-mean relationship $w\left(\frac{K_{ij}}{s_j}\right)$ fitted by DESeq2 as follows (Anders & Huber, 2010):

$$T(K) = \int^{K} \frac{d\frac{K_{ij}}{s_j}}{\sqrt{w\left(\frac{K_{ij}}{s_j}\right)}}$$

PCA was then performed on the transformed count data of top 500 genes with the highest variance, and the PC1 scores were plotted against PC2 scores for all samples using the package ggplot2 in R (Wickham, 2016).

DESeq2 models raw read counts K_{ij} using GLMs with a logarithmic link $log_2(q_{ij}) = \sum_r x_{jr}\beta_{ir}$, where q_{ij} is a quantity that is proportional to the expected true concentration of reads for sample j, with design matrix elements x_{jr} that was specified in the

DESeqDataSet object and coefficients β_{ir} for differential gene expression analysis. The design matrix is where comparisons are specified. For example, to identify differentially expressed genes (DEGs) between VR and VS samples, the simplest design matrix for this dataset is a 113×2 matrix, where the rows represent the samples and the first column represents the intercept, the second column represents whether a sample is from the VR or VS group. For each gene, a NB GLM is fitted, after estimating s_j and α_i, where group is a coefficient and its significance is assessed using a Wald test within the 'DESeq' function, that also adjusts for multiple testing using Benjamini–Hochberg (BH) method. The coefficient itself indicates the $\log_2$ fold change of the gene that was modeled between the groups VR and VS (Love et al., 2014).

The most complex design matrix for this dataset contains location, batch, group (or subgroup), varroa mite counts, diseases, pairwise disease interactions and interactions between varroa mite count and diseases. Results from models fitted with this design matrix are difficult to interpret as the matrix contains 31 coefficients. Therefore, a variety of design matrices ranging in complexity (number of coefficient) were used to fit models and the models were tested against each other for differences in residual deviance, using LRTs within the 'DESeq' function. Insignificant differences in residual deviance between two models, built with two different design matrices, fitting one gene, indicate that the two models fit the count data of that gene similarly. The total number of genes that fit better with the complex vs. simpler models were counted, and if the count is ≤ 100 genes, the simpler model, and hence the simpler design matrix, was chosen over the complex one, for ease of interpretability. The chosen design matrix with group as a main effect is: ~ location + SB + DWV + KV + group + SB:group + KV:group + DWV:group. The chosen design matrix with subgroup as a main effect is: ~ location + SB + DWV + KV + subgroup + SB:subgroup + KV:subgroup + DWV:subgroup.

Two DESeqDataSet objects using the 'DESeqDataSetFromMatrix' function with the chosen design matrix were constructed using the first design matrix and genes with 10 counts or higher in at least 7 samples were retained. In the first object, the base group and SB were specified as VS and negative, respectively, to identify DEGs between VS and VR with SB negative samples. In the second object, the SB base group was specified as positive to identify DEGs between VS and VR with SB positive samples. Another set of two Two DESeqDataSet objects were also created to identify DEGs between VRL and VSH with SB negative and positive samples. log2 fold change estimates (LFC) were

shrunk using apeglm, an adaptive Bayesian shrinkage estimator (Zhu et al., 2019). This method shrinks LFCs with high variance (arising from low or highly variable read counts) without overly shrinking large LFCs, thus reducing any variance that is associated with other diseases/infections. Genes were considered differentially expressed if their shrunken LFC > $|1|$ and adjusted p-value < 0.05. Common DEGs and their mean LFCs across all comparisons were retained for further analysis. Venn diagrams were drawn for common DEGs using ggvenn v0.1.9 (Yan, 2021). LFCs for common DEGs were calculated by taking the mean of LFC across all comparisons.

3.5.3. Network analysis and identification of candidate biomarkers

Protein-protein interaction (PPI) networks were constructed from common differentially expressed genes using the STRING plugin (stringApp) (Doncheva et al., 2019) in Cytoscape v3.8.2 (Shannon et al., 2003), with species set as *Apis mellifera* and a confidence (score) cutoff of 0.40. A PPI network constructed via STRING is a graph $G = (V, E)$ with a set of vertices or nodes V (referred to as proteins) and a set of undirected edges E (referred to as interactions) connecting the nodes together with weight corresponding to the confidence score of the interaction between the two proteins. STRING is a comprehensive database containing co-expression, text-mining, gene fusion, neighborhood, and experimental data for constructing PPI networks (Szklarczyk et al., 2021). Nodes that were not part of any network were discarded. Network- and node-level topological features were calculated using the built-in Cytoscape plugin NetworkAnalyzer. Nodes were then ranked based on their degree and betweenness centrality. Degree centrality C_D of a protein v is the number of interactions that protein is a part of, calculated using the following formula:

$$C_D(v) = \sum_{u \in K_u} w(u, v)$$

where K_u is the set of proteins u neighboring protein v, and $w(u, v) = 1$ if there is an interaction between protein u and protein v. The node degree distribution for each network was plotted to determine if the networks are scale-free. Scale-free networks are characterized by the abundance of small-degree nodes with a slow decrease in the frequency of high-degree nodes (Albert, 2005). This node degree distribution is said to

follow a power law. These high-degree nodes are called hubs and they represent proteins that are essential for the network as their loss causes its breakdown into isolated clusters (i.e., major loss of connectivity). Betweenness centrality C_B of a protein v is the frequency of participation that protein in acting as bridge along the shortest path between two other proteins, a and b, calculated using the following formula:

$$C_B(v) = \sum_{v \neq a \neq b} \frac{p(a, v, b)}{p(a, b)}$$

where $p(a, v, b)$ is the number of interactions between protein a and protein b, passing through protein v, and $p(a, b)$ is the total number of shortest paths from protein a to protein b. Betweenness centrality is an alternative measure for identifying hubs as it quantifies the amount of influence a protein has on the flow of information. Therefore, high-degree nodes with a high betweenness centrality in a scale-free network are considered candidate biomarkers for VR.

3.5.3. Functional enrichment analysis

The STRING Enrichment option provided by stringApp was used to map nodes in the networks constructed from the KEGG (Kyoto Encyclopedia of Genes and Genomes) PATHWAY database to identify pathways enriched in the networks. The KEGG PATHWAY database contains experimental data that is used to build networks for metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, and human diseases (Kanehisa & Goto, 2000). The stringApp was also used to retrieve enriched Gene Ontology (GO) terms that describe the molecular functions, biological processes in which those functions are necessary, and the cellular locations of the DEG products in the constructed networks (Ashburner et al., 2000; The Gene Ontology Consortium, 2021). The probabilities of GO terms and KEGG pathways enriched in the networks were calculated using an exact hypergeometric probability test automatically performed by stringApp on overlaps between nodes mapped to the list of genes participating in KEGG pathways or GO terms. P-values were corrected for multiple testing using the BH method. Furthermore, redundant pathways and terms were filtered out using Jaccard index which was set at 0.5. Pathways with adjusted p-value (FDR) < 0.05 were considered significantly enriched. Functional

annotations along with their negative log-transformed FDRs were plotted using ggplot2 (Wickham, 2016).

4. RESULTS

4.1. Prevalence of Diseases and Pathogens

The sampled *Varroa*-resistant (VR) honey bee colonies located on the Marmara Island, and *Varroa*-sensitive (VS) honey bee colonies from 5 locations within the Marmara region: Marmara Island, Mustafakemalpaşa, Çınarcık, Yalova, and Karacabey, were tested for *Varroa* mites using the sugar-roll method. At least one mite was present in 88.1% (37/42) of the VR colonies and 90.4% (66/73) of the VS colonies. Three mites or more were present in 44.2% (19/43) of the VR colonies and 68.5% (50/73) of the VS colonies. VR colonies had significantly less counts of *Varroa* mites than VS ones ($W = 852$, $p < 0.001$, Figure 2). Varroa mite counts were significantly different in VS colonies from different locations (Kruskal-Wallis test: $H=14.33$, $df = 4$, $p < 0.01$). VS colonies sampled from Yalova have significantly higher *Varroa* mite counts than VS colonies sampled from Çınarcık (Dunn's test: $p<0.05$), Karacabey (Dunn's test: $p < 0.01$), and Mustafakemalpaşa (Dunn's test: $p<0.05$).

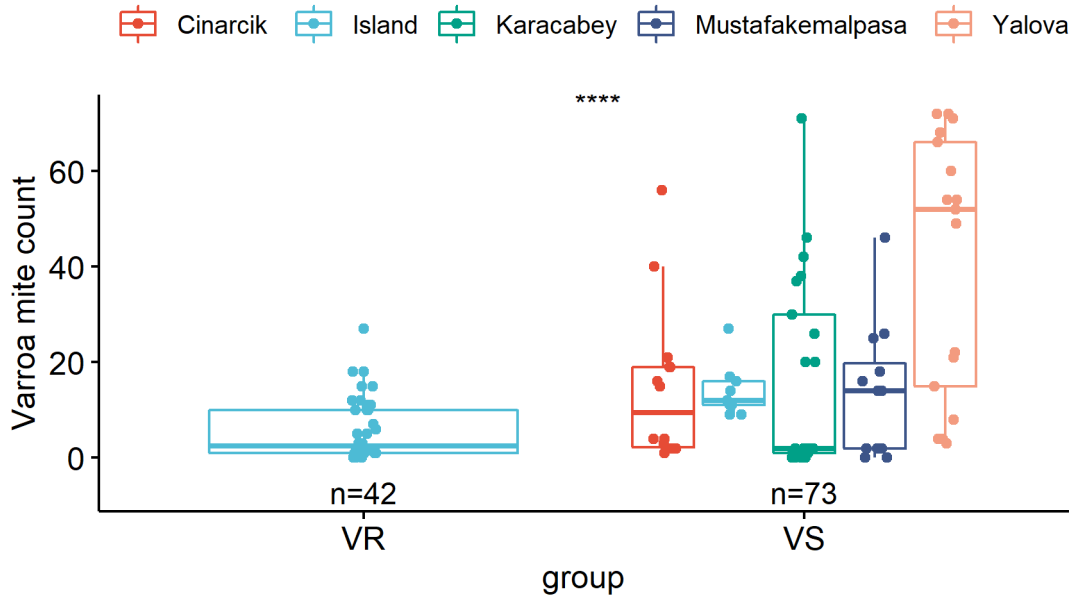

Figure 2. *Varroa* mite counts in VR and VS colonies across different locations. Box plots represent median, interquartile range (IQR), and 1.5 × IQR. Statistical significance was determined by Wilcoxon rank-sum test. ****p ≤ 0.0001, ***p ≤ 0.001, **p ≤ 0.01, *p ≤ 0.05, not significant (ns) p > 0.05.

Two fungal diseases, stonebrood (SB)—caused by a variety of *Aspergillus* species —and chalkbrood (CB)—caused by *Ascosphaera apis*—were screened for using conventional PCR assays. SB was detected in 52.4% (22/42) of the VR colonies and 72.6% (53/73) of the VS colonies (Table 6). The difference in SB prevalence between VR and VS colonies is significant (Fisher's exact test, $p < 0.05$). Differences in SB prevalence in VS colonies from different locations were also significant (Fisher's exact test, $p < 0.01$). SB was detected in 92.9% (13/14), 33.3% (3/9), 61.9% (13/21), 91.7% (11/12), and 82.4% (14/17) of VS colonies located in Çınarcık, Marmara Island, Karacabey, Mustafakemalpaşa, and Yalova regions, respectively (Table 6). CB was not detected in any of the sampled colonies.

Another widespread fungal disease, Nosema—caused by *N. ceranae* and *N. apis*, were also screened for using a triplex PCR assay. *N. ceranae* was detected in 58.1% (25/43) of VR colonies and 67.1% (49/73) of the VS colonies (Table 6). The *N. ceranae* load was significantly less in VR compared to VS colonies ($W = 1085$, $p < 0.01$, Figure 3). *N. ceranae* loads were also significantly different in VS colonies from different locations ($H=58.11$, $df = 4$, $p < 0.001$, Figure 3). VS colonies sampled from Çınarcık have significantly higher *N. ceranae* loads than VS colonies sampled from Marmara Island (Dunn's test: $p < 0.001$), Karacabey (Dunn's test: $p < 0.001$), and Mustafakemalpaşa

(Dunn's test: $p < 0.001$). VS colonies sampled from Karacabey have significantly higher *N. ceranae* loads than VS colonies sampled from Mustafakemalpaşa (Dunn's test: $p < 0.05$). VS colonies sampled from Yalova have significantly higher *N. ceranae* loads than VS colonies sampled from Marmara Island (Dunn's test: $p < 0.01$), Karacabey (Dunn's test: $p < 0.001$), and Mustafakemalpaşa (Dunn's test: $p < 0.01$). *N. apis* was not detected in any of the sampled colonies.

Two bacterial diseases, American foulbrood (AFB)—caused by *Paenibacillus larvae*—and European foulbrood (EFB)—caused by *Melissococcus plutonius*—were screened for using two different qPCR assays. AFB was detected in 19% (8/42) of the VR colonies and 38.4% (28/73) of the VS colonies (Table 6). The difference in AFB prevalence between VR and VS colonies is significant (Fisher's exact test, $p < 0.05$). Differences in AFB prevalence in VS colonies from different locations were also significant (Fisher's exact test, $p < 0.01$). AFB was detected in 64.3% (9/14), 22.2% (2/9), 33.3% (7/21), 58.3% (7/12), and 17.6% (3/17) of VS colonies located in Çınarcık, Marmara Island, Karacabey, Mustafakemalpaşa, and Yalova regions, respectively (Table 6). EFB was detected in 4.8% (2/42) of the VR colonies and 6.8% (5/73) of the VS colonies (Table 6). Due to its low prevalence, EFB was left out of further analyses.

Table 6. Prevalence of diseases screened for in VR colonies sampled from the Marmara Island (Island), and VS colonies sampled form Çınarcık, Island, Karacabey, Mustafakemalpaşa, and Yalova regions.

	VR	VS				
	Island	Çınarcık	Island	Karacabey	MKP	Yalova	Total
SB	52.4% (22/42)	13/14	3/9	13/21	11/12	14/17	72.6% (53/73)
Nosema	58.1% (25/43)	13/14	7/9	2/21	9/12	17/17	67.1% (49/73)
AFB	19% (8/42)	9/14	2/9	7/21	7/12	3/17	38.4% (28/73)
EFB	4.8% (2/42)	2/14	0/9	2/21	1/12	0/17	6.8% (5/73)

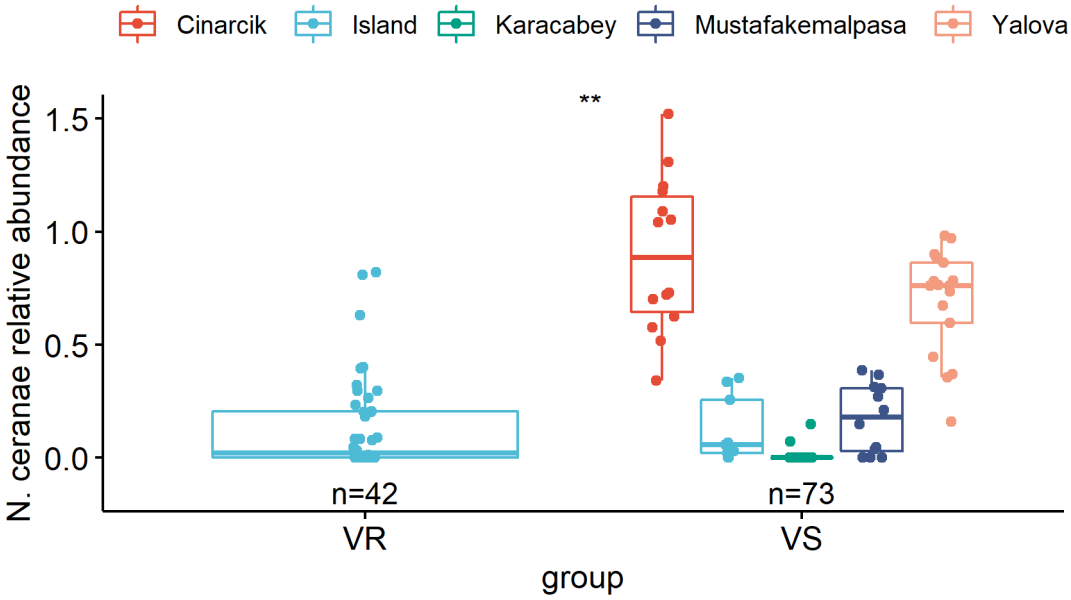

Figure 3. *N. ceranae* relative abundance in VR and VS colonies across different locations. Box plots represent median, interquartile range (IQR), and 1.5 × IQR. Statistical significance was determined by Wilcoxon rank-sum test. ****p ≤ 0.0001, ***p ≤ 0.001, **p ≤ 0.01, *p ≤ 0.05, not significant (ns) p > 0.05.

Fourteen different honey bee-infecting viruses screened for using RNA sequencing (RNA-seq). Chronic bee paralysis virus (CBPV), Israeli acute paralysis virus (IAPV), and slow bee paralysis virus (SBPV) were not detected in any of the sampled colonies. At least 10 reads aligning to the genomes of acute bee paralysis virus (ABPV), aphid lethal paralysis virus (ALPV), bee macula-like virus (BeeMLV), black queen cell virus (BQCV), Big Sioux River virus (BSRV), deformed wing virus (DWV), Kashmir bee virus (KBV), Kakugo virus (KV), Lake Sinai virus (LSV), sacbrood virus (SBV), and Varroa destructor virus-1 (VDV1) were detected in 78.6% (33/42), 23.8% (10/42), 26.2% (11/42), 100% (42/42), 0% (0/42), 100% (42/42), 2.4% (1/42), 97.6% (41/42), 50% (21/42), 2.4% (1/42), and 100% (42/42) of the VR colonies, respectively; and 72.6% (53/73), 15.1% (11/73), 27.4% (20/73), 100% (73/73), 1.4% (1/73), 100% (73/73), 0% (0/73), 100% (42/42), 60.3% (44/73), 9.6% (7/73), and 100% (73/73) of the VS colonies, respectively (Table 7). ALPV, BeeMLV, BSRV, KBV, LSV, and SBV were left out of further analyses due to their low prevalence and/or reads.

Table 7. Prevalence of viruses screened for using RNA-seq in VR colonies sampled from the Marmara Island (Island), and VS colonies sampled form Çınarcık, Island, Karacabey, Mustafakemalpaşa, and Yalova regions.

	VR	VS				
	Island	Çınarcık	Island	Karacabey	MKP	Yalova	Total

ABPV	78.6% (33/42)	12/14	4/9	21/21	7/12	9/17	72.6% (53/73)
ALPV	23.8% (10/42)	1/14	0/9	1/21	3/12	6/17	15.1% (11/73)
BeeMLV	26.2% (11/42)	2/14	1/9	7/21	4/12	10/17	27.4% (20/73)
BQCV	100% (42/42)	14/14	9/9	21/21	12/12	17/17	100% (73/73)
BSRV	0% (0/42)	0/14	0/9	0/21	0/12	1/17	1.4% (1/73)
DWV	100% (42/42)	14/14	9/9	21/21	12/12	17/17	100% (73/73)
KBV	2.4% (1/42)	0/14	0/9	0/21	0/12	0/17	0% (0/73)
KV	97.6% (41/42)	14/14	9/9	21/21	12/12	17/17	100% (42/42)
LSV	50% (21/42)	9/14	1/9	12/21	12/12	11/17	60.3% (44/73)
SB	2.4% (1/42)	0/14	0/9	2/21	1/12	4/17	9.6% (7/73)
VDV1	100% (42/42)	14/14	9/9	21/21	12/12	17/17	100% (73/73)

Reads aligning to the genomes of ABPV, BQCV, DWV, KV, and VDV1 were divided by the total number of reads to generate viral level estimates. ABPV levels were significantly higher in VR colonies compared to VS colonies ($W = 2157.5$, $p < 0.001$, Figure 4). ABPV levels were also significantly different in VS colonies from different locations ($H=33.62$, $df = 4$, $p < 0.001$, Figure 4). VS colonies sampled from Karacabey have significantly higher ABPV levels than VS colonies sampled from Çınarcık (Dunn's test: $p < 0.01$), Marmara Island (Dunn's test: $p < 0.001$), Mustafakemalpaşa (Dunn's test: $p < 0.001$), and Yalova (Dunn's test: $p < 0.001$). BQCV levels were not significantly different between VR and VS colonies ($W = 1517$, $p = 0.93$, Figure 4), nor were they significantly different in VS colonies from different locations ($H = 5.90$, $df = 4$, $p = 0.21$, Figure 4). DWV, KV, and VDV1 levels were significantly higher in VS colonies compared to VR colonies (DWV: $W = 658$, $p < 0.001$; KV: $W = 630$, $p < 0.001$; VDV1: $W = 1041$, $p < 0.01$; Figure 4). DWV and KV levels were not significantly different in VS colonies from different locations (DWV: $H = 8.8$, $df = 4$, $p = 0.07$, KV: $H = 6.62$, $df = 4$, $p = 0.16$, Figure 4). VDV1 levels were significantly different in VS colonies from different locations ($H= 14.5$, $df = 4$, $p < 0.01$, Figure 4). VS colonies sampled from Çınarcık have significantly lower VDV1 levels than VS colonies sampled from Marmara Island (Dunn's test: $p < 0.05$), and Karacabey (Dunn's test: $p < 0.05$). In addition, VS

colonies sampled from Marmara have significantly lower VDV1 levels than VS colonies sampled from Mustafakemalpaşa (Dunn's test: $p < 0.05$).

