Purdue University
Purdue e-Pubs

Open Access Dissertations | Theses and Dissertations

Winter 2014

Changes in Gene Expression and Viral Titer in Varroa Jacobsoni Mites After a Host Shift Asian to European Honey Bees

Gladys K. Andino Bautista
Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations

Part of the Bioinformatics Commons, Biology Commons, and the Entomology Commons

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Gladys K. Andino Bautista

Entitled
Changes in gene expression and viral titer in Varroa jacobsoni mites after a host shift from Asian to
European honey bees

For the degree of   Doctor of Philosophy

Is approved by the final examining committee:

Greg J. Hunt

Michael Gribskov

Jeffrey J. Stuart

Catherine Hill

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement,
Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation
adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of
copyrighted material.

Greg J. Hunt

Approved by Major Professor(s): _____

_____

Approved by: Steve Yaninek   12/05/2014

Head of the Department Graduate Program   Date

CHANGES IN GENE EXPRESSION AND VIRAL TITER IN VARROA

JACOBSONI MITES AFTER A HOST SHIFT FROM ASIAN TO EUROPEAN

HONEY BEES

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Gladys K. Andino Bautista

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2014

Purdue University

West Lafayette, Indiana

To my husband, Jesse T. Hoteling

"a partner, friend and the one whom I love…"

To my parents, Reinaldo Andino and Ernestina Bautista

"Honor your father and mother" which is the first commandment with a promise "so

that it may go well with you and that you may enjoy long life on the earth"

-Ephesians 6:2-3

To my dear sister, Silvia Bautista

"Many women do noble things, but you surpass them all" – Proverbs 31:29

To my family

ACKNOWLEDGMENTS

Above all I want to thank my heavenly Father for carrying me through ALL during these years at Purdue. This time has been a life changing experience filled with all kinds of experiences and in every step my Lord and Savior Jesus has been faithful. Thank you Lord Jesus that Purdue University was the place for me, thank you for all the blessings and challenges along the way, they have made me stronger, humble and they have taught me more about who you are and your unfailing love for me.

I am grateful to my advisor, Dr. Greg Hunt for giving me this wonderful opportunity and for being supportive of my ideas and projects. I am very grateful to my committee members, Dr. Catherine Hill, Dr. Jeff Stuart, Dr. Maureen McCaan for all their support throughout my program. I am grateful for their advice and encouragement. I am especially grateful to Dr. Michael Gribskov for all his support and encouragement on both academic and life changing issues. I am very thankful that he made the time to mentor me and guide me all the way to the end of my program. I am thankful that almost everything that I have learned about bioinformatics is because of his teaching. Thank you Dr. Gribskov for challenging me and believing in me, I am forever grateful.

I thank Krispn Given for being a friend and for all his help during my time at the bee lab, thank you for teaching me how to take care of the bees, I learned a lot from you.

I am very thankful to all my friends and colleagues, that I have met during these years at Purdue, these years would not have been wonderful without you guys! Thank you, Cynthia, Gloria, Diego, Julia, Victoria, Oscar, Helena, Ulianova, Lucio, Marie and Karen. I thank my home group from Upper Room Christian Fellowship, for being my family, thank you for all your prayers, encouragement and support. Thank you for being a clear example of God's goodness and love. You will be always close to my heart!

I am very blessed and thankful to my family; my parents, my sister Silvia, my brother Rey, Luis and Marcela, they have been a support and encouragement to me, Thank you for believing in me and for loving me with all that you have. I am so blessed to have the family that I have, I would not change them for anything, I love you all!