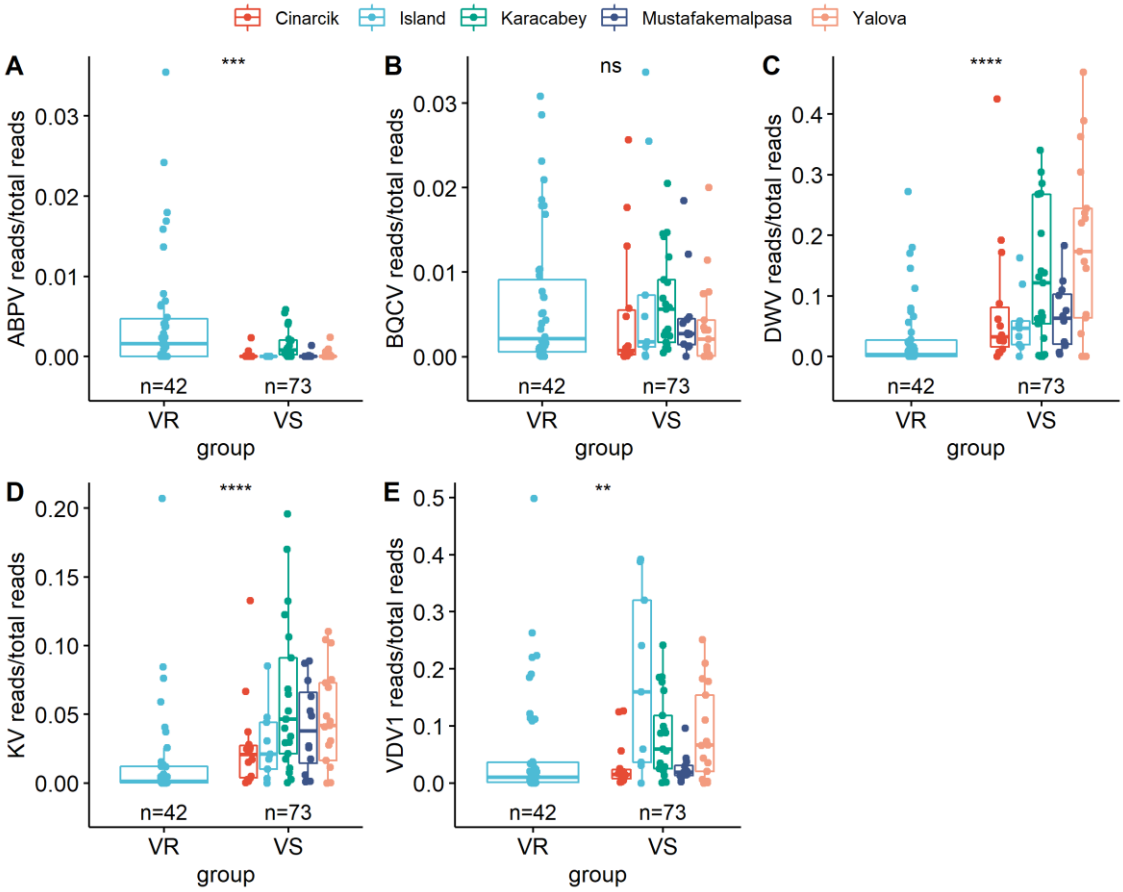

Figure 4. Levels of ABPV (A), BQCV (B), DWV (C), KV (D), and VDV1 (E) in VR and VS colonies across different locations. Box plots represent median, interquartile range (IQR), and 1.5 × IQR. Statistical significance between groups, not locations, was determined by Wilcoxon rank-sum test. ****$p \leq 0.0001$, ***$p \leq 0.001$, **$p \leq 0.01$, *$p \leq 0.05$, not significant (ns) $p > 0.05$.

Principal component analysis (PCA) of ABPV, BQCV, DWV, KV, VDV1, varroa mite, and *N. ceranae* levels was performed and principal component (PC) scores for the first and second principal components (PC1 and PC2) of the sampled colonies were plotted against each other. PC1 and PC2 described 38.5% and 17% of the total variance, respectively. Based on the loadings, PC1 was comprised mainly of KV, DWV, VDV1 and varroa mite counts, suggesting possible interaction effects between levels of those viruses. PC2 was largely described by ABPV, BQCV, and *N. ceranae*, again suggesting possible pairwise interaction effects (Figure 5). In this PC space, VR and VS colonies and their respective subgroups do not form entirely separate clusters (Figure 5). However, VRL and VSH colonies tend to cluster in the upper left quarter and right half of the PC space, respectively.

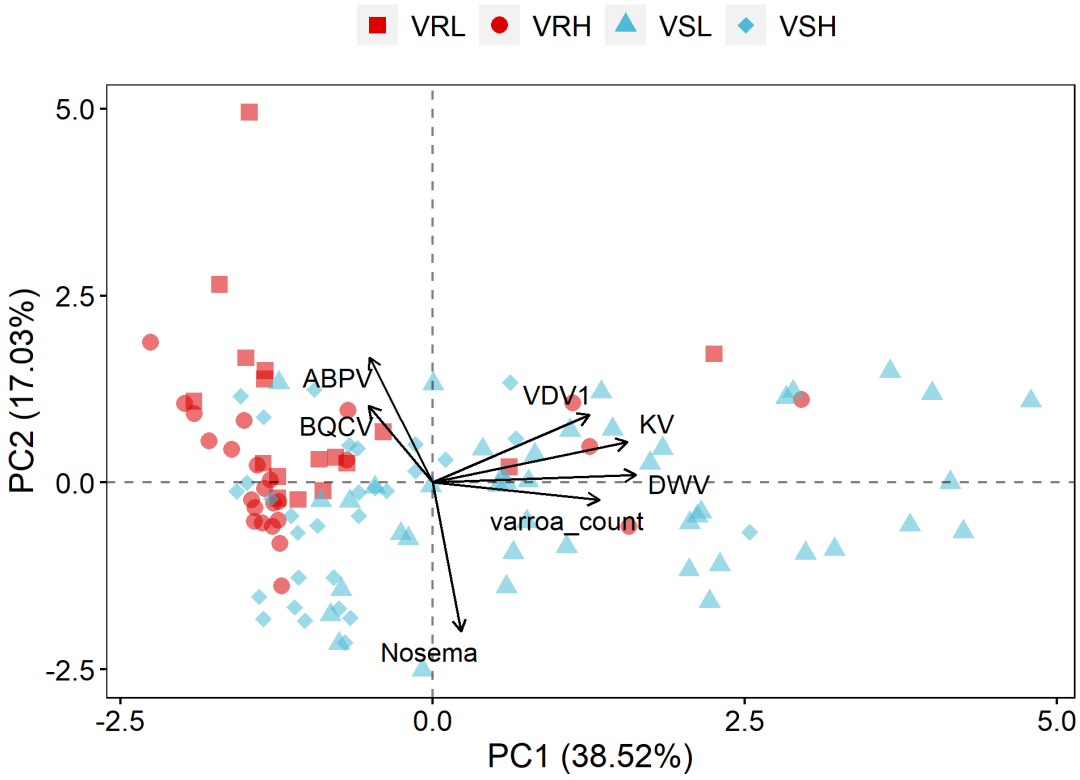

Figure 5. PCA results of viral loads, varroa mite counts, and N. ceranae levels quantified in sampled honey bee colonies for the first two PCs. Each dot represents a colony that is VRL (red square), VRH (red circle), VSL (cyan triangle), or VSH (cyan diamond). Arrows represent the directions of diseases/pathogen loadings according to first two PCs. Percentage of variance explained by each PC is depicted on its respective axis.

4.2. Relationships Between Diseases and Pathogens

4.2.1. Varroa Mites

Negative binomial (NB) generalized mixed-effect linear models (GLMMs) were built to examine relationships between *Varroa* mite counts and other diseases or pathogens. First, two models with *Varroa* mite counts as a response variable were built, one with and another without DWV-KV-VDV1 and ABPV-BQCV interactions as fixed variables, and subsequently compared for their goodness of fit. Difference in residual deviance between models with and without interactions across all colonies was not significant ($\chi^2 = 6.34$, df = 4, $p = 0.18$). Therefore, the simpler model (without interactions) was examined further. *Varroa* mite counts were associated positively with sensitivity to *Varroa* (NB Varroa

GLMM: groupVS estimate = 1.22, $p < 0.001$; Figure 6), SB prevalence (NB Varroa GLMM: SB estimate = 0.88, $p < 0.001$; Figure 6), and ABPV levels (NB Varroa GLMM: ABPV estimate = 0.27, $p < 0.01$; Figure 6). On the other hand, *Varroa* mite counts were associated negatively with *N. ceranae* levels (NB Varroa GLMM: Nosema estimate = -0.39, $p < 0.05$; Figure 6)

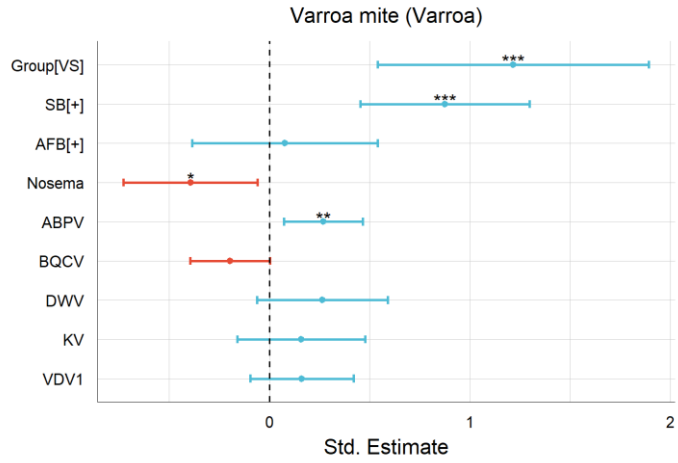

Figure 6. Size and direction of the relationships (standardized coefficient estimates) between *Varroa* mites counts, *Varroa* resistance, and other diseases. Positive and negative estimates and 95% CI whiskers are colored in cyan and red, respectively. *p < 0.05, **p < 0.01 and ***p < 0.001.

4.2.2. Fungal diseases

4.2.2.1. Stonebrood

Multivariate mixed-effect binomial models (binomial GLMM) were built to examine relationships between SB and other diseases or pathogens. First, two models with SB as a response variable were built, one with and another without DWV-KV-VDV1 and ABPV-BQCV interactions as fixed variables, and subsequently compared for their goodness of fit. Difference in residual deviance between models with and without interactions across all colonies was not significant ($\chi^2 = 4.98$, df = 4, $p = 0.29$). Therefore, the simpler model (without interactions) was examined further. SB was associated positively with *Varroa* mite counts (Binomial SB GLMM: varroa_count estimate = 1.5, $p < 0.001$; Figure 7A), but not with resistance to *Varroa* (Binomial SB GLMM: groupVS estimate = 0.02, $p = 0.97$; Figure 7A).

4.2.2.1. Nosema

Multivariate mixed-effect zero-inflated Gamma models were built to examine relationships between *N. ceranae* (herein referred to as Nosema) levels and other diseases or pathogens. First, two models with *Nosema* as a response variable were built, one with and another without DWV-KV-VDV1 and ABPV-BQCV interactions as fixed variables, and subsequently compared for their goodness of fit. Difference in residual deviance between models with and without interactions across all colonies was not significant (χ^2 = 2.44, df = 4, p = 0.66). Therefore, the simpler model was examined further. Across all colonies, Nosema was not associated with any other disease or pathogen (Figure 7B).

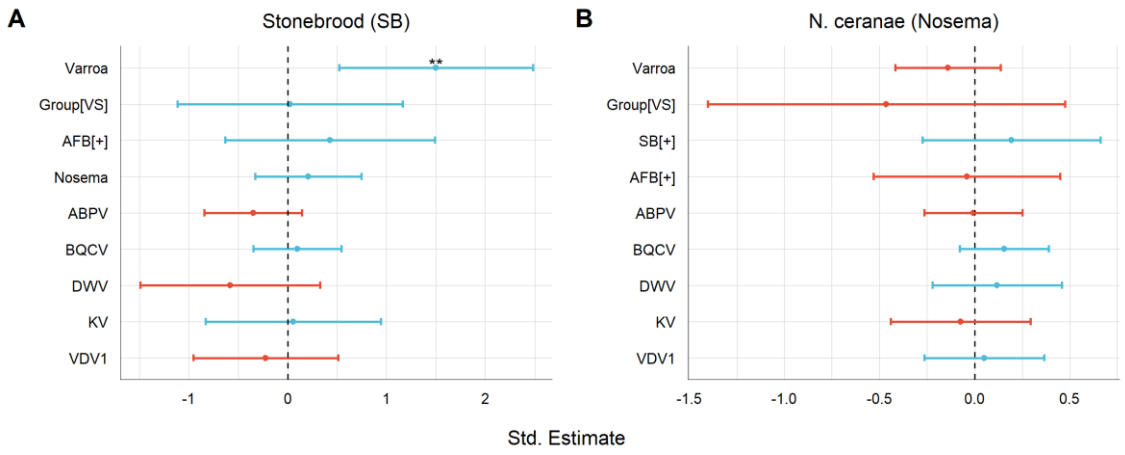

Figure 7. Size and direction of the relationships (standardized coefficient estimates) between SB (A) *Varroa* resistance, and other diseases, and Nosema (B) *Varroa* resistance, and other diseases. Positive and negative estimates and 95% CI whiskers are colored in cyan and red, respectively. *p < 0.05, **p < 0.01 and ***p < 0.001.

4.2.3. American Foulbrood

Another set of multivariate mixed-effect binomial models were built to examine relationships between AFB and other diseases or pathogen. First, two models with AFB as a response variable were built, one with and another without DWV-KV-VDV1 and ABPV-BQCV interactions as fixed variables, and subsequently compared for their goodness of fit. Difference in residual deviance between models with and without

interactions across all colonies was not significant (χ^2 = 2.97, df = 4, p = 0.57). Therefore, the simpler model was examined further. There was no significant difference between the two models across all colonies. Therefore, the simpler models were examined for disease associations. Across all colonies, AFB was not associated with any other disease or pathogen (Figure 8).

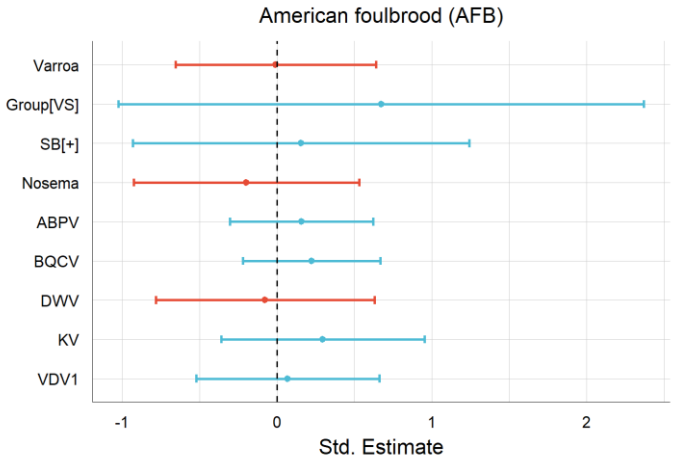

Figure 8. Size and direction of the relationships (standardized coefficient estimates) between AFB, *Varroa* resistance, and other diseases. Positive and negative estimates and 95% CI whiskers are colored in cyan and red, respectively. *p < 0.05, **p < 0.01 and ***p < 0.001.

4.2.5. Viral diseases

Another set of multivariate mixed-effect negative binomial models (NB GLMMs) were built to examine relationships between BQCV, DWV, or KV levels and other diseases or pathogens. As with other models, two models for each virus variable were built, one with and another without interactions between viral levels as fixed variables, and subsequently compared for their goodness of fit. Difference in residual deviance between models with and without interactions across all colonies was not significant for BQCV (χ2 = 0.34, df = 4, p = 0.99), but were significant for DWV (χ2 = 16.86, df = 2, p < 0.001) and (χ2 = 11.78, df = 3, p= p < 0.01). Therefore, the simpler model was further examined for BQCV, and the complex models were further examined for DWV and KV. Across all colonies, BQCV was not associated with any disease or pathogen (Figure 9A). DWV levels were positively associated with sensitivity to *Varroa* (NB DWV GLMM: groupVS estimate =

0.71, p < 0.001; Figure 9B), KV levels (NB DWV GLMM: KV estimate = 1.45, p < 0.001; Figure 9B), and VDV1 levels (NB DWV GLMM: VDV1 estimate = 0.5, p < 0.001; Figure 9B). DWV was negatively associated with the interaction between KV and VDV1 levels (NB DWV GLMM: KV × VDV1 estimate = -0.93, p < 0.001; Figure 9B). KV levels were positively associated with sensitivity to *Varroa* (NB KV GLMM: groupVS estimate = 1.32, p < 0.001; Figure 9C), VDV1 levels (NB KV GLMM: VDV1 estimate = 0.84, p < 0.001; Figure 9C), and DWV levels (NB KV GLMM: DWV estimate = 1.17, p < 0.001; Figure 9C). KV levels were negatively associated with the interaction between DWV and VDV1 levels (NB KV GLMM: DWV × VDV1 estimate = -0.78, p < 0.001; Figure 9C).

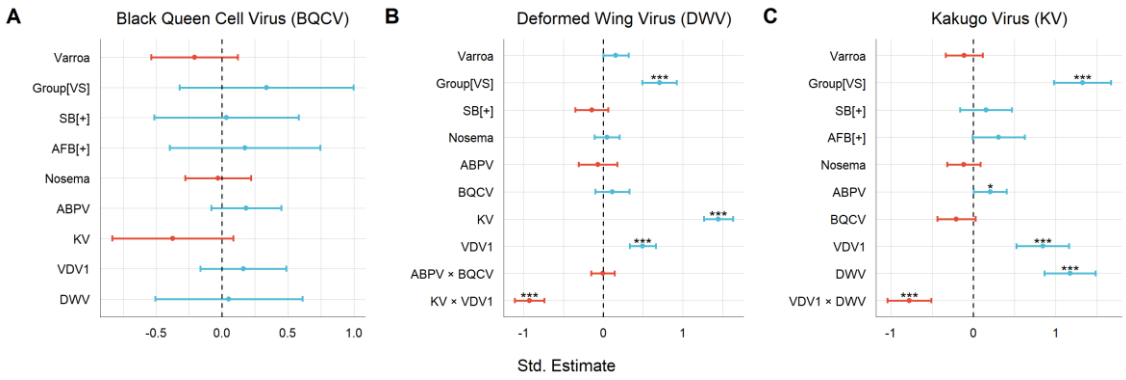

Figure 9. Size and direction of the relationships (standardized coefficient estimates) between BQCV (A) *Varroa* resistance, and other diseases; DWV (B) *Varroa* resistance, and other diseases; and KV (C) *Varroa* resistance, and other diseases. Positive and negative estimates and 95% CI whiskers are colored in cyan and red, respectively. *p < 0.05, **p < 0.01 and ***p < 0.001.

Multivariate generalized negative binomial models (NB GLMs) were built to examine relationships between ABPV or VDV1 levels and other diseases or pathogens. In contrast to other models, ABPV and VDV1 models included DWV-KV-VDV1 and KV-DWV interactions, respectively, and were not compared with simpler models. They also included location and batch as fixed rather random effects. ABPV was associated positively with *Varroa* mite counts (NB ABPV GLM: groupVS estimate = -1.44, p < 0.001; Figure 10A) and negatively with sensitivity to *Varroa* (NB ABPV GLM: groupVS estimate = -1.09, p < 0.01; Figure 10A). As for viruses, ABPV was negatively associated with DWV levels (NB ABPV GLM: DWV estimate = -1.23, p < 0.01; Figure 10A) and the interaction between KV and VDV1 levels (NB ABPV GLM: KV × VDV1 estimate = -1.97, p < 0.001; Figure 10A), and positively associated with the interaction between

VDV1 and DWV (NB ABPV GLM: VDV1 × DWV estimate = 1.24, $p < 0.01$; Figure 10A). VDV1 levels were positively associated with *Varroa* mite count (NB VDV1 GLM: varroa_count estimate = 0.42, $p < 0.05$; Figure 10B), KV levels (NB VDV1 GLM: KV estimate = 0.78, $p < 0.001$; Figure 10B), and DWV levels (NB VDV1 GLM: DWV estimate = 0.87, $p < 0.001$; Figure 10B). VDV1 levels were negatively associated KV and DWV levels (NB VDV1 GLM: KV × DWV estimate = -0.62, $p < 0.001$; Figure 10B).

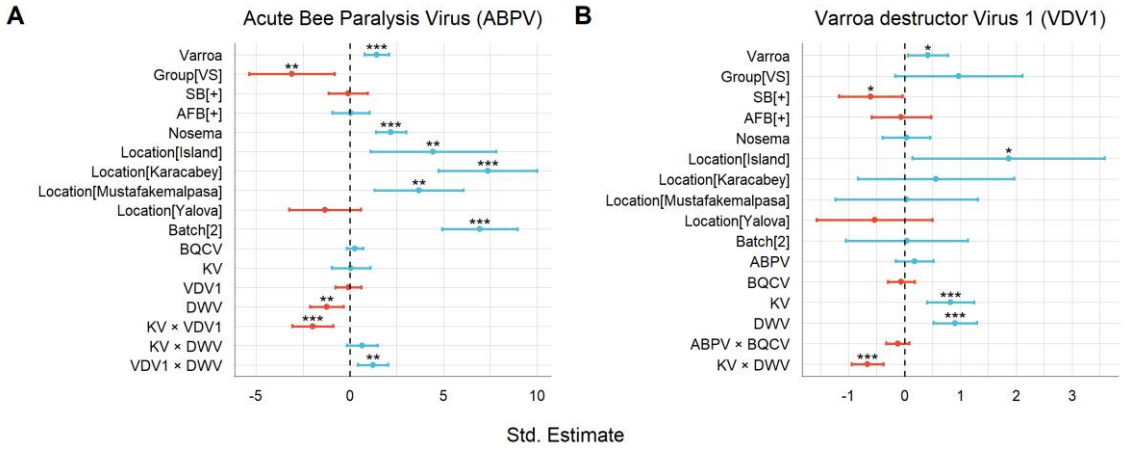

Figure 10. Size and direction of the relationships (standardized coefficient estimates) between ABPV (A) *Varroa* resistance, and other diseases, and VDV1 (B) *Varroa* resistance, and other diseases. Positive and negative estimates and 95% CI whiskers are colored in cyan and red, respectively. *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$.

4.6. Sequencing and Alignment Statistics

At least 30 million 150 bp paired-end reads were generated per sample, except for two due to failure in library construction (Appendix A, Table S1). Quality control (QC) analyses using FastQC/MultiQC of FASTQ files revealed characteristics typical of RNA-seq reads. All files containing the reads pass the Per Base Sequence Quality (Appendix B, Figure S1), Per Sequence Quality Scores (Appendix B, Figure S2), Per Base N Content (Appendix B, Figure S3), and Sequence Length Distribution modules. On the other hand, they all failed in passing the Per Base Sequence Content module. This is expected for RNA-seq reads as nearly all RNA-seq libraries are produced by priming using random hexamers that carry an intrinsic bias in the positions at the beginning of RNA-seq reads (Hansen et al., 2010). This true technical bias cannot be corrected. However, there are no adverse effects on downstream analyses for differential gene expression analysis (Hansen

et al., 2010). Furthermore, most files failed in passing the Per Sequence GC Content module, a feature of RNA-seq libraries as the distribution of mean GC content of reads depends on the transcripts. All files fail to pass the Duplicate Sequences modules (Appendix B, Figure S4), a result that is expected for RNA-seq libraries as they can contain truly overrepresented sequences such as very abundant transcripts. In some cases, some transcripts are so abundant in samples that they register as overrepresented sequences. Therefore, it is also normal for the Overrepresented Sequences module to issue a warning or failure for RNA-seq libraries. Finally, all files passed the Adapter Content module (Appendix B, Figure S5). However, it was observed that up to 0.5% of sequences contain partial adapter sequences at their 3'-ends (Appendix B, Figure S5). Absence of adapter contamination was confirmed after trimming with Cutadapt using FastQC/MultiQC.

The average alignment rates to the concatenation of the *Apis mellifera* and 14 viral genomes were 76.7% for the VS group and 85.4% for the VR group (Appendix A, Table S3). The overall average alignment rate was 80%, with 2 samples reaching alignment rates < 50%, 4 samples reaching alignment rates ranging between 50-60%, 8 samples reaching alignment rates ranging between 60-70%, 25 samples reaching alignment rates ranging between 70-80%, and 55 samples reaching alignment rates ranging between 80-90% using HISAT2 (Appendix A, Table S3). rRNA content reported by Novogene varied between 1-23% (Appendix A, Table S4).

4.7. Differentially Expressed Genes in VR and VRL Honey Bees

PCA on counts of the top 500 genes with high variance was performed and principal PC scores for PC1 and PC2 of the sampled colonies were plotted against each other. PC1 and PC2 described 40% and 18% of the total variance, respectively. In this PC space, VR and VS colonies and their respective subgroups do not form entirely separate clusters as there are VS colonies that have expression profile of the top 500 genes similar to that of some VR colonies (Figure 10).

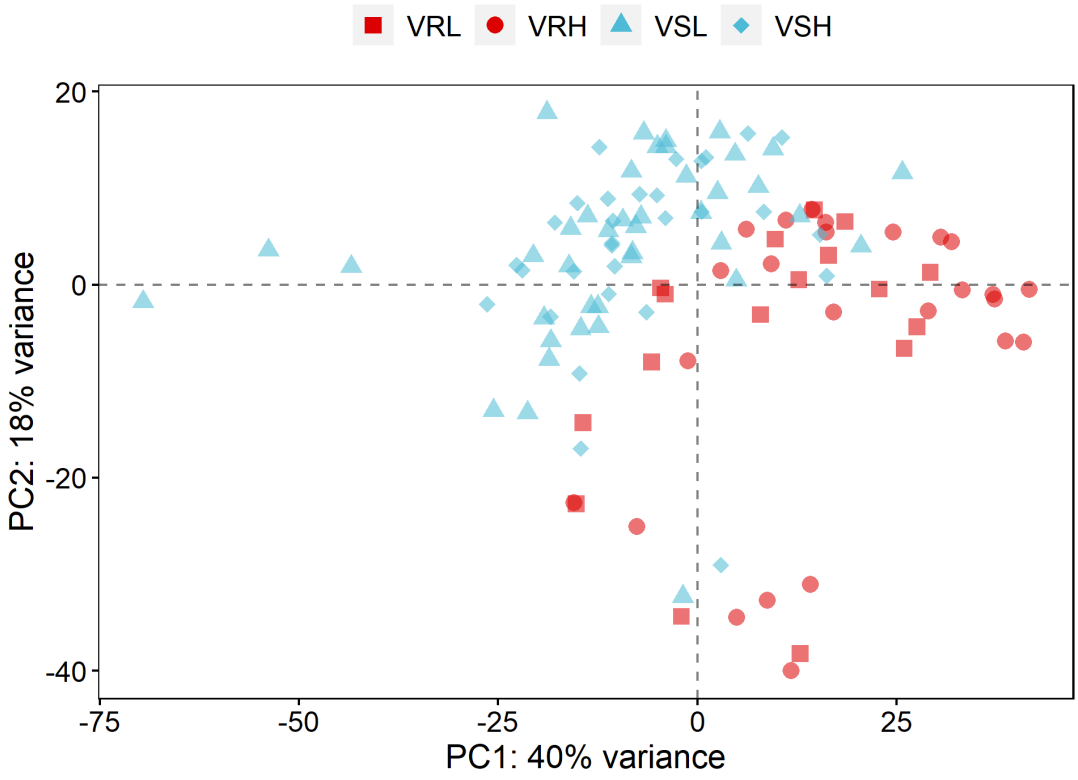

Figure 11. PCA results of top 500 genes with the highest variance across all 113 samples for the first two PCs. Each dot represents a colony that is VRL (red square), VRH (red circle), VSL (cyan triangle), or VSH (cyan diamond). Percentage of variance explained by each PC is depicted on its respective axis.

Differential expression analysis revealed 587 significantly upregulated and 44 significantly downregulated genes in VR compared to VS colonies that are SB-negative, and 329 significantly upregulated and 54 significantly downregulated genes in VR compared to VS colonies that are SB-positive while accounting for differences in location and other diseases through the design matrix and log2 fold change (LFC) shrinkage. In addition, differential expression analysis revealed 537 significantly upregulated and 147 significantly downregulated genes in VRL compared to VSH colonies that are SB-negative, and 361 significantly upregulated and 127 significantly downregulated genes in VRL compared to VSH colonies that are SB-positive while accounting for differences in location and other diseases through the design matrix and LFC shrinkage. Across all comparisons, 278 significantly upregulated and 23 significantly downregulated genes were common (Figure 12, Table S2, Table S3). The top significantly upregulated protein-coding gene in VR, CCAAT/enhancer-binding protein gamma (LOC726550), is considered a candidate biomarker.

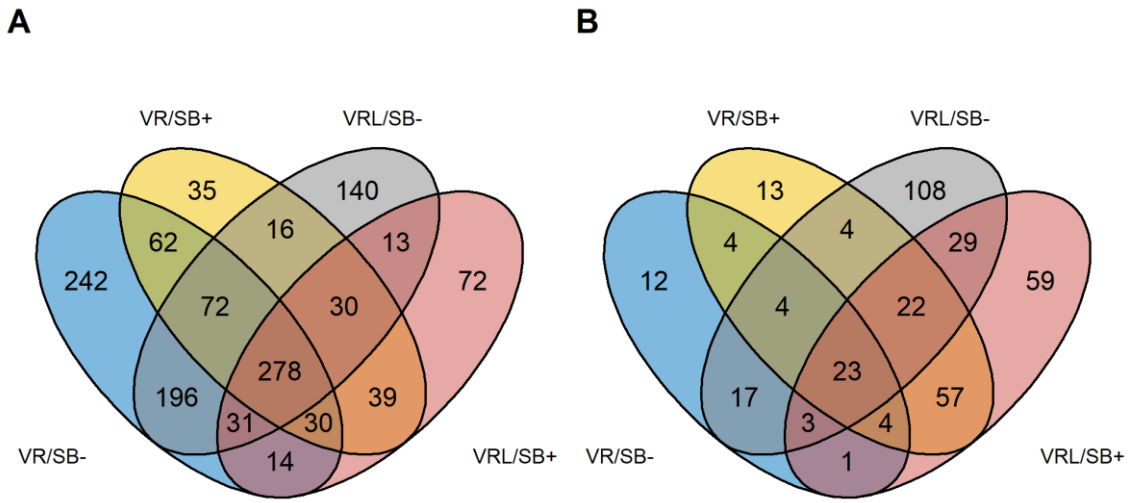

Figure 12. Common upregulated genes in VR colonies that are SB-positive and SB-negative (VR/SB+ and VR/SB-), and VRL colonies that SB-positive and SB-negative (VRL/SB+ and VRL/SB-) (A); and common downregulated genes in VR colonies that are SB-positive and SB-negative (VR/SB+ and VR/SB-), and VRL colonies that SB-positive and SB-negative (VRL/SB+ and VRL/SB-).

4.8. Protein-protein Interaction Networks and Enriched Functions in VR Colonies

Common significantly upregulated and downregulated genes were used to build two separate protein-protein interaction (PPI) networks by querying the STRING database for *Apis mellifera* with interaction confidence score cutoff of 0.4 in Cytoscape. Proteins mapped to common downregulated genes in VR colonies did not form a network and were thus not considered further. Of the 278 common upregulated genes in VR, 94 genes were mapped to proteins that were part of a large network, while 3 genes were mapped to proteins that formed a network independent of the large one (Figure 13).

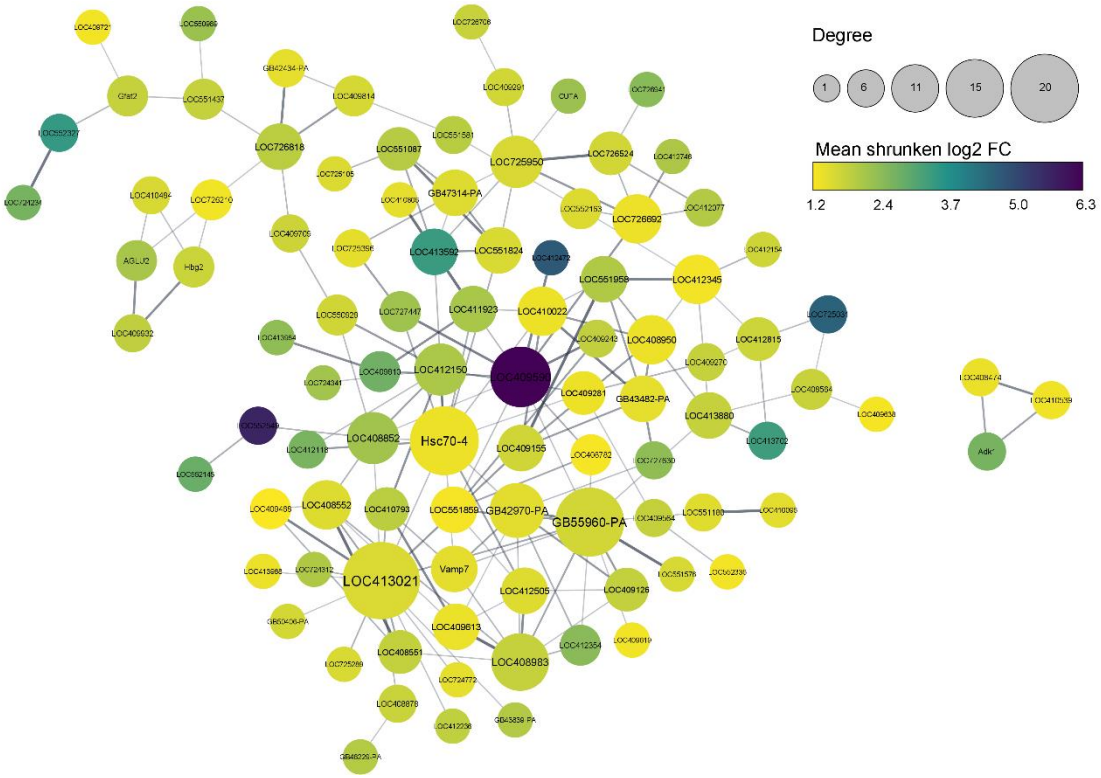

Figure 13. PPI network built from common upregulated genes in VR colonies. The node size and color vary according to node degree and mean shrunken log2 fold change, respectively. The edge thickness varies according to the confidence score of interactions.

Biological significance of the upregulated network in VR colonies was determined by identifying enriched KEGG pathways and GO terms. Enriched KEGG pathways were associated with metabolic pathways in general and more specialized metabolic pathways such as amino acid and nucleotide sugar metabolism (Figure 14). The network was also enriched in nodes participating in protein processing in the endoplasmic reticulum, lysosomal pathways, and fatty acid elongation (Figure 14). Enriched GO terms are mostly associated with metabolism of a variety of organic molecules (Figure 14). The network was also enriched in GO terms associated with regulation of protein localization, proteolysis, protein secretion extracellular matrix organization, and positive regulation of developmental pigmentation (Figure 14).

Figure 14. GO cellular components, GO molecular functions, GO biological processes, and KEGG pathways enriched in the upregulated Varroa resistance network.

4.9. Candidate Biomarkers for *Varroa* Resistance

Topological features of the largest network were calculated, including degree and betweenness centrality. The node degree distribution follows a power law (Figure 15), indicating its scale-free property.

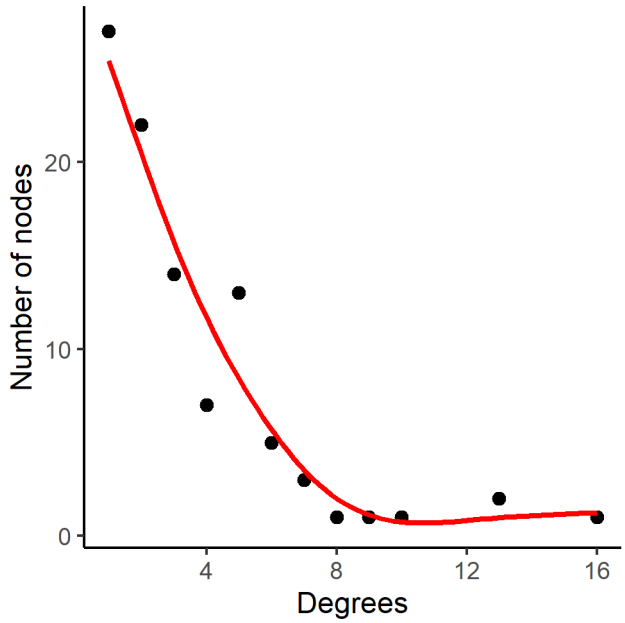

Figure 15. The degree distribution of nodes plotted against the number of nodes for the constructed upregulated PPI in VR colonies. The red line follows a power law.

To identify the hubs, nodes were ranked by degree and betweenness centrality, and the top nodes were considered potential biomarker candidates for *Varroa* resistance. The first-ranking node in terms of degrees is LOC413021, a gene that encodes papilin, a proteoglycan-like sulfated glycoprotein found in the extracellular matrix, with a degree of 16 and an LFC of 1.55 (Table 8). The second-ranking node in terms of degrees is GB55960-PA, a gene that encodes a Ras-related protein Rab-8A-like protein, with a degree of 13 and an LFC of 1.64 (Table 8). Sharing the second rank is Hsc70-4, a gene that encodes the heat shock protein cognate 4, with a LFC of 1.31 (Table 8). The third-ranking node in terms of degrees is LOC409590, a gene that encodes cyclin-dependent kinase 1, with a degree of 10 and an LFC of 6.27, the highest of all four nodes (Table 8). Hsc70-4 and LOC413021 rank first and second (Table 9), respectively, in terms of betweenness centrality and are therefore considered candidate biomarkers for VR in honey bees.

Table 8. Common genes mapped to the top 27 high-degree nodes in the largest VR upregulated network and their mean shrunken LFCs.

Gene	Description	C_D	LFC
LOC413021	papilin	16	1.55
GB55960-PA	ras-related protein Rab-8A-like	13	1.64
Hsc70-4	heat shock protein cognate 4	13	1.31
LOC409590	cyclin-dependent kinase 1	10	6.27
LOC408983	protein transport protein Sec23A	9	1.78

GB42970-PA	ras-related protein rab-5c	8	1.44
LOC408852	hypoxia-inducible factor 1-alpha homolog	7	2.20
LOC412150	heat shock protein 90kDa beta member 1, endoplasmin	7	2.09
LOC725950	PHD finger protein 20	7	1.57
LOC408552	collagen alpha-1(IV) chain	6	1.50
LOC408950	malate dehydrogenase, mitochondrial	6	1.33
LOC726692	histone H3.3	6	1.32
LOC410022	mitochondrial-processing peptidase subunit beta	6	1.31
LOC412345	exosome complex component RRP41	6	1.25
LOC413592	probable ATP-dependent RNA helicase DHX34	5	3.42
LOC411923	alanine-tRNA ligase, cytoplasmic	5	2.12
LOC551958	---	5	2.04
LOC726818	beta-hexosaminidase subunit beta	5	1.91
LOC413880	peroxisomal targeting signal 1 receptor	5	1.79
LOC409155	dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	5	1.66
LOC551824	putative pre-mRNA-splicing factor ATP-dependent RNA helicase PRP1	5	1.55
LOC412505	translocation protein SEC63 homolog	5	1.50
GB47314-PA	ATP-dependent RNA helicase DDX5-like	5	1.50
Vamp7	vesicle-associated membrane protein 7	5	1.42
GB43482-PA	ATP synthase subunit b, mitochondrial-like	5	1.39
LOC409613	GTP-binding protein SAR1	5	1.32
LOC551859	calmodulin	5	1.23

Table 9. Common genes mapped to the top 27 nodes with high betweenness centralities in the largest VR upregulated network and their mean shrunken LFCs.

Gene	Description	C_B	LFC
Hsc70-4	heat shock protein cognate 4	0.27	1.31
LOC413021	papilin	0.26	1.55
LOC726818	beta-hexosaminidase subunit beta	0.24	1.91
LOC408852	hypoxia-inducible factor 1-alpha homolog	0.21	2.20
GB55960-PA	ras-related protein Rab-8A-like	0.21	1.64
LOC550929	serine/threonine-protein kinase PINK1, mitochondrial	0.19	1.66
LOC409590	cyclin-dependent kinase 1	0.18	6.27
LOC409709	putative glucosylceramidase 4	0.18	1.71
LOC726692	histone H3.3	0.14	1.32
LOC552163	dolichyl pyrophosphate Man9GlcNAc2 alpha-1,3-glucosyltransferase	0.11	1.51
LOC551437	alpha-N-acetylglucosaminidase	0.10	1.80
LOC413880	peroxisomal targeting signal 1 receptor	0.09	1.79
LOC412345	exosome complex component RRP41	0.09	1.25
LOC725950	PHD finger protein 20	0.09	1.57
LOC726210	myogenesis-regulating glycosidase	0.08	1.26

LOC412150	Heat shock protein 90kDa beta member 1, endoplasmin	0.08	2.09
LOC551581	N-acetylgalactosamine kinase	0.08	1.93
LOC409814	beta-glucuronidase	0.07	1.59
LOC409281	protein sel-1 homolog 1	0.07	1.31
Gfat2	glucosamine-fructose-6-phosphate aminotransferase 2	0.06	1.79
LOC413592	probable ATP-dependent RNA helicase DHX34	0.05	3.42
LOC551824	putative pre-mRNA-splicing factor ATP-dependent RNA helicase PRP1	0.05	1.55
GB42970-PA	Ras-related protein rab-5c	0.04	1.44
LOC411923	alanine-tRNA ligase, cytoplasmic	0.04	2.12
LOC410022	mitochondrial-processing peptidase subunit beta	0.04	1.31
LOC408564	long-chain fatty acid transport protein 4	0.03	1.76
LOC412815	fatty acid synthase	0.03	1.70

5. DISCUSSION

5.1. Honey Bee Disease and Pathogen Prevalence Across Turkey's Marmara Region

Varroa destructor, also known as *Varroa* mite, is an ectoparasitic mite that is one of the major drivers of the global decline in honey bee health, leading to drastic reduction in honey production and lifespan of honey bees. To combat the global spread of *Varroa* mites and their negative impacts, several projects have been carried out to generate *Varroa*-resistant (VR) honey bee populations via selective breeding for low levels or absence of *Varroa* mite infestation, or traits associated with reduced *Varroa* mite levels, for generations around the globe. One such ongoing project is the establishment of VR honey bee lines via natural mating based on low *Varroa* mite levels from an isolated Anatolian honey bee (*Apis mellifera anatoliaca*) population on Marmara Island, Turkey. In this project, we evaluate *Varroa* mite and disease/pathogen prevalence across all sampled colonies and between *Varroa*-sensitive (VS) and VR honey bee colonies.

Here, we show that at least one *Varroa* mite is present in 90.4% of colonies sampled located in the southern Marmara region. It is important to point out that the sugar-roll method used here can recover up to 94% of *Varroa* mites from sampled nurse bees per colony (Çakmak et al., 2011). Consequently, colonies with low levels of mites cannot be classified as *Varroa* mite-positive and mite counts from *Varroa* mite-positive colonies are 94% accurate. Therefore, it is possible that all colonies sampled are positive for *Varroa* mites and that mite counts are larger than what we found. Nevertheless, the prevalence of *Varroa* mites reported here is alarmingly higher than ones reported by two large-scale studies that examined *Varroa* mite prevalence across all regions of Turkey in different years (Çakmak, Aydin, et al., 2003; Ütük et al., 2011), and a study that examined honey bee diseases in the southern Marmara region (Çakmak, Aydın, et al., 2003). Two

more recent large-scale studies identified a *Varroa* mite prevalence rate as low as 14.7% in Kırklareli (Bayrakal et al., 2020) and as high as 100% (D. Muz & Muz, 2017) in Tekirdağ from the northern Marmara region. The increase in *Varroa* mite prevalence over the years has been also observed in Hatay and Adana from the Mediterranean region, where it was as low as 32% in 2003 (Sahinler & Gul, 2005) and as high as 100% in 2006-2007 (M. N. Muz et al., 2012; Yalçınkaya & Keskin, 2010). This is similar to the increase seen in the Eastern Anatolia region, where it was as low as 25.6% between 2002-2004 in Elaziğ (Şimşek, 2005) and as high as 93-100% in Erzurum (Balkaya, 2016), Hakkari (Aydin, 2012), and Kars (Önk & Kılıç, 2014). The current *Varroa* mite prevalence, as well as its increase over the years, is consistent with that reported for other countries such as Estonia (Mõtus et al., 2016), Norway (Dahle, 2010), Uruguay (Anido et al., 2015; Antúnez et al., 2017), and the US (Kulhanek et al., 2017; Seitz et al., 2015; Traynor et al., 2016). We also show that VR colonies had significantly less counts of *Varroa* mites than VS ones, suggesting the ability of VR colonies to reduce *Varroa* mite burden (i.e., resistance). However, it possible that, since the VR population is located on the Marmara Island, its isolation from the mainland could play a role in lower *Varroa* mite levels. Twelve sampled VS colonies located on the Island had a median *Varroa* mite level higher than that of VR colonies. However, we believe this does not entirely eliminate the possibility of isolation affecting *Varroa* mite levels.