And last but never least; I am very thankful to my husband Jesse Hoteling. He has been there with me through all, the ugly and the pretty, the laughs and the tears, the success and the failures, the good nights of sleep and the ones without sleep, thank you for taking this journey with me. I am so thankful to you and for you. Your support, encouragement and waiting have been very precious to me. Thank you for your patience and help, I can say with all certainty that we accomplished this together. I could not have done it without you, love you! I am forever thankful!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABPV	Acute Bee Paralysis Virus
Ac	*A. cerana*
Am	*A. mellifera*
BLAST	Basic Local Alignment Search Tool
BP	Base Pairs
BRF	Biomolecular Resource Facility
BQCV	Black Queen Cell Virus
CDEMG	Consistently Differentially Expressed Mite Genes
cDNA	Complementary DNA
CEGMA	Core Eukaryotic Genes Mapping Approach
cox1	Cytochrome Oxidase 1
cpm	Counts Per Million
CPU	Central Processing Unit
DE	Differentially Expressed
DEG	Differentially Expressed Genes
DEMVG	Differentially Expressed Mite Virus Genes
DWV	Deformed Wing Virus
EC	Expected Counts
edgeR	Empirical Analysis of Digital Gene Expression Data in R
FDR	False Discovery Rate
FPKM	Fragments Per Kilobase Per Million
GO	Gene Ontology
IAPV	Israeli Acute Paralysis Virus
J haplotype	Japan Haplotype
KBV	Kashmir Bee Virus
K haplotype	Korean Haplotype
mtDNA	Mitochondrial DNA
NaCl	Sodium Chloride
NCBI	National Center for Biotechnology Information
NR	Non- Redundant
PASA	Program to Assemble Spliced Alignments
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PFAM	Protein Families
PGCF	Purdue University Genomics Core Facility
PMS	Parasitic Mite Syndrome

PNG	Papua New Guinea
PPDE	Posterior Probability of Being Differentially Expressed
RSEM	RNA-Seq by Expectation-Maximization
SBV	Sacbrood Virus
SC	St. Christobel Islands
SI	Solomon Islands
tmHMM	Hidden Markov Model for predicting Transmembrane Helices
TMM	Trimmed Mean of M-values
VD	*V. destructor*
VDV-1	*V. destructor* Virus-1
VJ	*V. jacobsoni*

ADDITIONAL FILES

Additional file 1: Figure S1. Distribution of normalized FPKM. Histograms of the

normalized FPKM values in the different samples showing the distribution of these

values about the median

Additional file 2: Tables S1.*Varroa jacobsoni*, 1013 differentially expressed genes (FDR

< 0.05 and absolute logFC ≥ 2) using EdgeR

Additional file 3: Tables S2.*Varroa jacobsoni*, 586 differentially expressed genes (FDR

< 0.05) using DESeq2

Additional file 3: Tables S3.*Varroa jacobsoni*, 6908 differentially expressed genes (FDR

< 0.05) using EBEseq

Additional file 4: Tables S4. Selected target genes (37,661) used as a reference set for

Fisher's exact test

Additional file 5: Tables S5 GO terms significantly over-represented in mite genes that

are down-regulated in the *A. mellifera* host, cluster 1 (23 genes)

Additional file 6: Table S6. Genes that had a significant Blastx similarity hit (e-value ≤

1e-06) to the *M. occidentali*s predicted peptideds (cluster 1)

Additional file 6: Table S8. Genes that had a significant Blastx similarity hit (e-value ≤

1e-06) to the *M. occidentali*s predicted peptideds (cluster 2)

Additional file 6: Table S10. Genes that had a significant Blastx similarity hit (e-value ≤ 1e-06) to the *M. occidentali*s predicted peptideds (cluster 3)

Additional file 7: Tables S7. GO terms significantly over-represented in mite genes that are down-regulated in the *A. mellifera* host, cluster 2 (208 genes)

Additional file 8: Table S9. GO terms significantly over-represented in mite genes that are down-regulated in the *A. mellifera* host, cluster 3 (56 genes)

Additional file 9: Table S11. Blastn results DWV_PA isolates vs. *V. jacobsoni* virus assembly (E-value ≤ 1e-06)

ABSTRACT

Andino, Gladys K. Ph.D., Purdue University, December 2014. Changes in gene expression and viral titer in *Varroa jacobsoni* mites after a host shift from Asian to European honey bees. Major Professor: Greg J. Hunt.