Varroa mites are active and passive vectors of more than a dozen bee-infecting viruses, the most common of which are deformed wing virus (DWV), Kashmir bee virus (KBV), Israeli acute paralysis virus (IAPV), chronic bee paralysis virus (CBPV), acute bee paralysis virus (ABPV), Bee Macula-like virus (BeeMLV), Lake Sinai virus (LSV), Varroa destructor virus-1 (VDV-1), black queen cell virus (BQCV), sacbrood virus (SBV), Kakugo virus (KV) (Fujiyuki et al., 2006), slow bee paralysis virus (SBPV), aphid lethal paralysis virus (ALPV), and Big Sioux River virus (BSRV). Therefore, we screened for those viruses using RNA sequencing (RNA-seq). Contrary to studies that reported a prevalence rate between 0-35.5% for ABPV, 20.2-32% for BQCV, 23-44.7% for DWV, 0-25% for CBPV, 0-6.5% for IAPV, 2.7-22.3% for SBV and 0% for KBV in different regions of Turkey (Çağirgan & Yazıcı, 2021; Kalayci et al., 2020; Okur Gumusova et al., 2010; Rüstemoğlu & Sipahioğlu, 2019); we found alarmingly higher prevalence rate for ABPV (74.8%), BQCV (100%), DWV (100%), and absence of CBPV and IAPV. We detected SBV and KBV in 7% and 2.4% of the sampled colonies respectively.

Additionally, we report a prevalence rate of 56.5% for LSV, a virus that was very recently detected in *Varroa* mites sampled from İzmir and Muğla (Çağirgan et al., 2022), 99.1% for KV, and 100% for VDV1. Except for ABPV, BQCV, DWV, KV, and VDV1, viral levels were very low, suggesting that those infections are covert (i.e., lack of symptoms with minimal or no effect on performance and lifespan). ABPV levels were significantly higher in VR colonies compared to VS colonies, while DWV, KV, and VDV1 levels were significantly higher in VS colonies compared to VR colonies. Interestingly, BQCV levels were not significantly different between VR and VS colonies.

Stonebrood (SB) is a honey bee disease caused by *Aspergillus* spp., two of which, *A. flavus* and *A. fumigatus*, are the primary pathogens. SB is rare because natural *A. flavus* and *A. fumigatus* infections are unsuccessful, though they can occasionally multiply, in honey bee colonies (Bailey, 1968). Thus, SB is considered to be of minor importance to beekeepers. Here, we show that *Aspergillus* spp. that cause SB are present in 65.2% of colonies, from which nurse bees were examined, located in the southern Marmara region. SB is significantly more prevalent in VS compared to VR colonies. Previous studies reported SB prevalence of 2%, 4.5%, 6.4%, and 14.3% in Erzurum from the Eastern Anatolia region (Balkaya, 2016), and Tekirdağ (Sıralı & Doğaroğlu, 2005), Istanbul (Dümen et al., 2013), and Kırklareli (Bayrakal et al., 2020) from the northern Marmara region, respectively. This contrast between prevalence rates based on clinical symptoms of SB and the presence of *Aspergillus* spp. that cause SB is expected because *Aspergillus* spp. can be present in adult bees from non-SB-infected colonies. Alone, these *Aspergillus* spp. cannot establish a disease outbreak unless colonies are affected by multiple stressors (e.g., *Varroa* mite infestation, viral infections, etc.) that compromise individual and social immunity. However, it has been shown that *A. flavus* can overcome immune responses and establish an infection (K. Foley et al., 2014). Since we did not examine colonies for the presence of clinical symptoms of SB, we cannot ascertain its true prevalence. Nevertheless, for the purpose of this study, it is sufficient to detect the presence of SB-causing *Aspergillus* spp. to account for their presence in differential gene expression analysis.

Chalkbrood (CB) is a honey bee brood disease caused by the fungi *Ascosphaera apis*, often regarded as an opportunistic pathogen. Both its outbreak and severity depends on a multitude of interacting stress-related factors, as in SB (Evison, 2015). Adult bees, though responsible for spore transmission, are not susceptible to *A. apis*. The pathogen, however,

can reside in adult bees (Borum & Ülgen, 2010; Maxfield-Taylor et al., 2015) as nurse bees transmit the spore-contaminated food, in addition to being in close proximity to the infected brood, and removing the infected/dead larvae. Therefore, we screened for CB in nurse bees using PCR and found none of the sampled colonies to be infected. This is the first study that reports the absence of CB in the Marmara region, as many studies that examined CB prevalence located in the Marmara region between 2003-2013 reported a prevalence rate up to 36.3% (Borum & Ülgen, 2010; Çakmak, Aydın, et al., 2003; Dümen et al., 2013; Sıralı & Doğaroğlu, 2005). However, our results are in concordance with that from a recent nation-wide study which reported a prevalence rate of 2.2% for CB (Sevim et al., 2022).

American foulbrood (AFB) and European foulbrood (EFB) are two honey bee brood diseases caused by the bacteria *Paenibacillus larvae* and *Melissococcus plutonius*, respectively. Worker bees, though not susceptible, can be used for the monitoring of AFB and EFB (Gillard et al., 2008; Roetschi et al., 2008). We therefore screened for AFB and EFB in nurse bees using qPCR and found a prevalence rate of 31.3% and 6.1%, respectively, in sampled colonies from the southern Marmara region, with AFB being significantly more prevalent in VS compared to VR colonies. A southern Marmara-region wide study examining AFB and EFB prevalence microbiologically in 2001 did not detect *P. larvae*, but detected *M. plutonius* in 5% of the sampled colonies (Çakmak, Aydın, et al., 2003). An Istanbul-wide study examining honey samples with combs detected *P. larvae* and *M. plutonius* in 3.2% and 5.8% of the samples, respectively (Dümen et al., 2013). Both of these studies point out to the low prevalence of AFB and EFB compared to that reported by a large-scale study that found 29% and 19% prevalence for AFB and EFB, respectively, in Hatay and Adana between 2006-2007 (Yalçınkaya & Keskin, 2010). As with SB, molecular detection of *P. larvae* and *M. plutonius* is far more sensitive than microbiological or visual examinations, and their presence in colonies does not indicate that those colonies are displaying clinical symptoms. Since we did not examine colonies for the presence of clinical symptoms of AFB and EFB, we cannot ascertain their prevalence. Nevertheless, for the purpose of this study, it is sufficient to detect the presence of *P. larvae* and *M. plutonius* to account for their presence in differential gene expression analysis.

Nosema is a globally prevalent adult honey bee disease caused by the two microsporidian parasite species *Nosema ceranae* and *Nosema apis*. In this study, we found none of the

sampled colonies to be infected with *N. apis*, and *N. ceranae* in 64.3% of all sampled colonies located in the southern Marmara region. The *N. ceranae* prevalence rate reported here is higher than what was reported by previous studies in the Marmara region (L. Aydin et al., 2005; Çakmak, Aydın, et al., 2003; D. Muz & Muz, 2017; Sıralı & Doğaroğlu, 2005) but consistent with that reported by recent studies in the Aegean (Kartal et al., 2021), Black Sea (Yilmaz et al., 2018), and Eastern Anatolia region (Oğuz et al., 2017). Our results all also consistent with that reported from Belgium (Matthijs et al., 2020), Iran (Mohammadian et al., 2018), Bulgaria (Shumkova et al., 2018), Canada (Emsen et al., 2016, 2020), and the US (Traynor et al., 2016). More importantly, we found that *N. ceranae* is significantly less abundant in VR compared to VS colonies.

Altogether, we show that the majority of colonies are infested with *Varroa* mites and infected with *N. ceranae*, *Aspergillus* spp., ABPV, BQCV, DWV, KV, and VDV1. Since *Aspergillus* spp. can only cause SB in colonies that are weakened by multiple stressors, it is possible that those colonies with *Varroa* mite infestation and multiple infections to develop symptoms of SB. In addition, we show that the prevalence of SB and EFB, and the levels of *Varroa* mites. *N. ceranae*, DWV, KV, and VDV1 are significantly less in VR compared to VS colonies, suggesting that VR colonies have a mechanism that combats stressors in general. Interestingly, we found that VS colonies sampled from Karacabey, compared to VS colonies sampled from other locations, have the lowest prevalence of *Varroa* mites, SB, and *N. ceranae*; and lowest levels of *Varroa* mites and *N. ceranae*. This reflects the efficacy of treatments in controlling spread and levels of pathogens in colonies located in Karacabey.

To further evaluate disease-*Varroa* resistance/sensitivity as well as disease-disease relationships, we built a generalized linear mixed model (GLMM) or generalized linear model (GLM) for each pathogen while including location and batch as random or fixed effects, respectively. The *Varroa*-specific GLMM revealed a positive relationship between *Varroa* mite counts and sensitivity to *Varroa* as well as SB prevalence. The former confirms our initial observation that VR colonies have lower *Varroa* mite counts, and the latter, also seen in the SB-specific GLMM, indicates a robust relationship between *Varroa* mites and SB. This also suggests an indirect relationship between SB and *Varroa* resistance, as VR colonies have significantly less *Varroa* mite counts. We also found a positive relationship between *Varroa* mite counts and ABPV levels from the *Varroa*-specific GLMM, a relationship that was seen in the ABPV-specific GLM and thus

suggesting its robustness. We found no relationships between AFB or Nosema with *Varroa* resistance or the other pathogens. This could be explained by the use of medications in VS colonies, especially the ones located in Karacabey, to treat bacterial and fungal bee diseases such as AFB/EFB and Nosema, respectively.

As for viruses, ABPV is positively associated with *Varroa* mite counts and the interaction between VDV1 and DWV. The former was not seen in the *Varroa*-specific GLMM, indicating the relationship's lack of robustness. On the other hand, ABPV is negatively associated with sensitivity to *Varroa*, DWV, and the interaction between KV and VDV1. The Relationship between ABPV and *Varroa* resistance confirms our initial results which show that VR colonies have significantly higher levels of ABPV, while the relationship between ABPV and DWV was not seen in the DWV-specific GLMM but rather in the KV-specific one, suggesting that ABPV-DWV and KV-ABPV relationships are not robust. The negative relationship between ABPV and the interaction between VDV1 and KV could be explained by a competition between KV-VDV1 and ABPV for resources. BQCV, similar to AFB and Nosema, was not associated with *Varroa* mite counts, *Varroa* resistance, or other pathogens. Interestingly, both DWV and KV are associated positively with sensitivity to *Varroa*, confirming our initial results that show VR colonies having lower levels of both DWV and KV. Furthermore, DWV, KV, and VDV1 levels are positively associated with each other, as revealed by the DWV-, KV-, and VDV1-specific GLMMs. Such relationships are a likely outcome if one virus weakens the immunity of the host, allowing the other virus to thrive. The DWV-, KV-, and VDV1-specific GLMMs revealed negative association between DWV and the interaction between KV and VDV1, KV and the interaction between DWV and VDV1, and VDV1 and the interaction between DWV and KV, respectively. These results show that high levels of two of the three viruses both have positive and negative effects on the levels of the third virus. Surprisingly, VDV1 was not positively associated with sensitivity to *Varroa*, but rather with *Varroa* mite counts, a relationship that was not seen in the *Varroa*-specific GLMM. However, VDV1 levels are positively associated with the island as a location, where most of the colonies are VR. This points to an indirect relationship between VDV1 and *Varroa* resistance that needs to be further examined using a larger sample size. VDV1 is also associated negatively with SB, a relationship that is not robust as it was not seen in the SB-specific GLMM.

5.2. Candidate Biomarkers from Differentially Expressed Genes in *Varroa* Resistant Colonies

We performed differential gene expression (DGE) analysis using DESeq2, and R package that models raw read counts for each gene using a generalized linear model (GLM). We found that the best model, selected through a series of comparisons between full and reduced models via likelihood ratio tests, contained geographical location, SB presence, DWV levels, KV levels, *Varroa* resistance, interaction between SB presence and *Varroa* resistance, interaction between DWV levels and interaction between KV levels and *Varroa* resistance as features. This indicates that *Varroa* mite infestation, AFB and EFB presence, and levels of *N. ceranae*, ABPV, BQCV, and VDV1 do not have significant explanatory effects on gene expression. We identified 587 significantly upregulated and 44 significantly downregulated genes in VR compared to VS colonies that are SB-negative, and 329 significantly upregulated and 54 significantly downregulated genes in VR compared to VS colonies that are SB-positive. We accounted for additional variance that could be attributed to disease and other elements via log2 fold change (LFC) shrinkage. Additionally, we further stratified VR colonies into *Varroa* resistant with low load (VRL) if their *Varroa* mite count ≤ 3 and *Varroa* resistant with high load (VRH), and VS colonies into *Varroa* sensitive with low load (VSL) if their *Varroa* mite counts < 10 and *Varroa* sensitive with high load (VSH) otherwise and performed DGE on VRL and VSH colonies. We identified 537 significantly upregulated and 147 significantly downregulated genes in VRL compared to VSH colonies that are SB-negative, and 361 significantly upregulated and 127 significantly downregulated genes in VRL compared to VSH colonies that are SB-positive. Finally, to identify genes differentially expressed regardless of SB and *Varroa* mite count, we grouped significantly upregulated genes in VR common across all comparisons into one set, and significantly downregulated genes in VR that are also common across all comparisons into another. We found 278 common significantly upregulated and 23 common significantly downregulated gene.

The top common upregulated gene in VR encodes CCAAT-enhancer-binding protein gamma (C/EBP-γ), a transcription factor from the C/EBP family that participates in a wide range of physiologic processes. In insects, the C/EBP family was shown to play a role in embryonic development in the fruit fly *Drosophila melanogaster* (Rørth &

Montell, 1992), and nitric oxide synthase (NOS) transcription in the silkworm, *Bombyx mori* (Furukawa et al., 2012). NOS generates nitric oxide (NO), a molecular that is important for the induction of innate immunity in *Drosophila* (E. Foley & O'Farrell, 2003; Nappi et al., 2000) and honey bees (Negri et al., 2013). NO was shown to be emitted by European beewolf *Philanthus Triangulum* to protect the brood and paralyzed honey bees against mold fungi (Strohm et al., 2019). Interestingly, NO was also shown to be implicated in learning, memory, and chemosensory processing in honey bees (Müller, 1997). Heightened chemosensing in nurse bees is crucial for the removal of diseased (*Varroa* mite-infested, viral infected, etc.) brood (Wagoner et al., 2019). A more directed link between C/EBP and insect immunity comes from a study which showed that C/EBP is required for the transcription of several Defensin genes, a class of antimicrobial peptides, in the Aedes aegypti (Meredith et al., 2006). All these studies establish a link between the C/EBP family, not C/EBP-γ specifically, and insect immunity. In the invertebrate *Caenorhabditis elegans*, C/EBP-γ was shown to be required for the activation of surveillance or "effector-triggered" immunity to reduce pathogen levels and promote survival (Reddy et al., 2016). *C. elegans*, like insects, lack adaptive immune system and thus relies on the innate one. Altogether, it appears that C/EBP-γ is important for immune response activation, reduction of pathogen levels, and survival. Thus, it could serve as a candidate biomarker for overall resistance against pathogens. However, C/EBP-γ is universally understudied (Renfro et al., 2022), and further experiments are required to elucidate the effects of C/EBP-γ on survival and immune response against multiple pathogens, specifically, *Varroa* mites, and its utility as a resistance biomarker in honey bees.

The top two common downregulated genes in VR encode the Nkx-6.1 and Sox-21-B. Nkx-6.1 a homeobox protein that is implicated in embryonic central nervous system development in *Drosophila* (Uhler et al., 2002). Sox-21-B is a transcription factor—homologous to the mammalian Sox2—that is implicated in honey bee embryogenesis and development, sensory processing, and memory formation (Wilson & Dearden, 2008). Both genes are understudied, and thus further experiments are required to elucidate their functions in adult honey bees and its relation to stress response and survival.

5.3. Candidate Biomarkers from an Upregulated Biological Network in *Varroa* Resistant Colonies

In contrast to DGE analysis-based biomarker identification, a network-based approach for biomarker identification focuses on the overall change of DEGs from VS to VR on a system-level. To that end, we queried common significantly upregulated and downregulated genes in VR separately against the STRING database to build protein-protein interaction (PPI) networks. Of the 278 common significantly upregulated genes, only 182 were found in the STRING database, of which 94 formed a large network and 3 formed a small one. The common significantly downregulated genes did not form a network. The upregulated PPI network was enriched in KEGG pathways associated with metabolic pathways, specifically amino acid and nucleotide sugar metabolism. These enriched pathways were also observed in the brains of *Apis cerana*, the native host of *Varroa* mite (Wu et al., 2017). It is worth noting that *A. cerana* survives *Varroa* mite infestation with minimal damage, suggesting that the rewiring of several metabolic pathways could be essential for *Varroa* resistance. The network was also enriched in genes participating processing, folding, glycosylating, and re-folding proteins, and directing terminally-misfolded proteins for degradation in the endoplasmic reticulum (Bravo et al., 2013). We found enriched GO terms associated with different metabolic processes, regulation of protein localization, proteolysis, protein secretion, extracellular matrix organization. A study that examined the proteome of hemolymph, mushroom bodies, and antennae from honey bees displaying *Varroa* sensitive hygiene revealed similarly enriched pathways and GO terms (Hu et al., 2016).

Topological features of the network revealed its scale-free property (i.e., the vast majority of the network's nodes are of low degree, while a small number of nodes are of high degree). The high-degree nodes are called hubs and they are of high importance because their loss causes the breakdown of the network (Albert & Barabasi, 2002). The top two genes in terms of degree encode Papilin (LOC413021) and Hsc70-4. Both of those genes rank in the top in terms of betweenness centrality, a measure that quantifies the amount of influence a gene has on the flow of information across the network. Hsc70-4 is involved in the heat shock response, a cellular mechanism that prevents proteotoxicity, that is induced in response to viral infections in honey bees (McMenamin et al., 2020). Papilin is a large proteoglycan-like sulfated glycoprotein found in the extracellular matrix

that is important for embryonic development of *D. melanogaster* and *C. elegans* (Fessler et al., 2004). It was shown that migrating hemocytes, cells that play a role in the immune system of insects, produce and secrete papilin in abundance prior to the formation of basal membranes around tissues during development (Kramerova et al., 2003), and when immune response is activated to synthesize a temporary extracellular matrix that acts as a scaffold (Kramerova et al., 2000). It was recently proposed that papilin is involved in wound repair and tissue remodeling in the silkworm *B. mori* (Feng et al., 2022), a set of processes required for fast recovery from the effects of *Varroa* mite feeding in honey bees.

6. CONCLUSION

In conclusion, we discovered that the parasitic mite *Varroa destructor* (the *Varroa* mite), the fungal pathogens *Aspergillus* spp. (SB) and *N. ceranae* (Nosema), the bacterium *Paenibacillus larvae* (AFB), and the viruses acute bee paralysis virus (ABPV), black queen cell virus (BQCV), deformed wing virus (DWV), Kakugo virus (KV), Lake Sinai virus (LSV), and Varroa destructor virus 1 (VDV1) to be highly prevalent in colonies located in the southern Marmara region of Turkey. More importantly, we found that honey bee colonies that have been bred for lower levels of *Varroa* mites—but not traits linked to reduced levels of *Varroa* mites—on the Marmara Island for 15 years have lower prevalence of SB, higher levels of ABPV, and lower levels of *Varroa* mites, Nosema, DWV, KV, and VDV1. These results suggest that these colonies are resistant rather than tolerant to *Varroa* mites. We also found *Varroa* mite infestation to be associated directly with higher prevalence SB and higher levels of ABPV and indirectly with higher levels of DWV, KV, and VDV1. Interestingly, we observed that higher levels of any two viruses in the DWV-KV-VDV1 complex have synergistic (through the individual effect of each virus) and antagonistic (through the interaction effect between the two viruses) effects on the level of the third virus. Using differential gene expression analysis on RNA-seq from whole-bee samples coupled with a network-based approach, we built a protein-protein interaction (PPI) network associated with *Varroa* resistance. We found this PPI network to be enriched in multiple pathways and gene sets related to metabolism, indicating rewiring of metabolism, as well as protein processing. Most of the enriched pathways and gene sets were in concordance with multiple studies that examined molecular mechanisms in specific tissues of eastern honey bees (naturally resistant to *Varroa*) and *Varroa* resistant western honey bees. We calculated the topological features of the network and found papilin (LOC413021) and heat shock protein cognate 4 (Hsc70-4) to be the most important hubs. Papilin is involved in embryonic development and innate immunity. It is also involved in tissue remodeling, specifically wound healing, a possibly

important process for fast recovery from the effects of *Varroa* mite feeding in honey bees. Hsc70-4 is involved in antiviral response. We revisited the differentially expressed genes in *Varroa* resistant (VR) colonies and found the top upregulated gene in VR colonies encodes the CCAAT-enhancer-binding protein gamma (C/EBP-γ), a transcription factor that is involved in embryonic development, immunity, and survival. Our analyses, coupled with literature search, suggest that C/EBP-γ, papilin, and Hsc70-4 are candidate biomarkers for *Varroa* resistance. Although the approach we used here is novel and allows the detection of biomarkers of greater utility, further investigation is needed to identify the roles of these genes and the pathways they participate in. It is also of greater importance to examine whether their downregulation would result in higher levels of *Varroa* mites before utilizing them for marker-assisted selection for *Varroa* resistance.