Honey bees (*Apis mellifera* L.) are the most important insects for the pollination of crops and wildflowers. However, they have experienced increasing colony die-offs during the past two decades. Multiple species of parasitic mites have been described that affect honey bees. The most important species in beekeeping belong to the genus *Varroa* (*Varroa jacobsoni* and *Varroa destructor*). *Varroa* mite parasitism of honey bees is thought to be the most significant cause of colony mortality worldwide, and mite resistance to active ingredients of acaricides has become common. *V. destructor* causes direct impacts on bee production as well as indirect effects on bee health by vectoring viruses and other pathogens. These large ectoparasitic mites are associated with a condition known as parasitic mite syndrome, or PMS. When colonies exhibit PMS pathogens, brood diseases and viruses are present at unusually high levels. The open wound caused during feeding can allow microorganisms to enter and weaken the host, and mites themselves are vectors for viruses and perhaps other bee pathogens. There are a number of studies that also suggest that the primary cause of colony mortality is the viruses associated with the mites. Until recently, *Varroa jacobsoni* was known to only live and reproduce in Asian honey bee (*Apis cerana*) colonies while *V. destructor*

successfully reproduces in both *A. cerana* and *A. mellifera* colonies. However, we have sampled an island population of *V. jacobsoni* that is highly destructive to *A. mellifera*, the primary species used for pollination and honey production. These recently discovered populations of mites represent an enormous threat to apiculture.

For the first part of this project, we focused on investigating the differences in gene expression between populations of *V. jacobsoni* mites reproducing on *A. cerana* and those reproducing on *A. mellifera* (detailed description of the methods and results are discussed in chapter two). Briefly, we sequenced and assembled a *de novo* transcriptome of *V. jacobsoni*. We also performed a differential gene expression analysis contrasting biological replicates of *V. jacobsoni* populations that differed in their ability to parasitize *A. mellifera*. Using the edgeR, EBSeq, and DESeq R packages for the differential gene expression analysis, we found 287 differentially expressed genes (FDR ≤ 0.05), of which 91% were up-regulated in mites paraistizing *A. mellifera.* Furthermore, we searched for orthologous genes in public databases and were able to associate 100 of these 287 differentially expressed genes with a functional description. The mites found parasitizing *A. mellifera* showed substantially more variation in expression among replicates. While a small set of genes including, putative transcription factors and digestive tract developmental genes showed reduced expression in the mites, the vast majority of differentially expressed genes were up-regulated. These up-regulated genes are associated with mitochondrial respiratory function and apoptosis, suggesting that mites on this host may have experienced higher stress levels and were less optimally adapted to parasitize

them. Some genes involved in reproduction and oogenesis were also over-expressed, which should be further studied in regards to this host shift.

The second part of the study was carried out to survey, for the first time, the viruses associated with *V. jacobsoni* and to determine whether these viruses played a role in mite colonization (detailed description of methods and results are discussed in chapter three). Briefly, we assembled a virus transcriptome of *V. jacobsoni* to provide the first survey of pathogens in this species. Among the list of putative viruses found were Deformed Wing Virus (DWV), Dragonfly Cyclovirus 1, Farmington Virus, Formica Exsecta Virus 2, Halyomorpha Halys Virus, Heliconius Erato Iflavirus, Kakugo Virus, Kirsten Murine Sarcoma Virus, Sacbrood Virus, Spodoptera Exigua Iflavirus 1. Our findings suggest that overall all the mite samples had similar viruses, with slight variations in the abundance of certain sequences. A search against a honey bee associated microbe database revealed the likely presence of Macula-like virus (Tymoviridae), a microsporidian and a spiroplasma, all of which have been previously reported for *A. mellifera*. The results from the expression analysis suggested that there are four different viruses that are differentially expressed between mites on *A. cerana* and *A. mellifera*, three of which were found up-regulated on the *A. mellifera* hosts. Among those three sequences, we found a match to Dragonfly Cyclovirus PK5222. Although this exact sequence was not similarly expressed across all samples of the two hosts, this Cyclovirus appears to be one of the most abundant viruses that were common to all samples. This analysis also showed a clear geographical clustering of the samples according to the expression patterns. Samples collected in Solomon Islands (SI) clustered together and the samples collected in Papua