7. BIBLIOGRAPHY

Albert, R. (2005). Scale-free networks in cell biology. *Journal of Cell Science*, *118*(21), 4947–4957. https://doi.org/10.1242/jcs.02714

Albert, R., & Barabasi, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, *74*(1), 47–97. https://doi.org/10.1103/RevModPhys.74.47

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), R106. https://doi.org/10.1186/gb-2010-11-10-r106

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Anido, M., Branchiccela, B., Castelli, L., Harriet, J., Campá, J., Zunino, P., & Antúnez, K. (2015). Prevalence and distribution of honey bee pests and pathogens in Uruguay. *Journal of Apicultural Research*, *54*(5), 532–540. https://doi.org/10.1080/00218839.2016.1175731

Antúnez, K., Invernizzi, C., Mendoza, Y., vanEngelsdorp, D., & Zunino, P. (2017). Honeybee colony losses in Uruguay during 2013–2014. *Apidologie*, *48*(3), 364–370. https://doi.org/10.1007/s13592-016-0482-2

Aranda, P. S., LaJoie, D. M., & Jorcyk, C. L. (2012). Bleach gel: A simple agarose gel for analyzing RNA quality. *Electrophoresis*, *33*(2), 366–369. PubMed. https://doi.org/10.1002/elps.201100335

Aronstein, K. A., & Murray, K. D. (2010). Chalkbrood disease in honey bees. *Journal of Invertebrate Pathology*, *103*, S20–S29. https://doi.org/10.1016/j.jip.2009.06.018

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Aydin, A. (2012). Hakkari Yöresinde Varroasis'in yaygınlığı. *Yüzüncü Yıl Üniversitesi Veteriner Fakültesi Dergisi*, *23*(3), 129–130.

Aydin, L., Cakmak, I., Gulegen, E., & Wells, H. (2005). Honey bee nosema disease in the Republic of Turkey. *Journal of Apicultural Research*, *44*(4), 196–197. https://doi.org/10.1080/00218839.2005.11101179

Aydın, A. (1998). *Van yöresinde bal arılarında Varroa jacobsoni'nin epidemiyolojisi üzerine araştırmalar*. Yüzüncü Yıl Üniversitesi Sağlık Bilimleri Enstitüsü.

Bailey, L. (1968). Honey Bee Pathology. *Annual Review of Entomology*, *13*(1), 191–212. https://doi.org/10.1146/annurev.en.13.010168.001203

Balkaya, İ. (2016). Erzurum Yöresi Arıcılarının Karşılaştıkları Bal Arısı Hastalıkları. *Atatürk Üniversitesi Veteriner Bilimleri Dergisi*, *11*(3), 273–281. https://doi.org/10.17094/ataunivbd.282981

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, Articles*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bayrakal, G., Dümen, E., Ekici, G., Akkaya, H., & Sezgin, F. H. (2020). Detection and Molecular Examination of Pathogens in Honey and Bees in the Northern Marmara Region, Turkey. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, *26*(3), 313–319. https://doi.org/10.9775/kvfd.2019.22845

Ben Abdelkader, F., Çakmak, İ., Çakmak, S. S., Nur, Z., İncebıyık, E., Aktar, A., & Erdost, H. (2021). Toxicity assessment of chronic exposure to common insecticides and bee medications on colony development and drones sperm parameters. *Ecotoxicology*, *30*(5), 806–817. https://doi.org/10.1007/s10646-021-02416-3

Beyazıt, A., Akkoca, N., Eskiizmirliler, S., Albayrak, H., Özan, E., Özden, M., Selver, M., & Tunalıgil, S. (2012). Ege Bölgesi illerinde önemli arı hastalıklarının yaygınlığının araştırılması. *Hayvan Sağlığı Program Değerlendirme Kitapçığı*, 366.

Bienefeld, K. (2016). Breeding Success or Genetic Diversity in Honey Bees? *Bee World*, *93*(2), 40–44. https://doi.org/10.1080/0005772X.2016.1227547

Borum, A. E., & Ülgen, M. (2010). Güney Marmara Bölgesindeki Bal Arılarının Chalkbrood (Ascosphaera apis) İnfeksiyonunda Predispozisyon Faktörleri. *Uludağ Arıcılık Dergisi*, *10*(2), 56–69.

Bravo, R., Parra, V., Gatica, D., Rodriguez, A. E., Torrealba, N., Paredes, F., Wang, Z. V., Zorzano, A., Hill, J. A., Jaimovich, E., Quest, A. F. G., & Lavandero, S. (2013). Endoplasmic Reticulum and the Unfolded Protein Response: Dynamics and Metabolic Integration. *International Review of Cell and Molecular Biology*, *301*, 215–290. https://doi.org/10.1016/B978-0-12-407704-1.00005-1

Brooks, M. E., Kristensen, K., Benthem, K. J. van, Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). GlmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378–400.

Büchler, R. (1994). Varroa Tolerance in Honey Bees—Occurrence, Characters and Breeding. *Bee World*, *75*(2), 54–70. https://doi.org/10.1080/0005772X.1994.11099201

Burnham, A. J. (2019). Scientific Advances in Controlling Nosema ceranae (Microsporidia) Infections in Honey Bees (Apis mellifera). *Frontiers in Veterinary Science*, *6*. https://www.frontiersin.org/articles/10.3389/fvets.2019.00079

Çağirgan, A. A., Kaplan, M., Pekmez, K., & Arslan, F. (2022). Türkiye'de Varroa Akarlarında Lake Sinai Virus (LSV)' un İlk Tespiti. *Atatürk Üniversitesi Veteriner Bilimleri Dergisi*, *17*(1), 16–19.

Çağirgan, A. A., & Yazıcı, Z. (2021). The prevalence of seven crucial honeybee viruses using multiplex RT-PCR and theirphylogenetic analysis. *Turkish Journal of Veterinary & Animal Sciences*, *45*(1), 44–55. https://doi.org/10.3906/vet-2004-139

Çakmak, İ. (2010). The over wintering survival of highly Varroa destructor infested honey bee colonies determined to be hygienic using the liquid nitrogen freeze killed brood assay. *Journal of Apicultural Research*, *49*(2), 197–201. https://doi.org/10.3896/IBRA.1.49.2.09

Çakmak, İ., Aydin, L., Gulegen, E., & Wells, H. (2003). Varroa (Varroa destructor) and tracheal mite (Acarapis woodi) incidence in the Republic of Turkey. *Journal of Apicultural Research*, *42*(4), 57–60. https://doi.org/10.1080/00218839.2003.11101093

Çakmak, İ., Aydın, L., & Güleğen, A. E. (2003). Güney marmara bölgesinde balarısı zararlı ve hastalıkları. *Uludag Bee Journal*. http://acikerisim.uludag.edu.tr/jspui/handle/11452/15797

Çakmak, İ., Çakmak, S. S., Fuchs, S., & Yeninar, H. (2011). Balarısı Kolonilerinde Varroa Bulaşıklık Seviyesinin Belirlenmesinde Pudra Şekeri ve Deterjan Yönteminin Karşılaştırılması. *Uludağ Arıcılık Dergisi*, *11*(2), 63–68.

Cakmak, I., & Fuchs, S. (2013). Exploring a treatment strategy for long-term increase of varroa tolerance on Marmara Island, Turkey. *Journal of Apicultural Research*, *52*(5), 242–250. https://doi.org/10.3896/IBRA.1.52.5.11

Calderón, R. A., van Veen, J. W., Sommeijer, M. J., & Sanchez, L. A. (2010). Reproductive biology of Varroa destructor in Africanized honey bees (Apis mellifera). *Experimental and Applied Acarology*, *50*(4), 281–297. https://doi.org/10.1007/s10493-009-9325-4

Carreck, N. L., Ball, B. V., & Martin, S. J. (2010). Honey bee colony collapse and changes in viral prevalence associated with Varroa destructor. *Journal of Apicultural Research*, *49*(1), 93–94. https://doi.org/10.3896/IBRA.1.49.1.13

Celle, O., Blanchard, P., Olivier, V., Schurr, F., Cougoule, N., Faucon, J.-P., & Ribière, M. (2008). Detection of Chronic bee paralysis virus (CBPV) genome and its replicative RNA form in various hosts and possible ways of spread. *Virus Research*, *133*(2), 280–284. https://doi.org/10.1016/j.virusres.2007.12.011

Chantawannakul, P., Ward, L., Boonham, N., & Brown, M. (2006). A scientific note on the detection of honeybee viruses using real-time PCR (TaqMan) in Varroa mites collected from a Thai honeybee (Apis mellifera) apiary. *Journal of Invertebrate Pathology, 91*(1), 69–73. https://doi.org/10.1016/j.jip.2005.11.001

Chen, Y., Evans, J. D., Smith, I. B., & Pettis, J. S. (2008). Nosema ceranae is a long-present and wide-spread microsporidian infection of the European honey bee (Apis mellifera) in the United States. *Journal of Invertebrate Pathology*, *97*(2), 186–188. https://doi.org/10.1016/j.jip.2007.07.010

Chen, Y., Pettis, J. S., Evans, J. D., Kramer, M., & Feldlaufer, M. F. (2004). Transmission of Kashmir bee virus by the ectoparasitic mite Varroa destructor. *Apidologie*, *35*(4), 441–448. https://doi.org/10.1051/apido:2004031

Conte, Y. L., Vaublanc, G. de, Crauser, D., Jeanne, F., Rousselle, J.-C., & Bécard, J.-M. (2007). Honey bee colonies that have survived Varroa destructor. *Apidologie*, *38*(6), 566–572. https://doi.org/10.1051/apido:2007040

Dagenais, T. R. T., & Keller, N. P. (2009). Pathogenesis of Aspergillus fumigatus in Invasive Aspergillosis. *Clinical Microbiology Reviews*, *22*(3), 447–465. https://doi.org/10.1128/CMR.00055-08

Dahle, B. (2010). The role of Varroa destructor for honey bee colony losses in Norway. *Journal of Apicultural Research*, *49*(1), 124–125. https://doi.org/10.3896/IBRA.1.49.1.26

Daughenbaugh, K. F., Martin, M., Brutscher, L. M., Cavigli, I., Garcia, E., Lavin, M., & Flenniken, M. L. (2015). Honey Bee Infecting Lake Sinai Viruses. *Viruses*, *7*(6), 3285–3309. https://doi.org/10.3390/v7062772

de Miranda, J. R., Cordoni, G., & Budge, G. (2010). The Acute bee paralysis virus–Kashmir bee virus–Israeli acute paralysis virus complex. *Journal of Invertebrate Pathology*, *103*, S30–S47. https://doi.org/10.1016/j.jip.2009.06.014

De Miranda, J. R., Cornman, R. S., Evans, J. D., Semberg, E., Haddad, N., Neumann, P., & Gauthier, L. (2015). Genome Characterization, Prevalence and Distribution of a Macula-Like Virus from Apis mellifera and Varroa destructor. *Viruses*, *7*(7), 3586–3602. https://doi.org/10.3390/v7072789

Di Prisco, G., Annoscia, D., Margiotta, M., Ferrara, R., Varricchio, P., Zanni, V., Caprio, E., Nazzi, F., & Pennacchio, F. (2016). A mutualistic symbiosis between a parasitic mite and a pathogenic virus undermines honey bee immunity and health. *Proceedings of the National Academy of Sciences*, *113*(12), 3203–3208. https://doi.org/10.1073/pnas.1523515113

Di Prisco, G., Pennacchio, F., Caprio, E., Boncristiani, H. F., Evans, J. D., & Chen, Y. 2011. (2011). Varroa destructor is an effective vector of Israeli acute paralysis virus in the honeybee, Apis mellifera. *Journal of General Virology*, *92*(1), 151–155. https://doi.org/10.1099/vir.0.023853-0

Doncheva, N. T., Morris, J. H., Gorodkin, J., & Jensen, L. J. (2019). Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *Journal of Proteome Research*, *18*(2), 623–632. https://doi.org/10.1021/acs.jproteome.8b00702

Dümen, E., Akkaya, H., Öz, G., & Sezgin, F. (2013). Microbiological and parasitological quality of honey produced in İstanbul. *Turkish Journal of Veterinary & Animal Sciences*, *37*(5), 602–607. https://doi.org/10.3906/vet-1301-46

El-Nahhal, Y. (2020). Pesticide residues in honey and their potential reproductive toxicity. *Science of The Total Environment*, *741*, 139953. https://doi.org/10.1016/j.scitotenv.2020.139953

Emsen, B., De la Mora, A., Lacey, B., Eccles, L., Kelly, P. G., Medina-Flores, C. A., Petukhova, T., Morfin, N., & Guzman-Novoa, E. (2020). Seasonality of Nosema ceranae Infections and Their Relationship with Honey Bee Populations, Food Stores, and Survivorship in a North American Region. *Veterinary Sciences*, *7*(3), 131. https://doi.org/10.3390/vetsci7030131

Emsen, B., Guzman-Novoa, E., Hamiduzzaman, M. M., Eccles, L., Lacey, B., Ruiz-Pérez, R. A., & Nasr, M. (2016). Higher prevalence and levels of Nosema ceranae than Nosema apis infections in Canadian honey bee colonies. *Parasitology Research*, *115*(1), 175–181. https://doi.org/10.1007/s00436-015-4733-3

Erkan, C. (1998). *Van ili Bahçesaray ilçesi arıcılık faaliyetleri ve sorunları* [Yüzüncü Yıl Üniversitesi]. https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=gySaxfRwibLHFpB0kALQcg&no=gySaxfRwibLHFpB0kALQcg

Evison, S. E. (2015). Chalkbrood: Epidemiological perspectives from the host–parasite relationship. *Current Opinion in Insect Science*, *10*, 65–70. https://doi.org/10.1016/j.cois.2015.04.015

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Feng, M., Swevers, L., & Sun, J. (2022). Hemocyte Clusters Defined by scRNA-Seq in Bombyx mori: In Silico Analysis of Predicted Marker Genes and Implications for Potential Functional Roles. *Frontiers in Immunology*, *13*, 852702. https://doi.org/10.3389/fimmu.2022.852702

Fessler, J. H., Kramerova, I., Kramerov, A., Chen, Y., & Fessler, L. I. (2004). Papilin, a novel component of basement membranes, in relation to ADAMTS metalloproteases and ECM development. *The International Journal of Biochemistry & Cell Biology*, *36*(6), 1079–1084. https://doi.org/10.1016/j.biocel.2003.12.010

Foley, E., & O'Farrell, P. H. (2003). Nitric oxide contributes to induction of innate immune responses to gram-negative bacteria in Drosophila. *Genes & Development*, *17*(1), 115–125. https://doi.org/10.1101/gad.1018503

Foley, K., Fazio, G., Jensen, A. B., & Hughes, W. O. H. (2014). The distribution of Aspergillus spp. Opportunistic parasites in hives and their pathogenicity to honey bees. *Veterinary Microbiology*, *169*(3), 203–210. https://doi.org/10.1016/j.vetmic.2013.11.029

Forsgren, E. (2010). European foulbrood in honey bees. *Journal of Invertebrate Pathology*, *103*, S5–S9. https://doi.org/10.1016/j.jip.2009.06.016

Forsgren, E., Budge, G. E., Charrière, J.-D., & Hornitzky, M. A. Z. (2013). Standard methods for European foulbrood research. *Journal of Apicultural Research*, *52*(1), 1–14. https://doi.org/10.3896/IBRA.1.52.1.12

Fries, I., Imdorf, A., & Rosenkranz, P. (2006). Survival of mite infested (Varroa destructor) honey bee (Apis mellifera) colonies in a Nordic climate. *Apidologie*, *37*(5), 564–570. https://doi.org/10.1051/apido:2006031

Fujiyuki, T., Ohka, S., Takeuchi, H., Ono, M., Nomoto, A., & Kubo, T. (2006). Prevalence and Phylogeny of Kakugo Virus, a Novel Insect Picorna-Like Virus That Infects the Honeybee (Apis mellifera L.), under Various Colony Conditions. *Journal of Virology*, *80*(23), 11528–11538. https://doi.org/10.1128/JVI.00754-06

Furukawa, S., Tanaka, H., Sagisaka, A., Ishibashi, J., & Yamakawa, M. (2012). Both κB and C/EBP binding sites are indispensable for full expression of a nitric oxide synthase gene in the silkworm, Bombyx mori. *Journal of Insect Biotechnology and Sericology*, *81*(1), 1_013-1_020. https://doi.org/10.11416/jibs.81.1_013

Genersch, E. (2010). American Foulbrood in honeybees and its causative agent, Paenibacillus larvae. *Journal of Invertebrate Pathology*, *103*, S10–S19. https://doi.org/10.1016/j.jip.2009.06.015

Gillard, M., Charriere, J. D., & Belloy, L. (2008). Distribution of Paenibacillus larvae spores inside honey bee colonies and its relevance for diagnosis. *Journal of Invertebrate Pathology*, *99*(1), 92–95. https://doi.org/10.1016/j.jip.2008.05.010

Gisder, S., Aumeier, P., & Genersch, E. 2009. (2009). Deformed wing virus: Replication and viral load in mites (Varroa destructor). *Journal of General Virology*, *90*(2), 463–467. https://doi.org/10.1099/vir.0.005579-0

Govan, V. A., Allsopp, M. H., & Davison, S. (1999). A PCR detection method for rapid identification of Paenibacillus larvae. *Applied and Environmental Microbiology*, *65*(5), 2243–2245. PubMed. https://doi.org/10.1128/AEM.65.5.2243-2245.1999

Hamiduzzaman, M. Md., Guzman-Novoa, E., & Goodwin, P. H. (2010). A multiplex PCR assay to diagnose and quantify Nosema infections in honey bees (Apis mellifera). *Journal of Invertebrate Pathology*, *105*(2), 151–155. https://doi.org/10.1016/j.jip.2010.06.001

Hansen, K. D., Brenner, S. E., & Dudoit, S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research*, *38*(12), e131. https://doi.org/10.1093/nar/gkq224

Harbo, J. R., & Hoopingarner, R. A. (1997). Honey Bees (Hyntenoptera: Apidae) in the United States That Express Resistance to Varroa jacobsoni (Mesostigmata: Varroidae). *Journal of Economic Entomology*, *90*(4), 893–898. https://doi.org/10.1093/jee/90.4.893

Hartig, F. (2022). *DHARMa—Residual Diagnostics for HierARchical Models* (0.4.5) [R]. https://github.com/florianhartig/DHARMa (Original work published 2015)

Higes, M., Martín-Hernández, R., Botías, C., Bailón, E. G., González-Porto, A. V., Barrios, L., del Nozal, M. J., Bernal, J. L., Jiménez, J. J., Palencia, P. G., & Meana, A. (2008). How natural infection by Nosema ceranae causes honeybee colony collapse. *Environmental Microbiology*, *10*(10), 2659–2669. https://doi.org/10.1111/j.1462-2920.2008.01687.x

Higes, M., Martín-Hernández, R., Garrido-Bailón, E., González-Porto, A. V., García-Palencia, P., Meana, A., Del Nozal, M. J., Mayo, R., & Bernal, J. L. (2009). Honeybee colony collapse due to Nosema ceranae in professional apiaries. *Environmental Microbiology Reports*, *1*(2), 110–113. https://doi.org/10.1111/j.1758-2229.2009.00014.x

Higes, M., Meana, A., Bartolomé, C., Botías, C., & Martín-Hernández, R. (2013). Nosema ceranae (Microsporidia), a controversial 21st century honey bee pathogen. *Environmental Microbiology Reports*, *5*(1), 17–29. https://doi.org/10.1111/1758-2229.12024

Hu, H., Bienefeld, K., Wegener, J., Zautke, F., Hao, Y., Feng, M., Han, B., Fang, Y., Wubie, A. J., & Li, J. (2016). Proteome Analysis of the Hemolymph, Mushroom Body, and Antenna Provides Novel Insight into Honeybee Resistance against Varroa Infestation. *Journal of Proteome Research*, *15*(8), 2841–2854. https://doi.org/10.1021/acs.jproteome.6b00423

James, R. R., & Skinner, J. S. (2005). PCR diagnostic methods for Ascosphaera infections in bees. *Journal of Invertebrate Pathology*, *90*(2), 98–103. https://doi.org/10.1016/j.jip.2005.08.004

Jensen, A. B., Aronstein, K., Flores, J. M., Vojvodic, S., Palacio, M. A., & Spivak, M. (2013). Standard methods for fungal brood disease research. *Journal of Apicultural Research*, *52*(1), 1–20. https://doi.org/10.3896/IBRA.1.52.1.13

Jensen, A. B., Welker, D. L., Kryger, P., & James, R. R. (2012). Polymorphic DNA sequences of the fungal honey bee pathogen Ascosphaera apis. *FEMS Microbiology Letters*, *330*(1), 17–22. https://doi.org/10.1111/j.1574-6968.2012.02515.x

Kaftanoğlu, O., Kumova, U., Yeninar, H., & Özkök, D. (1995). *Türkiye'de balarısı (Apis mellifera L.) hastalıklarının dağılımı, koloniler üzerindeki etkileri ve entegre kontrol yöntemlerinin uygulanması* [Kesin Sonuç Raporu]. Ankara. Türkiye Bilimsel ve Teknik Araştırma Kurumu Veterinerlik ve Hayvancılık Araştırma Grubu. https://www.beyazkovan.com/?p=17138

Kalayci, G., Çağirgan, A. A., Kaplan, M., Pekmez, K., Beyazit, A., Ozkan, B., Yesıloz, H., & Arslan, F. (2020). The Role of Viral and Parasitic Pathogens Affected By Colony Losses in Turkish Apiaries. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, *26*(5), 671–677. https://doi.org/10.9775/kvfd.2020.24154

Kandemir, I., Kence, M., & Kence, A. (2000). Genetic and morphometric variation in honeybee (Apis mellifera L.) populations of Turkey. *Apidologie*, *31*(3), 343–356. https://doi.org/10.1051/apido:2000126

Kandemir, I., Kence, M., Sheppard, W. S., & Kence, A. (2006). Mitochondrial DNA variation in honey bee (Apis mellifera L.) populations from Turkey. *Journal of Apicultural Research*, *45*(1), 33–38. https://doi.org/10.1080/00218839.2006.11101310

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30.

Karapınar, Z., Oğuz, B., Dinçer, E., & Öztürk, C. (2018). Phylogenetic analysis of black queen cell virus and deformed wing virus in honeybee colonies infected by mites in Van,

Eastern Turkey. *MEDYCYNA WETERYNARYJNA-VETERINARY MEDICINE-SCIENCE AND PRACTICE*, *74*(7). https://doi.org/10.21521/mw.5990

Kartal, S., Tunca, R. İ., Özgül, O., Karabağ, K., & Koç, H. (2021). MICROSCOPIC AND MOLECULAR DETECTION OF NOSEMA SP. IN THE SOUTHWEST AEGEAN REGION. *Uludağ Arıcılık Dergisi*, *21*(1), 8–20. https://doi.org/10.31467/uluaricilik.880380

Kaskinova, M. D., Gaifullina, L. R., Saltykova, E. S., Poskryakov, A. V., & Nikolenko, A. G. (2020). Genetic markers for the resistance of honey bee to Varroa destructor. *Vavilovskii Zhurnal Genetiki I Selektsii*, *24*(8), 853–860. https://doi.org/10.18699/VJ20.683

Kassambara, A. (2020). *ggpubr: "ggplot2" Based Publication Ready Plots*. https://CRAN.R-project.org/package=ggpubr

Kekeçoğlu, M., Rasgele, P. G., Acar, F., & Kaya, S. T. (2013). Investigation of Colony Losses, Honey Bee Diseases and Pests and Fighting Methods Seen in the Beekeeping Enterprises in the Province of Duzce. *Journal of the Institute of Science and Technology*, *3*(3), 99–108.

Kesik, H. K., Kutlu, M. A., Kilinç, Ş. G., Gul, A., & Şimşek, S. (2022). Bingöl Yöresi Bal Arılarında Varroasis'in Yaygınlığı. *Türk Tarım ve Doğa Bilimleri Dergisi*, *9*(1), 26–32. https://doi.org/10.30910/turkjans.1024821

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Klee, J., Besana, A. M., Genersch, E., Gisder, S., Nanetti, A., Tam, D. Q., Chinh, T. X., Puerta, F., Ruz, J. M., Kryger, P., Message, D., Hatjina, F., Korpela, S., Fries, I., & Paxton, R. J. (2007). Widespread dispersal of the microsporidian Nosema ceranae, an emergent pathogen of the western honey bee, Apis mellifera. *Journal of Invertebrate Pathology*, *96*(1), 1–10. https://doi.org/10.1016/j.jip.2007.02.014

Klein, A.-M., Vaissière, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., & Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1608), 303–313. https://doi.org/10.1098/rspb.2006.3721

Kramerova, I. A., Kawaguchi, N., Fessler, L. I., Nelson, R. E., Chen, Y., Kramerov, A. A., Kusche-Gullberg, M., Kramer, J. M., Ackley, B. D., Sieron, A. L., Prockop, D. J., & Fessler, J. H. (2000). Papilin in development; a pericellular protein with a homology to the ADAMTS metalloproteinases. *Development (Cambridge, England)*, *127*(24), 5475–5485. https://doi.org/10.1242/dev.127.24.5475

Kramerova, I. A., Kramerov, A. A., & Fessler, J. H. (2003). Alternative splicing of papilin and the diversity of Drosophila extracellular matrix during embryonic morphogenesis. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, *226*(4), 634–642. https://doi.org/10.1002/dvdy.10265

Kulhanek, K., Steinhauer, N., Rennich, K., Caron, D. M., Sagili, R. R., Pettis, J. S., Ellis, J. D., Wilson, M. E., Wilkes, J. T., Tarpy, D. R., Rose, R., Lee, K., Rangel, J., & vanEngelsdorp, D. (2017). A national survey of managed honey bee 2015–2016 annual colony losses in the USA. *Journal of Apicultural Research*, *56*(4), 328–340. https://doi.org/10.1080/00218839.2017.1344496

Le Conte, Y., Alaux, C., Martin, J.-F., Harbo, J. R., Harris, J. W., Dantec, C., Séverac, D., Cros-Arteil, S., & Navajas, M. (2011). Social immunity in honeybees (Apis mellifera): Transcriptome analysis of varroa-hygienic behaviour. *Insect Molecular Biology*, *20*(3), 399–408. https://doi.org/10.1111/j.1365-2583.2011.01074.x

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Locke, B., Forsgren, E., & Miranda, J. R. de. (2014). Increased Tolerance and Resistance to Virus Infections: A Possible Factor in the Survival of Varroa destructor-Resistant Honey Bees (Apis mellifera). *PLOS ONE*, *9*(6), e99998. https://doi.org/10.1371/journal.pone.0099998

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lüdecke, D. (2021). *sjPlot: Data Visualization for Statistics in Social Science* (2.8.10) [Computer software]. https://CRAN.R-project.org/package=sjPlot

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*. https://doi.org/10.14806/ej.17.1.200

Matthijs, S., De Waele, V., Vandenberge, V., Verhoeven, B., Evers, J., Brunain, M., Saegerman, C., De Winter, P. J. J., Roels, S., de Graaf, D. C., & De Regge, N. (2020). Nationwide Screening for Bee Viruses and Parasites in Belgian Honey Bees. *Viruses*, *12*(8), 890. https://doi.org/10.3390/v12080890

Maxfield-Taylor, S. A., Mujic, A. B., & Rao, S. (2015). First Detection of the Larval Chalkbrood Disease Pathogen Ascosphaera apis (Ascomycota: Eurotiomycetes: Ascosphaerales) in Adult Bumble Bees. *PLoS ONE*, *10*(4), e0124868. https://doi.org/10.1371/journal.pone.0124868

McCarthy, S. D., Dugon, M. M., & Power, A. M. (2015). "Degraded" RNA profiles in Arthropoda and beyond. *PeerJ*, *3*, e1436–e1436. PubMed. https://doi.org/10.7717/peerj.1436

McMenamin, A. J., Daughenbaugh, K. F., & Flenniken, M. L. (2020). The Heat Shock Response in the Western Honey Bee (Apis mellifera) is Antiviral. *Viruses*, *12*(2), 245. https://doi.org/10.3390/v12020245

McMenamin, A. J., & Flenniken, M. L. (2018). Recently identified bee viruses and their impact on bee pollinators. *Current Opinion in Insect Science*, *26*, 120–129. https://doi.org/10.1016/j.cois.2018.02.009

Meredith, J. M., Munks, R. J. L., Grail, W., Hurd, H., Eggleston, P., & Lehane, M. J. (2006). A novel association between clustered NF-κB and C/EBP binding sites is required for immune regulation of mosquito Defensin genes. *Insect Molecular Biology*, *15*(4), 393–401. https://doi.org/10.1111/j.1365-2583.2006.00635.x

Mohammadian, B., Bokaie, S., Moharrami, M., Nabian, S., & Forsi, M. (2018). Distribution of Nosema Spp. In climatic regions of Iran. *Veterinary Research Forum*, *9*(3), 259–263. https://doi.org/10.30466/vrf.2018.32082

Mondet, F., Alaux, C., Severac, D., Rohmer, M., Mercer, A. R., & Le Conte, Y. (2015). Antennae hold a key to Varroa-sensitive hygiene behaviour in honey bees. *Scientific Reports*, *5*, 10454. https://doi.org/10.1038/srep10454

Mondet, F., Miranda, J. R. de, Kretzschmar, A., Conte, Y. L., & Mercer, A. R. (2014). On the Front Line: Quantitative Virus Dynamics in Honeybee (Apis mellifera L.) Colonies along a New Expansion Front of the Parasite Varroa destructor. *PLOS Pathogens*, *10*(8), e1004323. https://doi.org/10.1371/journal.ppat.1004323

Mõtus, K., Raie, A., Orro, T., Chauzat, M.-P., & Viltrop, A. (2016). Epidemiology, risk factors and varroa mite control in the Estonian honey bee population. *Journal of Apicultural Research*, *55*(5), 396–412. https://doi.org/10.1080/00218839.2016.1251081

Müller, U. (1997). The Nitric Oxide System in Insects. *Progress in Neurobiology*, *51*(3), 363–381. https://doi.org/10.1016/S0301-0082(96)00067-6

Muz, D., & Muz, M. N. (2009). Survey of the occurrence of Deformed Wing Virus and multiple parasites of queens (Apis mellifera L.) in apiaries with collapsed colonies in Hatay, Turkey. *Journal of Apicultural Research*, *48*(3), 204–208. https://doi.org/10.3896/IBRA.1.48.3.09

Muz, D., & Muz, M. N. (2017). Investigation of Some Pathogens "Colony Loss Syndrome" Resembled Losses Apiaries in Tekirdağ. *Kocatepe Veterinary Journal*, *10*(1), 21–28.

Muz, M. N., Yaman, M., Karakavuk, M., & Solmaz, H. (2012). Kış Salkımı erken bozulan Arı Kolonilerinde Paraziter ve Bakteriyel Patojenler. *Yüzüncü Yıl Üniversitesi Veteriner Fakültesi Dergisi*, *23*(3), 147–150.

Nappi, A. J., Vass, E., Frey, F., & Carton, Y. (2000). Nitric Oxide Involvement in Drosophila Immunity. *Nitric Oxide*, *4*(4), 423–430. https://doi.org/10.1006/niox.2000.0294

Nasri, T., Hedayati, M. T., Abastabar, M., Pasqualotto, A. C., Armaki, M. T., Hoseinnejad, A., & Nabili, M. (2015). PCR-RFLP on β-tubulin gene for rapid identification of the most clinically important species of Aspergillus. *Journal of Microbiological Methods*, *117*, 144–147. https://doi.org/10.1016/j.mimet.2015.08.007

Nazzi, F., Brown, S. P., Annoscia, D., Piccolo, F. D., Prisco, G. D., Varricchio, P., Vedova, G. D., Cattonaro, F., Caprio, E., & Pennacchio, F. (2012). Synergistic Parasite-Pathogen Interactions Mediated by Host Immunity Can Drive the Collapse of Honeybee Colonies. *PLOS Pathogens*, *8*(6), e1002735. https://doi.org/10.1371/journal.ppat.1002735

Nazzi, F., & Le Conte, Y. (2016). Ecology of Varroa destructor, the Major Ectoparasite of the Western Honey Bee, Apis mellifera. *Annual Review of Entomology*, *61*(1), 417–432. https://doi.org/10.1146/annurev-ento-010715-023731

Negri, P., Maggi, M., Correa-Aragunde, N., Brasesco, C., Eguaras, M., & Lamattina, L. (2013). Nitric oxide participates at the first steps of Apis mellifera cellular immune activation in response to non-self recognition. *Apidologie*, *44*(5), 575–585. https://doi.org/10.1007/s13592-013-0207-8

Oğuz, B., Karapinar, Z., Dinçer, E., & Değer, M. S. (2017). Molecular detection of Nosema spp. And black queen-cell virus in honeybees in Van Province, Turkey. *Turkish Journal of Veterinary and Animal Sciences*, *41*(2), 221–227.

Okur Gumusova, S., Albayrak, H., Kurt, M., & Yazici, Z. (2010). Prevalence of three honey bee viruses in Turkey. *Veterinarski Arhiv*, *80*(6), 779–785.

Ongus, J. R., Peters, D., Bonmatin, J.-M., Bengsch, E., Vlak, J. M., & van Oers, M. M. Y. 2004. (2004). Complete sequence of a picorna-like virus of the genus Iflavirus replicating in the mite Varroa destructor. *Journal of General Virology*, *85*(12), 3747–3755. https://doi.org/10.1099/vir.0.80470-0

Önk, K., & Kılıç, Y. (2014). Kars yöresindeki bal arılarında varroosis'in yaygınlığı. *Uludağ Arıcılık Dergisi*, *14(2)*, 69–73.

Oxley, P. R., Spivak, M., & Oldroyd, B. P. (2010). Six quantitative trait loci influence task thresholds for hygienic behaviour in honeybees (Apis mellifera). *Molecular Ecology*, *19*(7), 1452–1461. https://doi.org/10.1111/j.1365-294X.2010.04569.x

Özbilgin, N., Alataş, İ., Balkan, C., Öztürk, A. İ., & Karaca, Ü. (1999). Ege Bölgesi arıcılık faaliyetlerinin teknik ve ekonomik başlıca karakteristiklerinin belirlenmesi. *ANADOLU Ege Tarımsal Araştırma Enstitüsü Dergisi*, *9*(1), 149–170.

Parlakay, O., & Esengün, K. (2005). Tokat İli Merkez İlçede Arıcılık Faaliyetinin Ekonomik Analizi ve İşletmecilik Sorunları. *Gaziosmanpaşa Üniversitesi Ziraat Fakültesi Dergisi*, *22*(1). https://arastirmax.com/tr/publication/gaziosmanpasa-universitesi-ziraat-fakultesi-dergisi/22/1/tokat-ili-merkez-ilcede-aricilik-faaliyetinin-ekonomik-analizi-isletmecilik-sorunlari/arid/b8a93828-a43e-4182

Paxton, R. J., Klee, J., Korpela, S., & Fries, I. (2007). Nosema ceranae has infected Apis mellifera in Europe since at least 1998 and may be more virulent than Nosema apis. *Apidologie*, *38*(6), 558–565. https://doi.org/10.1051/apido:2007037

Potts, S. G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O., & Kunin, W. E. (2010). Global pollinator declines: Trends, impacts and drivers. *Trends in Ecology & Evolution*, *25*(6), 345–353. https://doi.org/10.1016/j.tree.2010.01.007

Potts, S. G., Imperatriz-Fonseca, V., Ngo, H. T., Aizen, M. A., Biesmeijer, J. C., Breeze, T. D., Dicks, L. V., Garibaldi, L. A., Hill, R., Settele, J., & Vanbergen, A. J. (2016). Safeguarding pollinators and their values to human well-being. *Nature*, *540*(7632), 220–229. https://doi.org/10.1038/nature20588

Pritchard, D. J. (2016). Grooming by honey bees as a component of varroa resistant behavior. *Journal of Apicultural Research*, *55*(1), 38–48. https://doi.org/10.1080/00218839.2016.1196016

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. R Foundation for Statistical Computing. https://www.R-project.org/

Ramsey, S. D., Ochoa, R., Bauchan, G., Gulbronson, C., Mowery, J. D., Cohen, A., Lim, D., Joklik, J., Cicero, J. M., Ellis, J. D., Hawthorne, D., & vanEngelsdorp, D. (2019). Varroa destructor feeds primarily on honey bee fat body tissue and not hemolymph. *Proceedings of the National Academy of Sciences*, *116*(5), 1792–1801. https://doi.org/10.1073/pnas.1818371116

Reddy, K. C., Dunbar, T. L., Nargund, A. M., Haynes, C. M., & Troemel, E. R. (2016). The C. elegans CCAAT-Enhancer-Binding Protein Gamma Is Required for Surveillance Immunity. *Cell Reports*, *14*(7), 1581–1589. https://doi.org/10.1016/j.celrep.2016.01.055

Renfro, Z., White, B. E., & Stephens, K. E. (2022). CCAAT enhancer binding protein gamma (C/EBP-γ): An understudied transcription factor. *Advances in Biological Regulation*, *84*, 100861. https://doi.org/10.1016/j.jbior.2022.100861

Rinderer, T. E., Harris, J. W., Hunt, G. J., & Guzman, L. I. de. (2010). Breeding for resistance to Varroa destructor in North America. *Apidologie*, *41*(3), 409–424. https://doi.org/10.1051/apido/2010015

Roetschi, A., Berthoud, H., Kuhn, R., & Imdorf, A. (2008). Infection rate based on quantitative real-time PCR of Melissococcus plutonius, the causal agent of European foulbrood, in honeybee colonies before and after apiary sanitation. *Apidologie*, *39*(3), 362–371. https://doi.org/10.1051/apido:200819

Rørth, P., & Montell, D. J. (1992). Drosophila C/EBP: A tissue-specific DNA-binding protein required for embryonic development. *Genes & Development*, *6*(12A), 2299–2311. https://doi.org/10.1101/gad.6.12a.2299

Rosenkranz, P., Aumeier, P., & Ziegelmann, B. (2010). Biology and control of Varroa destructor. *Journal of Invertebrate Pathology*, *103*, S96–S119. https://doi.org/10.1016/j.jip.2009.07.016

Rossi, F., Amadoro, C., Ruberto, A., & Ricchiuti, L. (2018). Evaluation of Quantitative PCR (qPCR) Paenibacillus larvae Targeted Assays and Definition of Optimal Conditions for Its Detection/Quantification in Honey and Hive Debris. *Insects*, *9*(4), 165. PubMed. https://doi.org/10.3390/insects9040165

Runckel, C., Flenniken, M. L., Engel, J. C., Ruby, J. G., Ganem, D., Andino, R., & DeRisi, J. L. (2011). Temporal Analysis of the Honey Bee Microbiome Reveals Four Novel Viruses and Seasonal Prevalence of Known Viruses, Nosema, and Crithidia. *PLoS ONE*, *6*(6), e20656. https://doi.org/10.1371/journal.pone.0020656

Rüstemoğlu, M., & Sipahioğlu, H. M. (2019). Occurrence and prevalence of six honey bee viruses in Hakkari (Turkey) and their genomic divergence. *Munis Entomology & Zoology*, *14*(2), 574–583.

Sahinler, N., & Gul, A. (2005). Investigation of bee diseases in beekeeping enterprice in Hatay province. *Uludag Bee Journal (Turkey)*.

Santillán-Galicia, M. T., Ball, B. V., Clark, S. J., & Alderson, P. G. (2014). Slow bee paralysis virus and its transmission in honey bee pupae by Varroa destructor. *Journal of Apicultural Research*, *53*(1), 146–154. https://doi.org/10.3896/IBRA.1.53.1.16

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Seitz, N., Traynor, K. S., Steinhauer, N., Rennich, K., Wilson, M. E., Ellis, J. D., Rose, R., Tarpy, D. R., Sagili, R. R., Caron, D. M., Delaplane, K. S., Rangel, J., Lee, K., Baylis, K., Wilkes, J. T., Skinner, J. A., Pettis, J. S., & vanEngelsdorp, D. (2015). A national survey of managed honey bee 2014–2015 annual colony losses in the USA. *Journal of Apicultural Research*, *54*(4), 292–304. https://doi.org/10.1080/00218839.2016.1153294

Seven Çakmak, S., & Çakmak, İ. (2016). Beekeeping and Recent Colony Losses in Turkey. *Uludağ Arıcılık Dergisi*, *16*(1), 31–48.

Sevim, A., Akpınar, R., Karaoğlu, Ş. A., Bozdeveci, A., & Sevim, E. (2022). Prevalence and phylogenetic analysis of Ascosphaera apis (Maassen ex Claussen) LS Olive & Spiltoir (1955) isolates from honeybee colonies in Turkey. *Biologia*. https://doi.org/10.1007/s11756-022-01114-7

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Shen, M., Cui, L., Ostiguy, N., & Cox-Foster, D. 2005. (2005). Intricate transmission routes and interactions between picorna-like viruses (Kashmir bee virus and sacbrood virus) with the honeybee host and the parasitic varroa mite. *Journal of General Virology*, *86*(8), 2281–2289. https://doi.org/10.1099/vir.0.80824-0

Shoreit, M. N., & Bagy, M. M. K. (1995). Mycoflora associated with stonebrood disease in honeybee colonies in Egypt. *Microbiological Research*, *150*(2), 207–211. https://doi.org/10.1016/S0944-5013(11)80058-3

Shumkova, R., Georgieva, A., Radoslavov, G., Sirakova, D., Dzhebir, G., Neov, B., Bouga, M., & Hristov, P. (2018). The first report of the prevalence of Nosema ceranae in Bulgaria. *PeerJ*, *6*, e4252. https://doi.org/10.7717/peerj.4252

Şimşek, H. (2005). Elazığ yöresi bal arılarında bazı parazit ve mantar hastalıklarının araştırılması. *Ankara Üniversitesi Veteriner Fakültesi Dergisi*, *52*(2), 123–126.

Sıralı, R., & Doğaroğlu, M. (2005). Survey Results on Honeybee Pests and Diseases in Thracian Region of Turkey. *Uludag Bee Journal*. https://dergipark.org.tr/en/download/article-file/143549

Slowikowski, K. (2021). *ggrepel: Automatically Position Non-Overlapping Text Labels with "ggplot2."* https://CRAN.R-project.org/package=ggrepel

Spiltoir, C. F. (1955). Life Cycle of Ascosphaera Apis (pericystis Apis). *American Journal of Botany*, *42*(6), 501–508. https://doi.org/10.1002/j.1537-2197.1955.tb11154.x

Spiltoir, C. F., & Olive, L. S. (1955). A Reclassification of the Genus Pericystis Betts. *Mycologia*, *47*(2), 238–244. https://doi.org/10.1080/00275514.1955.12024448

Spivak, M., Masterman, R., Ross, R., & Mesce, K. A. (2003). Hygienic behavior in the honey bee (Apis mellifera L.) and the modulatory role of octopamine. *Journal of Neurobiology*, *55*(3), 341–354. https://doi.org/10.1002/neu.10219

Strohm, E., Herzner, G., Ruther, J., Kaltenpoth, M., & Engl, T. (2019). Nitric oxide radicals are emitted by wasp eggs to kill mold fungi. *ELife*, *8*, e43718. https://doi.org/10.7554/eLife.43718

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2021). The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, *49*(D1), D605–D612. https://doi.org/10.1093/nar/gkaa1074

The Gene Ontology Consortium. (2021). The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Research*, *49*(D1), D325–D334. https://doi.org/10.1093/nar/gkaa1113

Thompson, G. J., Yockey, H., Lim, J., & Oldroyd, B. P. (2007). Experimental manipulation of ovary activation and gene expression in honey bee (Apis mellifera) queens and workers: Testing hypotheses of reproductive regulation. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, *307A*(10), 600–610. https://doi.org/10.1002/jez.415

Traynor, K. S., Rennich, K., Forsgren, E., Rose, R., Pettis, J., Kunkel, G., Madella, S., Evans, J., Lopez, D., & vanEngelsdorp, D. (2016). Multiyear survey targeting disease incidence in US honey bees. *Apidologie*, *47*(3), 325–347. https://doi.org/10.1007/s13592-016-0431-0

Tunca, R. İ., & Çimrin, T. (2012). Kırşehir İlinde Bal Arısı Yetiştiricilik Aktiviteleri Üzerine Anket Çalışması. *Journal of the Institute of Science and Technology*, *2*(2), 99–108.

Uhler, J., Garbern, J., Yang, L., Kamholz, J., & Mellerick, D. M. (2002). Nk6, a novel Drosophila homeobox gene regulated by vnd. *Mechanisms of Development*, *116*(1), 105–116. https://doi.org/10.1016/S0925-4773(02)00155-7

Ütük, A. E., Pişkin, F. Ç., Deniz, A., & Balkaya, İ. (2011). Varroosis ve nosemosis üzerine retrospektif bir çalışma. *Etlik Veteriner Mikrobiyoloji Dergisi*, *22*(1), 11–15.