New Guinea (PNG) also clustered together in their own group. The lack of clear expression differences between the two hosts suggests that the viruses are not critical for host acquisition or overcoming host defenses. We have solid evidence that DWV infects *V. jacobsoni* and it is surprising that it is most closely related to an isolate from North America. As far as we know this is the first time that DWV has been reported in *V. jacobsoni* and indeed the virus pathogens of this mite had not previously been determined.

CHAPTER 1 LITERATURE REVIEW

1.1 Background on *Varroa* parasitism of honey bees

Multiple species of parasitic mites have been described that affect honey bees. The most important species in beekeeping belong to the genus *Varroa* (*V. jacobsoni* and *V. destructor*). *Varroa* mite parasitism of honey bees is thought to be the most significant cause of colony mortality worldwide, and mite resistance to active ingredients of acaricides has become common (Elzen et al., 1999a; Elzen & Westervelt, 2002; Elzen et al., 1999b; Milani, 1999; Skinner et al., 2003; Spreafico et al., 2001; Thompson et al., 2002) *V. destructor* causes direct impacts on bee production as well indirect effects on bee health by transmitting viruses and other pathogens (Boecking & Spivak, 1999; Bowen-Walker et al., 1999; Webster & Delaplane, 2001). Recent survey results agree that one of the common denominators among the winter honey bee colony losses is *Varroa* mites. A survey of colony losses in 18 countries in the winter of 2007/2008 showed mortality levels ranging from 10-36% (Currie et al., 2010; vanEngelsdorp et al., 2010). Guzmán-Novoa (Guzmán-Novoa et al., 2010) reported that in Ontario, Canada, the best predictor of winter losses in a survey of 400 hives during 2007-2008 was *V. destructor*; around 27% of the colonies died and 85% of their winter mortality was significantly associated with high mite infestations. Results from a survey from Norway

where *V. destructor* has not yet reached the northern regions of the country, revealed an average of 10.3% colony loss during the winters of 1999-2008 in regions were *V. destructor* is present and 6.3% loss in regions where mites are absent (Dahle, 2010).

1.2 The life cycle of *Varroa* mites

The *Varroa* mite's life cycle can be divided in two phases: 1) the phoretic phase, during which the adult female mites live on the adult bee. The crab-like shape of the female mites allows them to hide beneath the abdominal sclerites of the honey bees and stay there until they are ready for reproduction, and 2) the reproductive phase, where the female mite reproduces inside the brood cell of the honey bee (Martin, 2001a). After the female mite invades the cell, the first egg laid becomes a haploid male, which will later mate with his sisters to give rise to the next generation. Mated female mites are phoretic on adult honey bees but male mites die after the bee emerges. *Varroa* mites feed on the hemolymph of the larva, pupa and adults. When the mites feed on the young they cause damage that results in weak adults that are smaller or diseased (Rosenkranz et al., 2010).

1.3 The Role of Viruses in Parasitic Mite Syndrome (PMS)

Mites are able to transmit different viral diseases, (Bowen-Walker et al., 1999) and the open wound caused during feeding can also allow other microorganisms to enter and kill the host (Bailey, 1981). The appearance of brood diseases and viral diseases in a colony collapsing from *Varroa* parasitism has been called parasitic mite syndrome. There are a number of studies that suggest that the primary cause of colony mortality is the viruses associated with the mites. Honey bee viruses were not considered a serious problem prior