Vanbergen, A. J., & the Insect Pollinators Initiative. (2013). Threats to an ecosystem service: Pressures on pollinators. *Frontiers in Ecology and the Environment*, *11*(5), 251–259. https://doi.org/10.1890/120126

Varga, J., & Samson, R. A. (2008). *Aspergillus in the genomic era*. Wageningen Academic Publishers.

Venables, B., & Ripley, B. (2002). *Modern Applied Statistics with S* (Fourth Edition). New York. Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Villa, J. D., Danka, R. G., & Harris, J. W. (2017). Repeatability of measurements of removal of mite-infested brood to assess Varroa Sensitive Hygiene. *Journal of Apicultural Research*, *56*(5), 631–634. https://doi.org/10.1080/00218839.2017.1369707

Wagoner, K., Spivak, M., Hefetz, A., Reams, T., & Rueppell, O. (2019). Stock-specific chemical brood signals are induced by Varroa and Deformed Wing Virus, and elicit hygienic response in the honey bee. *Scientific Reports*, *9*(1), 8753. https://doi.org/10.1038/s41598-019-45008-2

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wilfert, L., Long, G., Leggett, H. C., Schmid-Hempel, P., Butlin, R., Martin, S. J. M., & Boots, M. (2016). Deformed wing virus is a recent global epidemic in honeybees driven by Varroa mites. *Science*, *351*(6273), 594–597. https://doi.org/10.1126/science.aac9976

Wilson, M. J., & Dearden, P. K. (2008). Evolution of the insect Sox genes. *BMC Evolutionary Biology*, *8*, 120. https://doi.org/10.1186/1471-2148-8-120

Winnebeck, E. C., Millar, C. D., & Warman, G. R. (2010). Why does insect RNA look degraded? *Journal of Insect Science (Online)*, *10*, 159–159. PubMed. https://doi.org/10.1673/031.010.14119

Wu, J.-L., Zhou, C.-X., Wu, P.-J., Xu, J., Guo, Y.-Q., Xue, F., Getachew, A., & Xu, S.-F. (2017). Brain metabolomic profiling of eastern honey bee (Apis cerana) infested with the mite Varroa destructor. *PLoS ONE*, *12*(4), e0175573. https://doi.org/10.1371/journal.pone.0175573

Xiao, N. (2018). *ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for "ggplot2"* (2.9) [R]. https://github.com/nanxstats/ggsci (Original work published 2016)

Yalçınkaya, A., & Keskin, N. (2010). The investigation of honey bee diseases after colony losses in Hatay and Adana provinces of Turkey. *Mellifera*. https://www.cabi.org/ISC/abstract/20113131435

Yan, L. (2021). *ggvenn: Draw Venn Diagram by "ggplot2"* (0.1.9) [Computer software]. https://CRAN.R-project.org/package=ggvenn

Yaşar, N., Güler, A., Yeşiltaş, H. B., Bülüt, G., & Gökçe, M. (2002). Overall Structure of Beekeeping in the Black Sea Region of Turkey. *Mellifera*, *2*(3), 47.

Yilmaz, F., Öztürk, S. hasan, Kuvanci, A., Kayaboynu, Ü., Karataş, Ü., Kaya, S., Derebaşi, E., & Buldağ, M. (2018). Doğu Karadeniz Bölgesinde Nosema apis ve Nosema ceranae'nın Epidemiyolojisi. *Arıcılık Araştırma Dergisi*, *10*(2), 34–44.

Yue, C., & Genersch, E. 2005. (2005). RT-PCR analysis of Deformed wing virus in honeybees (Apis mellifera) and mites (Varroa destructor). *Journal of General Virology*, *86*(12), 3419–3424. https://doi.org/10.1099/vir.0.81401-0

APPENDIX A

Table S1. Number of reads generated in millions, percentages of bases with a quality score of 20 or higher (Q20) and 30 or higher (Q30), GC content and percentages of rRNA reads for all 115 samples.

Sample	group	Raw reads (million)	Q20	Q30	GC	rRNA rate (%)
A_003	VR	68.31	96.08	90.61	29.62	14.05
A_005	VR	49.94	95.44	89.13	35.05	2.95
A_006	VR	92.59	96.53	91.53	32.74	5.72
A_007	VR	60.41	96.76	91.71	36.29	1.7
A_009	VR	65.41	95.55	89.42	32.3	5.56
A_015	VR	50.94	96.33	91.04	35.82	3.51
A_017	VR	65.61	95.79	89.81	35.04	3.17
A_018	VR	65.10	96.76	91.65	36.8	1.85
A_031	VR	104.03	96.34	91.19	31.51	8.15
A_035	VR	62.34	97	92.17	37.04	2.01
A_092	VR	73.32	96.86	91.89	36.74	1.63
A_119	VR	62.69	96.26	90.85	35.48	3.23
A_189	VR	60.68	96.53	91.27	36.86	1.98
A_226	VR	88.10	96.67	91.66	35.08	6.53
B_009	VR	68.90	95.21	89.13	36.3	1.93
B_010	VR	57.50	95.73	90.31	33.21	6.16
B_012	VR	96.10	96.4	91.17	35.17	3.83
B_020	VR	64.71	95.43	90.02	37.32	2.72
B_021	VR	79.75	96.57	91.65	35.07	3.4
B_022	VR	65.49	95.5	89.46	34.09	4.94
B_027	VR	67.60	96.3	91.14	33.14	5.58
B_028	VR	43.40	94.95	88.41	31.81	7.49
B_029	VR	56.54	96.78	91.96	34.48	3.76
B_031	VR	63.20	95.67	89.62	31.73	6.78
B_032	VR	61.15	96.65	91.41	33.03	4.03
B_033	VR	62.34	97.09	92.3	36.28	1.8
B_034	VR	46.11	96.08	90.35	33.68	5.1
B_035	VR	57.71	96.87	91.68	35.77	2.45
B_036	VR	60.37	96.75	91.86	34.55	3.11
B_037	VR	55.06	96.44	91.05	31.1	8.46
B_038	VR	60.58	96.73	91.47	35.37	2.23
H_001	VR	79.45	95.75	90.05	32.54	5.79
H_002	VS	66.30	96.9	91.88	37.25	2.08
H_003	VS	62.76	96.16	90.34	27.95	18.44
H_004	VS	116.79	96.44	91.41	35	4.07
H_006	VS	96.54	95.46	89.78	32.23	8.12
H_010	VS	75.47	95.81	89.9	33.78	7.83
H_012	VS	61.37	96.5	91.21	37.45	1.73
H_013	VS	90.52	96.46	91.41	32.02	8.39
H_014	VS	84.16	96.9	91.85	36.1	4.67

H_016	VS	77.56	96.01	90.73	33.91	4.83
H_019	VS	63.92	95.03	88.47	36.59	11.55
H_020	VS	65.82	96.67	91.53	36.41	2.27
H_021	VS	66.78	96.84	91.85	37.79	2.56
H_022	VS	67.69	95.83	90.4	31.15	9.43
I_009	VS	82.07	96.19	90.94	33.05	13.38
I_013	VS	92.14	96.28	91.11	34.36	11.39
I_016	VS	58.89	97.1	92.41	33.17	12.09
I_020	VS	64.83	96.44	91.2	27.86	16.49
I_025	VS	82.77	95.92	90.56	30.08	13.43
I_027	VS	94.36	96.06	90.73	33.07	5.65
I_028	VS	59.63	95.04	88.39	34.57	4.95
I_034	VS	57.53	96.42	91	32.38	9.04
I_035	VS	50.12	95.12	88.43	33.55	7.01
I_046	VS	61.36	96.24	90.52	28.37	17.81
I_047	VS	62.16	96.12	90.35	29.08	16.02
I_048	VS	65.31	96.14	90.45	26.67	19.21
N_001	VS	47.37	97.35	92.65	36.95	3.37
N_002	VS	70.71	96.71	91.67	36.09	3.17
N_004	VS	63.65	96.13	90.47	27.31	16.67
N_010	VS	66.29	97.12	92.16	33.92	8.54
N_015	VS	68.34	96.14	90.58	29.42	14.65
N_016	VS	73.34	96	90.54	32.38	7.36
N_017	VS	70.46	96.34	91.2	33.72	5.12
N_021	VS	76.13	96.26	90.93	30.29	11.68
N_022	VS	62.88	97.17	92.29	37.16	2.62
N_026	VS	58.02	94.97	88.2	28.7	14.23
N_027	VS	88.20	96.77	91.8	36.26	10.67
N_028	VS	62.66	97.14	92.16	37.81	1.76
N_029	VS	75.74	97.77	93.36	37.63	3.21
N_034	VS	62.02	96.73	91.64	32.63	8.63
N_035	VS	69.89	95.74	89.8	34.98	3.23
N_037	VS	46.84	96.61	91.29	30.28	15.18
N_041	VS	30.00	94.97	88.3	32.41	9.06
N_043	VS	56.65	96.22	90.51	29.91	16.78
N_046	VS	86.47	96.65	91.78	36.48	4.69
N_047	VS	66.87	96.19	90.7	32.74	8.91
N_048	VS	53.63	96.7	91.26	36.82	3.13
R_002	VS	77.84	96.72	92.03	35.96	4.38
R_004	VS	61.26	96.28	90.22	25.47	23.48
R_005	VR	58.41	97.47	92.88	39.78	2.56
R_006	VS	94.25	96.17	91.02	34.54	5.21
R_007	VS	49.56	95.05	88.55	34.26	6.42
R_008	VS	84.65	96.97	92.24	38.95	2.65
R_009	VS	72.34	96.33	90.89	32.16	10.77
R_010	VS	93.80	96.55	91.69	35.1	5.36
R_014	VS	93.12	95.77	90.39	34.48	6.49
R_015	VS	62.78	97.49	92.91	38.65	0.88

R_016	VS	65.81	97.33	92.53	38.68	0.94
R_020	VS	61.62	96.71	91.57	36.29	3.62
R_022	VS	50.77	96.03	90.12	36.6	5.12
R_023	VS	68.95	98.21	94.44	35.83	5.15
R_024	VR	45.89	96.44	90.8	36.08	5.83
R_026	VS	58.07	97.57	92.93	39.23	1.01
R_029	VS	60.45	97.44	92.85	38.61	0.96
R_035	VS	64.28	95.35	89.72	31.27	11.91
R_037	VS	55.88	97.32	92.3	31.91	11.95
R_039	VR	83.99	96.01	90.64	33.63	5.26
R_082	VS	77.02	97.74	93.39	32	12.44
R_090	VR	53.06	96.58	90.77	30.24	15.6
R_093	VR	65.85	97.53	93.18	32.02	11.89
R_119	VR	62.76	97.39	92.5	30.97	13.32
R_133	VR	65.92	95.5	89.45	30.72	13.31
R_229	VR	43.85	95.84	89.59	34.65	6.65
R_245	VR	68.82	95.87	90.12	27.22	19.89
S_024	VR	41.48	95.3	88.95	31.66	8.93
S_025	VS	42.69	95.85	89.61	25.26	22.4
S_039	VS	3.10	95.75	90	33.39	3.99
S_106	VS	61.24	94.48	87.66	31.11	9.5
S_129	VS	19.42	96.66	91.44	34.33	5.59
S_170	VS	67.22	95.83	90.19	33.25	6.93
S_180	VS	88.11	96.34	91.26	34.63	4.3
S_200	VS	88.18	96.71	91.69	35.94	4.49
S_246	VS	106.22	97.43	92.56	32.23	12.06

Table S2. Common significantly upregulated genes and their shrunken log2 fold changes in VR and VRL compared to VS and VSH colonies that are SB-positive (SB+) and SB-negative (SB-). Some genes are uncharacterized.

Gene	Description	VR_SB-	VR_SB+	VRL_SB-	VR_SB+
18-w	18-wheeler	7.72	6.91	7.02	7.16
Adk1	adenylate kinase 1	2.69	2.79	2.20	3.09
AGLU 2	alpha glucosidase 2	1.97	1.88	2.54	2.12
Apd-2	apidermin 2	1.49	1.50	1.78	1.44
Cpap3-b	cuticular protein analogous to peritrophins 3-B	1.99	2.01	2.25	1.81
CUTA	cutA divalent cation tolerance homolog	2.60	2.00	1.90	2.61
Fstl5	follistatin-like 5	3.41	3.61	4.33	3.96
Gfat2	glucosamine-fructose-6-phosphate aminotransferase 2	1.57	1.68	2.28	1.63
Hbg2	alpha-glucosidase	1.62	1.62	1.87	1.83
Hsc70-4	heat shock protein cognate 4	1.08	1.44	1.22	1.52

LOC100576085	ras-related protein Rab-8A	1.41	1.80	1.66	1.68
LOC100577143	---	1.29	1.24	1.52	1.49
LOC100577156	chitinase-3-like protein 1	1.24	1.38	1.56	1.38
LOC100577198	---	1.37	1.38	1.50	1.56
LOC100577504	---	1.27	1.35	1.37	1.45
LOC100577527	---	1.46	1.66	1.97	1.59
LOC100577539	---	2.94	3.12	2.10	3.74
LOC100577587	cytospin-A	1.89	1.60	2.52	2.05
LOC100577764	ras-related protein Rab-5C	1.36	1.49	1.53	1.40
LOC100577901	---	2.72	2.47	2.59	2.09
LOC100578035	sphingolipid delta(4)-desaturase DES1	2.34	1.92	2.78	2.09
LOC100578100	uncharacterized LOC100578100	1.74	1.77	2.38	1.95
LOC100578144	mitotic apparatus protein p62	2.23	2.22	2.97	2.03
LOC100578253	Golgi integral membrane protein 4	1.71	1.28	2.00	1.56
LOC100578341	---	1.18	1.22	1.38	1.52
LOC100578352	ionotropic receptor 75a-like	5.06	5.63	1.02	7.71

LOC100578446	pituitary homeobox x	4.23	2.94	4.46	4.06
LOC100578483	probable DNA primase large subunit	1.80	1.88	2.08	2.15
LOC100578579	---	1.69	1.45	1.76	1.58
LOC100578953	dual specificity protein phosphatase 15	2.46	2.39	1.67	2.61
LOC100265416	transmembrane protein 132C	1.73	1.95	1.88	2.40
LOC100265440	---	1.37	1.61	1.76	1.56
LOC100265442	sodium-dependent nutrient amino acid transporter 1	1.45	1.18	1.72	1.20
LOC100265495	lipase 3	3.14	2.22	2.94	3.67
LOC100265495	ATP synthase subunit b, mitochondrial	1.38	1.38	1.45	1.36
LOC100265540	ATP-dependent RNA helicase p62-like	1.43	1.33	1.69	1.54
LOC100265559	---	1.48	1.79	1.72	1.82
LOC100265562	facilitated trehalose transporter Tret1	1.44	1.03	1.62	1.17
LOC100265636	---	3.31	2.69	3.25	3.16
LOC100796398	protein tweety	1.55	1.85	1.87	2.01
LOC100796483	rootletin	3.08	2.46	2.75	2.54
LOC100796496	---	1.73	1.46	2.13	1.62

LOC10796521	feline leukemia virus subgroup C receptor-related protein 2	1.95	1.79	2.57	1.79
LOC10796579	lactosylceramide 4-alpha-galactosyltransferase	1.63	1.74	1.86	1.84
LOC11321853	galectin-8-like	2.72	2.47	2.84	3.07
LOC11321855	solute carrier family 52, riboflavin transporter, member 3-A-like	2.29	1.88	2.80	2.57
LOC11321875	proteasome subunit alpha type-6-like	2.22	2.55	1.52	3.22
LOC11321889	---	5.75	6.85	6.36	7.12
LOC11321902	pancreatic triacylglycerol lipase-like	1.10	1.54	1.31	1.50
LOC11321926	trypsin-1-like	1.37	1.64	1.53	1.67
LOC11321941	---	3.08	3.22	3.46	3.31
LOC408272	monocarboxylate transporter 7	1.25	1.37	1.53	1.51
LOC408275	---	1.28	1.46	1.87	1.72
LOC408301	type 1 phosphatidylinositol 4,5-bisphosphate 4-phosphatase	1.58	1.55	1.66	1.60
LOC408302	solute carrier family 22 member 21	3.37	2.02	3.46	2.27
LOC408329	---	1.38	1.65	1.60	1.65
LOC408335	nuclear pore complex protein Nup98-Nup96	1.01	1.11	1.38	1.27
LOC408452	cytochrome P450 9e2	1.99	1.47	2.48	1.74
LOC408470	NAD kinase	1.08	1.17	1.38	1.38
LOC408474	apyrase	1.17	1.42	1.30	1.47
LOC408531	PRADC1-like protein	1.64	1.76	1.56	1.71
LOC408533	mitogen-activated protein kinase kinase kinase 15	4.23	3.51	3.22	3.61

LOC408534	trypsin	1.44	1.48	1.93	1.66
LOC408551	collagen alpha-5(IV) chain	1.72	1.91	1.85	1.75
LOC408552	collagen alpha-1(IV) chain	1.51	1.51	1.53	1.45
LOC408564	long-chain fatty acid transport protein 4	1.58	1.84	1.54	2.09
LOC408572	myophilin	1.10	1.22	1.30	1.39
LOC408592	---	1.60	1.61	1.81	1.72
LOC408594	triple functional domain protein	1.08	1.41	1.46	1.60
LOC408603	glucose dehydrogenase [FAD, quinone]	1.39	1.36	1.84	1.40
LOC408635	BTB/POZ domain-containing protein 1	2.22	1.55	2.48	2.07
LOC408656	RNA exonuclease 1 homolog	1.30	1.48	1.57	1.87
LOC408742	fatty-acid amide hydrolase 2-B	1.34	1.44	1.83	1.57
LOC408782	tubulin beta-1	1.38	1.05	1.25	1.11
LOC408790	protein croquemort	1.32	1.51	1.80	1.64
LOC408841	voltage-dependent calcium channel subunit alpha-2/delta-3	2.21	1.79	2.68	1.98
LOC408850	pre-mRNA-processing factor 39	1.45	1.71	1.73	1.53
LOC408852	protein similar	2.03	2.08	2.48	2.23
LOC408872	COUP transcription factor 2	1.02	1.24	1.24	1.56
LOC408878	protein mesh	1.55	1.56	1.98	1.91
LOC408937	SPARC	1.33	1.56	1.34	1.50
LOC408950	malate dehydrogenase, mitochondrial	1.32	1.22	1.56	1.20
LOC408957	---	1.05	1.09	1.04	1.24
LOC408961	apolipophorins	1.04	1.11	1.07	1.23
LOC408981	activating transcription factor 3	1.41	1.35	1.67	1.31
LOC408983	protein transport protein Sec23A	1.39	1.87	1.83	2.02

LOC409068	annexin B11	2.14	2.11	1.69	2.13
LOC409126	ras-related protein Rab-2	1.62	1.72	1.78	2.13
LOC409141	sugar transporter SWEET1	1.01	1.04	1.67	1.23
LOC409155	dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	1.74	1.36	2.08	1.47
LOC409180	4-hydroxybutyrate coenzyme A transferase-like	1.15	1.37	1.74	1.38
LOC409183	sialin	1.62	1.78	1.68	1.88
LOC409243	A-kinase anchor protein 90	1.47	1.76	2.20	1.89
LOC409256	armadillo repeat-containing protein 5	3.29	3.91	3.40	4.02
LOC409270	glycerol-3-phosphate acyltransferase 4	1.78	1.58	2.11	1.67
LOC409281	protein sel-1 homolog 1	1.09	1.36	1.31	1.45
LOC409291	protein fem-1 homolog CG6966	1.35	1.55	1.93	1.76
LOC409366	integral membrane protein 2C	2.22	1.85	2.41	1.83
LOC409468	venom metalloproteinase 32	1.12	1.01	1.18	1.32
LOC409563	---	1.57	1.25	1.77	1.55
LOC409564	ubiquitin conjugation factor E4 A	1.65	1.91	1.83	1.83
LOC409576	spastin	5.05	5.87	1.12	7.14
LOC409590	cyclin-dependent kinase 1	7.80	4.33	7.41	5.54
LOC409613	GTP-binding protein SAR1	1.12	1.27	1.57	1.30
LOC409619	putative aminopeptidase-2	1.21	1.27	1.21	1.35
LOC409626	chymotrypsin-2	1.78	2.01	1.82	2.23
LOC409638	elongation of very long chain fatty acids protein AAEL008004	1.14	1.04	1.42	1.27
LOC409650	solute carrier organic anion transporter family member 2A1	1.42	1.42	1.75	1.60

LOC409709	putative glucosylceramidase 4	1.81	1.33	2.20	1.50
LOC409721	uncharacterized protein YER152C	1.22	1.15	1.25	1.33
LOC409722	basement membrane-specific heparan sulfate proteoglycan core protein2	1.63	1.52	1.71	1.66
LOC409810	anaphase-promoting complex subunit 4	3.12	2.36	3.21	2.42
LOC409814	beta-glucuronidase	1.34	1.66	1.62	1.73
LOC409865	protein retinal degeneration B	1.27	1.27	1.74	1.56
LOC409899	putative sodium-dependent multivitamin transporter	2.36	2.63	2.35	2.82
LOC409919	excitatory amino acid transporter 1	1.58	1.59	1.77	2.06
LOC409924	synaptic vesicle glycoprotein 2B	1.57	1.39	1.75	1.46
LOC409932	b(0,+)-type amino acid transporter 1	1.60	1.89	2.11	1.83
LOC409934	vesicle-associated membrane protein-associated protein B	1.34	1.52	1.49	1.41
LOC410022	mitochondrial-processing peptidase subunit beta	1.27	1.18	1.50	1.28
LOC410044	rho GTPase-activating protein 20	3.68	3.07	4.23	3.89
LOC410096	glutathione hydrolase 1 proenzyme	1.29	1.28	1.74	1.43
LOC410229	toll-like receptor 6	4.61	4.05	3.62	4.90
LOC410241	ras-related protein Rab-10	1.53	1.92	1.86	1.74
LOC410271	BMP-binding endothelial regulator protein	1.62	1.60	1.55	1.54
LOC410451	venom serine carboxypeptidase	1.44	1.67	1.79	1.47
LOC410484	trehalase	1.45	1.50	1.56	1.73
LOC410489	myosin-IIIb	3.11	2.45	3.04	2.76
LOC410539	protein 5NUC	1.17	1.20	1.51	1.21
LOC410589	semaphorin-5A	1.03	1.45	1.27	1.44
LOC410621	multiple epidermal growth factor-like domains protein 10	1.56	1.54	1.42	1.58