to the spread of *Varroa* in Europe (Chen & Siede, 2007; de Miranda & Genersch, 2010; Francis et al., 2013a; Francis et al., 2013b; Genersch & Aubert, 2010). Most of the viruses associated with *Varroa* are positive-strand RNA viruses of the family Dicistroviridae. There are two species complexes in this family that commonly use *V. destructor* as a host and a vector. The first complex consists of deformed wing virus (DWV) and *V. destructor* virus-1 (VDV-1). Recombinants between DWV and VDV-1 have been found replicating in symptomatic bees parasitized by *Varroa* (Zioni et al., 2011). The second species complex consists of Israeli acute paralysis virus (IAPV), acute bee paralysis virus (ABPV) and Kashmir bee virus (KBV). Levels of black queen cell virus (BQCV) and sacbrood virus (SBV) have also been found to be elevated in colonies with high *Varroa* infestations. ABPV and IAPV have been associated with winter losses (Cox-Foster et al., 2007b; Nguyen et al., 2011).

DWV has become the most common honey bee virus worldwide. The incidence and abundance of DWV and VDV-1 correspond closely to the spread of *V. destructor*, which suggests that the mites act as reservoirs for the virus (Genersch & Aubert, 2010; Martin et al., 2012; Mondet et al., 2014b). The diversity of DWV isolates in honey bees can also drastically decrease with parasitism. Transmission of DWV by a mite or by experimental injection appears to select for a particularly virulent isolate (Ryabov et al., 2014). Whether the feeding of *Varroa* mites suppresses immune gene expression in honey bees is controversial, but the feeding or artificial wounding of the pupa increases DWV titers (Kuster et al., 2014; Nazzi et al., 2012; Ryabov et al., 2014; Yang & Cox-Foster, 2005). Although a lot has been learned about *Varroa*-virus interactions it is not yet known

whether the virus plays a role in allowing the mite to colonize the host, or whether these viruses are also present in *V. jacobsoni*.

1.4 Origin and spread of *V. jacobsoni* and *V. destructor*

V. jacobsoni was first found parasitizing the Asian honey bee *A. cerana*, in Java, Indonesia at the beginning of the 19th century. In 1957 these mites were reportedly found parasitizing the European honey bee *A. mellifera* in colonies from the Philippines. However, differences in virulence, morphology, reproductive isolation, and genetics were observed between *Varroa* mites infesting the two species of honey bees, leading to the conclusion that *V. jacobsoni* was a complex of two species; and a new name *V. destructor* was given to a group of haplotypes that were reproductively isolated from those called *V. jacobsoni* (Anderson & Trueman, 2000). Furthermore, a total of 18 haplotypes (mites with unique mtDNA sequences) were identified by analyzing an mtDNA fragments, and coding a portion of the cytochrome oxidase I (*cox1*) gene. The sequence of this fragment comprises 458 base pairs (bp). Each of the haplotypes that will be mentioned in this review were named after the country/island where they were found for the first time infesting *A. cerana* (Table 1.1).

Six out of 18 haplotypes identified (China 1, Japan, Korea, Nepal, Sri Lanka and Vietnam) were allocated to the new species *V. destructor*. Interestingly, two of these six haplotypes (Korea and Japan) were found infesting both *A. cerana* and *A. mellifera*. The Korea, or K haplotype is most common, and has been identified in *A. mellifera* colonies in Europe, Middle East, Africa, Asia, and the Americas. The Japan, or J haplotype has been found

infesting *A. mellifera* in Japan, Thailand and the Americas. The J and K haplotypes are reported to be the only two haplotypes that successfully reproduce in *A. mellifera* outside of Asia (Navajas et al., 2010). Furthermore, the K haplotype has been characterized as being more virulent than the J haplotype (Anderson & Trueman, 2000; De Guzman et al., 1997; de Guzman & Rinderer, 1999; Garrido et al., 2003). The remaining four haplotypes of *V. destructor* (China 1, Nepal, Sri Lanka and Vietnam) were found infesting *A. cerana* only. The haplotypes (Ambon, Bali, Borneo, Flores, Java, Lombok, Malaysia Peninsular, Sumatra, and Sumbawa) were described as *V. jacobsoni* affecting *A. cerana* only.