LOC410624	serine proteinase stubble	7.24	7.99	5.30	9.06
LOC410729	putative serine protease K12H4.7	1.44	1.61	1.39	1.76
LOC410793	---	1.67	1.58	2.47	1.90
LOC410806	histidine--tRNA ligase, cytoplasmic	1.20	1.28	1.52	1.56
LOC410829	---	1.32	1.50	1.41	1.72
LOC410894	chymotrypsin-1	1.03	1.13	1.47	1.24
LOC410982	protein croquemort	1.22	1.30	1.40	1.41
LOC411252	GTP-binding protein 2	2.53	2.34	1.81	2.69
LOC411387	O-acyltransferase like protein	4.23	4.10	4.18	3.07
LOC411631	liprin-beta-1	2.15	1.74	2.45	1.94
LOC411647	calumenin	1.93	1.83	1.98	1.86
LOC411813	NADP-dependent malic enzyme	1.17	1.19	1.30	1.11
LOC411846	neprilysin-4	2.13	2.22	1.91	2.17
LOC411923	alanine--tRNA ligase, cytoplasmic	1.69	2.26	2.13	2.42
LOC411983	putative fatty acyl-CoA reductase	2.07	1.63	2.81	1.87
LOC412077	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1	1.98	1.98	1.96	2.13
LOC412118	dnaJ homolog subfamily C member 16	4.19	2.00	2.74	1.72
LOC412150	endoplasmin	2.16	1.99	2.33	1.89
LOC412154	probable cytosolic oligopeptidase A	1.16	1.94	1.65	2.00
LOC412236	E3 ubiquitin-protein ligase LRSAM1	1.71	1.74	2.02	1.94
LOC412354	---	2.58	2.22	2.94	2.21
LOC412355	diacylglycerol kinase eta	6.71	5.79	6.87	4.69
LOC412394	alpha-L-fucosidase	1.51	2.14	1.41	1.92
LOC412430	major facilitator superfamily domain-containing protein 6	1.24	1.13	1.73	1.15

LOC412472	protein lethal(2)denticleless	5.09	3.88	5.19	4.61
LOC412505	translocation protein SEC63 homolog	1.39	1.49	1.49	1.64
LOC412643	---	3.99	2.72	4.30	3.52
LOC412663	laminin subunit alpha	2.24	1.86	2.19	2.14
LOC412746	calcineurin-binding protein cabin-1	1.84	2.20	2.02	2.23
LOC412787	protein TAPT1 homolog	2.88	2.94	3.39	3.17
LOC412799	moesin/ezrin/radixin homolog 1	1.58	1.68	1.91	1.90
LOC412815	fatty acid synthase	1.46	1.44	2.24	1.68
LOC412843	phosphoenolpyruvate carboxykinase	1.75	1.49	2.44	1.77
LOC413021	papilin	1.88	1.51	1.40	1.41
LOC413097	dynein heavy chain 12, axonemal	1.21	1.34	1.31	1.45
LOC413117	proton-coupled amino acid transporter-like protein pathetic	1.65	1.51	1.77	1.57
LOC413128	putative ferric-chelate reductase 1 homolog	2.52	2.28	2.63	2.01
LOC413132	---	2.67	2.20	2.78	2.45
LOC413138	probable phospholipid-transporting ATPase IIA	1.78	1.75	2.09	1.75
LOC413208	rab11 family-interacting protein 2	2.36	1.77	2.72	1.93
LOC413303	vascular endothelial growth factor receptor 1	3.40	2.64	2.95	3.25
LOC413332	spectrin beta chain, non-erythrocytic 5	1.77	1.57	2.14	1.86
LOC413363	F-box only protein 42	2.74	2.60	1.91	2.31
LOC413376	E3 ubiquitin-protein ligase TRIM37	2.03	1.83	2.01	2.03
LOC413551	mitochondrial coenzyme A transporter SLC25A42	2.77	2.23	4.05	2.63
LOC413592	probable ATP-dependent RNA helicase DHX34	3.07	3.51	2.94	4.13
LOC413702	hormone-sensitive lipase	2.79	3.58	2.97	4.08
LOC413755	attractin	1.65	2.02	1.95	2.10

LOC413880	peroxisomal targeting signal 1 receptor	1.44	1.79	2.04	1.90
LOC413968	dystrophin, isoforms A/C/F/G/H	1.16	1.35	1.28	1.48
LOC413984	serine-rich adhesin for platelets	2.85	1.97	2.25	2.55
LOC414015	abl interactor 2	2.67	2.14	2.51	2.79
LOC550929	serine/threonine-protein kinase PINK1, mitochondrial	1.42	1.55	2.05	1.63
LOC551061	multidrug resistance-associated protein 4	2.04	1.74	2.88	2.12
LOC551087	periodic tryptophan protein 2 homolog	1.43	2.06	2.32	2.00
LOC551180	aminopeptidase N	1.38	1.43	1.59	1.50
LOC551206	---	2.44	2.83	2.01	3.90
LOC551437	alpha-N-acetylglucosaminidase	1.52	1.95	1.65	2.06
LOC551524	zinc carboxypeptidase	1.88	2.17	2.13	2.35
LOC551563	cullin-5	4.83	5.62	5.19	7.38
LOC551576	exocyst complex component 7	1.39	1.76	1.33	1.86
LOC551581	N-acetylgalactosamine kinase	1.80	1.82	2.22	1.90
LOC551582	maternal protein exuperantia	2.00	2.09	1.87	2.45
LOC551714	cGMP-dependent protein kinase, isozyme 1	1.33	1.78	1.70	1.59
LOC551775	trifunctional enzyme subunit beta, mitochondrial	1.24	1.25	1.25	1.28
LOC551793	---	1.54	1.17	1.92	1.43
LOC551824	putative pre-mRNA-splicing factor ATP-dependent RNA helicase PRP1	1.45	1.44	1.57	1.75
LOC551837	long-chain-fatty-acid--CoA ligase ACSBG2	1.92	2.05	1.98	2.00
LOC551859	calmodulin	1.13	1.19	1.24	1.34
LOC551861	ATP synthase subunit e, mitochondrial	1.23	1.37	1.15	1.38
LOC551913	protein CREG1	1.55	1.28	1.74	1.42
LOC551958	---	1.79	2.10	1.91	2.35

LOC552145	repressor of RNA polymerase III transcription MAF1 homolog	2.78	2.88	2.19	3.56
LOC552163	dolichyl pyrophosphate Man9GlcNAc2 alpha-1,3-glucosyltransferase	1.60	1.41	1.60	1.42
LOC552257	bromodomain-containing protein DDB_G0270170	1.34	1.44	1.25	1.53
LOC552283	---	1.40	1.37	1.79	1.47
LOC552327	UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase	3.26	3.52	3.57	3.59
LOC552336	UBX domain-containing protein 1	1.29	1.28	1.31	1.11
LOC552375	NPC intracellular cholesterol transporter 1	1.11	1.10	1.36	1.27
LOC552392	serine palmitoyltransferase 1	1.29	1.66	1.66	1.49
LOC552549	ribosomal protein S6 kinase beta-1	6.34	5.94	5.37	4.96
LOC552647	E3 ubiquitin-protein ligase Bre1	2.36	2.06	2.92	2.25
LOC552708	major facilitator superfamily domain-containing protein 10	1.65	2.03	2.55	1.85
LOC552712	6-phosphogluconate dehydrogenase, decarboxylating	1.01	1.06	1.29	1.20
LOC552784	---	1.48	1.57	2.04	1.80
LOC552795	cholinephosphotransferase 1	2.11	2.30	3.35	2.27
LOC724160	---	1.90	1.65	2.38	1.93
LOC724199	early nodulin-75	1.27	1.33	1.64	1.42
LOC724211	cytochrome P450 9e2	1.53	1.91	1.49	2.20
LOC724234	dolichol kinase	2.87	2.32	3.23	2.38
LOC724286	---	1.14	1.42	1.06	1.49
LOC724308	serine protease 53	1.30	1.26	1.43	1.37
LOC724312	vanin-like protein 1	1.96	1.97	2.13	2.12
LOC724341	endoplasmic reticulum resident protein 29	2.07	2.06	1.72	2.93

LOC724471	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase-like	2.99	2.34	3.43	2.22
LOC724498	---	1.86	2.27	2.22	2.54
LOC724510	neuropathy target esterase sws	1.61	1.55	2.00	1.51
LOC724634	golgin subfamily B member 1	1.66	1.66	1.89	1.92
LOC724724	ATP-dependent 6-phosphofructokinase	2.34	1.70	2.01	1.55
LOC724743	PRKC apoptosis WT1 regulator protein	1.25	1.49	1.67	1.58
LOC724772	leucine-rich repeat neuronal protein 1	1.30	1.35	1.51	1.45
LOC724877	integral membrane protein GPR155	3.58	3.53	2.10	3.61
LOC724912	anoctamin-1-like	1.84	1.92	2.21	1.85
LOC725031	elongation of very long chain fatty acids protein 6	5.44	3.51	5.00	4.19
LOC725068	dynein heavy chain, cytoplasmic	2.33	2.35	3.02	2.48
LOC725087	probable cytochrome P450 6a14	1.92	1.69	2.35	1.94
LOC725105	UDP-sugar transporter UST74c	1.25	1.59	1.57	1.70
LOC725150	---	3.21	2.01	3.26	3.04
LOC725289	low-density lipoprotein receptor-related protein 2	1.28	1.74	1.48	1.70
LOC725320	N-alpha-acetyltransferase 35, NatC auxiliary subunit	2.11	1.64	2.82	1.68
LOC725389	box A-binding factor	1.52	1.96	1.82	2.01
LOC725396	THO complex subunit 4	1.21	1.49	1.40	1.54
LOC725456	sodium-dependent nutrient amino acid transporter 1	1.30	1.52	1.36	1.60
LOC725621	---	1.14	1.28	1.42	1.30
LOC725638	---	1.57	1.66	1.82	2.07
LOC725682	---	3.78	2.97	3.77	2.94
LOC725701	sialin-like	1.16	1.31	1.34	1.49
LOC725950	PHD finger protein 20	1.49	1.48	1.70	1.62

LOC726178	hrp65 protein	1.09	1.14	1.33	1.24
LOC726210	myogenesis-regulating glycosidase	1.26	1.13	1.36	1.27
LOC726230	cAMP-responsive element-binding protein-like 2	1.76	1.74	1.94	1.93
LOC726309	protein artichoke	1.94	1.83	2.72	1.93
LOC726315	sphingomyelin phosphodiesterase 1	1.95	1.69	2.17	1.82
LOC726352	transmembrane protease serine 11B-like protein	1.23	1.31	1.43	1.44
LOC726409	peptidyl-prolyl cis-trans isomerase H	1.84	2.44	1.71	3.19
LOC726524	chromodomain-helicase-DNA-binding protein 7	1.33	1.66	1.67	1.86
LOC726550	CCAAT/enhancer-binding protein gamma	9.78	9.95	9.88	10.23
LOC726609	N-alpha-acetyltransferase 11	1.04	1.39	1.18	1.28
LOC726616	---	1.32	1.36	1.34	1.31
LOC726692	histone H3.3	1.51	1.23	1.24	1.33
LOC726706	F-box/LRR-repeat protein 7	1.58	1.80	1.75	1.95
LOC726708	F-box only protein 28	3.08	2.49	2.94	2.49
LOC726736	laminin subunit beta-1	3.71	2.82	2.81	2.54
LOC726818	beta-hexosaminidase subunit beta	1.78	1.78	2.21	1.85
LOC726855	vegetative cell wall protein gp1	1.02	1.09	1.26	1.06
LOC726883	CD109 antigen	2.86	2.82	3.57	2.76
LOC726906	---	2.54	2.30	3.09	2.87
LOC726934	PI-PLC X domain-containing protein 1	1.12	1.25	1.24	1.35
LOC726941	ATP-binding cassette sub-family A member 2	2.15	2.45	2.86	2.45
LOC727243	---	1.33	1.97	1.68	2.07
LOC727293	guanine deaminase	1.67	1.91	1.77	1.93
LOC727447	nuclear pore complex protein Nup153	2.51	1.73	2.39	2.26
LOC727598	probable cytochrome P450 6a13	2.36	2.14	2.67	2.39

LOC727630	V-type proton ATPase 116 kDa subunit a	2.26	2.01	2.98	2.34
Pgrp-s2	peptidoglycan recognition protein S2	1.93	2.15	1.67	1.69
Rnft2	ring finger protein, transmembrane 2	1.79	2.19	1.68	2.03
Vamp7	vesicle-associated membrane protein 7	1.49	1.33	1.54	1.32

Table S3. Common significantly downregulated genes and their shrunken log2 fold changes in VR and VRL compared to VS and VSH colonies that are SB-positive (SB+) and SB-negative (SB-). Some genes are uncharacterized.

Gene	Description	VR_SB-	VR_SB+	VRL_SB-	VR_SB+
Ac3	adenylate cyclase 3	-1.10	-1.61	-1.49	-1.99
Amih	hyperpolarization-activated ion channel	-1.70	-1.88	-2.19	-2.36
GluCl	glutamate-gated chloride channel	-1.28	-1.62	-1.44	-2.31
LOC100576597	---	-1.63	-1.67	-3.77	-3.65
LOC100578818	amyloid beta A4 precursor protein-binding family B member 1-interacting protein	-1.02	-1.37	-1.37	-1.39
LOC102654975	zinc finger protein 628-like	-1.19	-1.54	-1.31	-1.72
LOC102655034	calcium and integrin-binding family member 3	-1.86	-1.63	-4.44	-4.58
LOC102656661	---	-1.48	-1.89	-1.70	-2.02
LOC406124	gamma-aminobutyric acid receptor subunit beta	-1.20	-1.80	-1.53	-2.68
LOC408914	BAI1-associated protein 3	-1.14	-1.34	-1.56	-1.86
LOC409791	cAMP-dependent protein kinase catalytic subunit	-1.11	-1.57	-1.29	-1.76
LOC410013	calexcitin-2	-1.14	-1.64	-1.55	-1.80
LOC410259	nuclear factor 1 X-type	-1.13	-1.33	-1.62	-1.73
LOC411199	sodium-dependent neutral amino acid transporter B(0)AT3	-2.20	-2.62	-3.37	-4.11
LOC411953	Peter pan	-1.05	-1.62	-1.53	-2.12

LOC551086	zinc finger protein chinmo	-1.42	-1.91	-1.93	-2.35
LOC724220	synaptotagmin-10	-1.96	-1.79	-1.69	-1.91
LOC724599	chromodomain-helicase-DNA-binding protein 1	-1.19	-2.06	-1.44	-2.24
LOC725127	---	-1.58	-1.48	-4.70	-4.76
LOC725220	homeobox protein Nkx-6.1	-4.20	-4.05	-5.13	-1.89
LOC726150	transcription factor Sox-21-B	-2.30	-2.27	-5.53	-5.04
LOC726399	transcriptional activator protein Pur-beta	-1.12	-1.45	-1.33	-1.74
nanos	nanos	-1.04	-1.10	-1.49	-1.61

Figure S1. The mean quality value (Phred score) across each base position in the read for all raw reads across all the samples. Base calls are either very good quality calls (green), calls of reasonable quality (orange), or calls of poor quality (red).

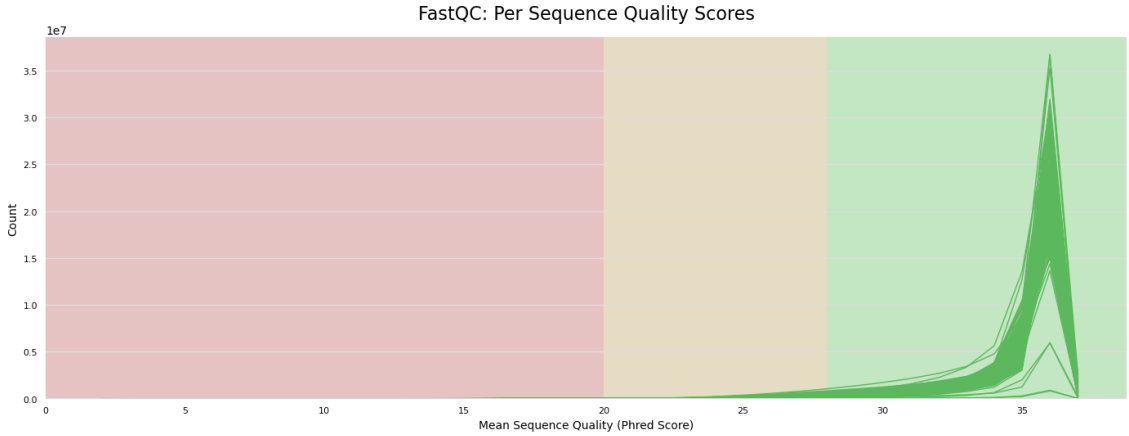

Figure S2. The mean quality value (Phred score) per sequence read for all raw reads across all the samples. Reads are either of very good quality (green), of reasonable quality (orange), or of poor quality (red).

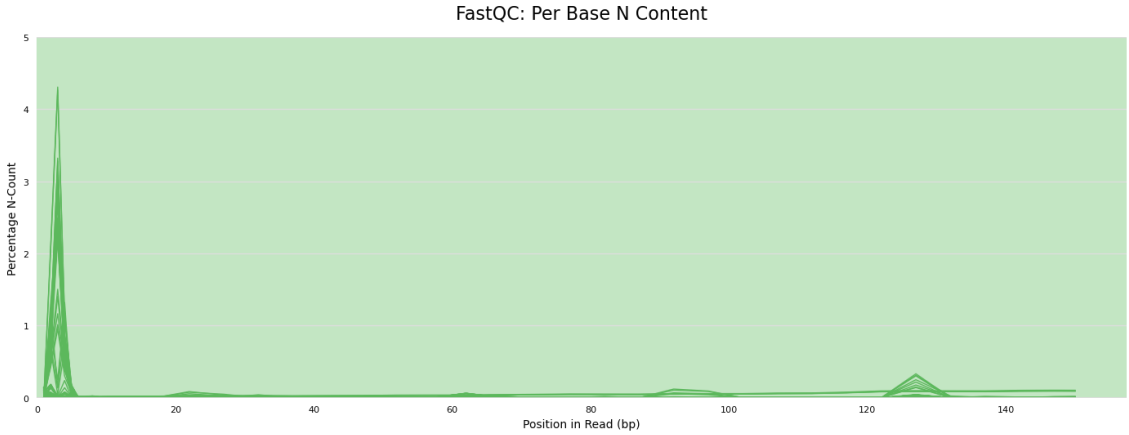

Figure S3. The percentage of base calls at each position along reads for which an ambiguous base ‘N’ was called.

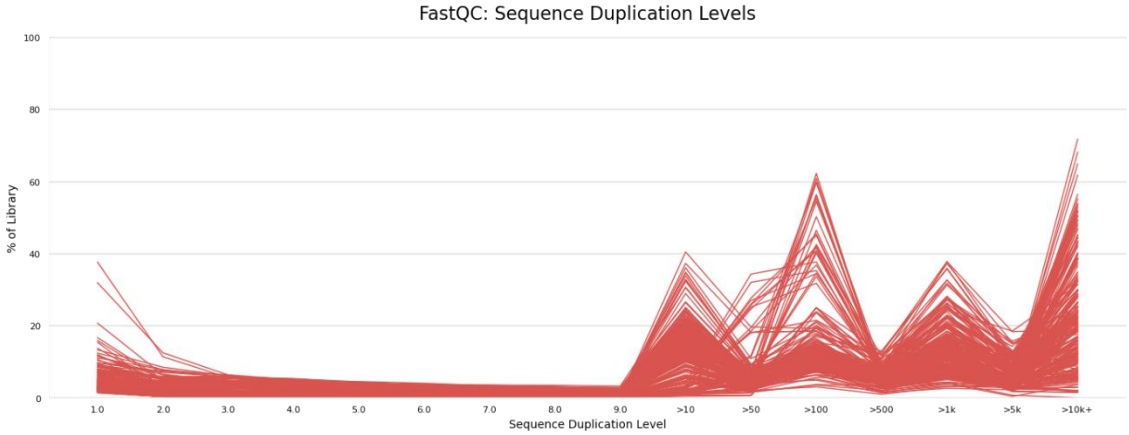

Figure S4. The relative level of duplication found for the first 100,000 sequences in each file.

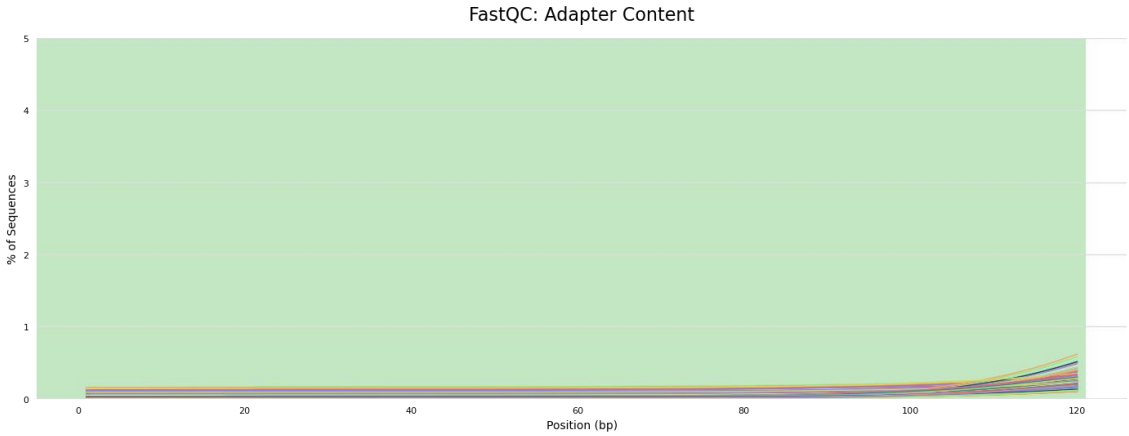

Figure S5. The cumulative percentage count of reads containing adapter sequences at each position. Only samples with ≥ 0.1% adapter contamination are shown