In China, *V. jacobsoni* has not been found. Only the K haplotype was found infesting *A. mellifera* in China but three other haplotypes of *V. destructor* were found infesting *A. cerana* colonies, suggesting reproductive isolation of haplotypes (Zhou et al., 2004). A comparable situation was also reported by Anderson and Trueman, (2000) in Java, Indonesia, where reproductive isolation was observed between sympatric populations of *Varroa* mites. The Korean haplotype of *V. destructor* was only found on *A. mellifera* and the Java haplotype of *V. jacobsoni* was found only on *A. cerana*. This isolation is mainly maintained by the inability of the Java haplotype to reproduce in *A. mellifera*. Another similar case is the one reported by (Fuchs & Anderson, 2000) where *Varroa* mites infesting *A. cerana* and *A. mellifera* belong to two different haplotypes, (Vietnam and Korea respectively), also suggesting genetic isolation.

1.5 Genetic variation in *V. destructor* infesting *A. mellifera*

Further studies to try to understand the genetic variability of *V. destructor* haplotypes affecting *A. mellifera* (J and K haplotypes) have been carried out using several molecular approaches. Throughout the years different groups of investigators have reached the same conclusion, that there is a remarkable absence of polymorphisms in the *Varroa* populations of Europe and USA collected on *A. mellifera* (Anderson & Trueman, 2000; Biasiolo, 1992; Kraus & Hunt, 1995). In addition, a widespread study using microsatellite markers in 45 different populations of *Varroa* mites from around the world showed a relative lack of polymorphisms within each of the two haplotypes that successfully infest *A. mellifera* outside of Asia. These results suggested that these two haplotypes, J and K, each correspond to a single host capture event, followed by a rapid spread worldwide. Two routes of invasion of *V. destructor* into the Americas and specifically into the USA have been proposed based on the dates and places where they were first detected (Navajas, 2010; Navajas et al., 2010; Solignac et al., 2005). The J haplotype first shifted from *A. cerana* to *A. mellifera* in Japan during the last century following the introduction of *A. mellifera*. From Japan, it spread to Thailand, then to Paraguay in (1971), and then to Brazil in 1972, and later North America in 1987. The K haplotype first shifted from *A. cerana* to *A. mellifera* near Vladivostok (north of the Korean peninsula), following the introduction of *A. mellifera* from Ukraine in the 1950's. Later, it spread from eastern Russia to western Russia, then to Bulgaria in 1972, and Germany in 1977, and then continued spreading around Europe and also the U.S.

1.6 Genetic variation of *V. destructor* infesting *A. mellifera* and *A. cerana*

To further understand the genetic variability of *V. destructor* haplotypes, researchers carried out a study to determine how variable *V. destructor* is on its primary host (*A. cerana*) and its new host (*A. mellifera*) in Asia where the host shift originated. A total of 21 samples of female *Varroa* mites were analyzed, 14 were found infesting *A. cerana* and seven were found infesting *A. mellifera*. Samples were first analyzed based on sequence obtained from the *cox1* gene (458 bp) and then further analyzed based on 2700 bp sequence of 4 mtDNA genes (*cox1*, *cox3*, *atp6* and *cytb*). Results revealed more genetic variability of *V. destructor* haplotypes infesting both honey bee species in Asia (Table 1.2). A total of 18 haplotypes were identified in Asia, which included a new China haplotype (China 3) found in *A. cerana,* and new variants of the J and K haplotypes found in well established infestations of *A. mellifera* (Navajas et al., 2010) (Table 1.2). These results supported the hypothesis that the only species of mite affecting *A. mellifera* is *V. destructor.* However, new variants of the J and K haplotypes were found affecting *A. mellifera* in Asia (specifically in China, Taiwan, Vietnam and Thailand). These findings suggest that these new haplotypes may represent a potential threat to *A. mellifera* outside of Asia (Navajas et al., 2010). Furthermore, these findings highlight the possibility that new *Varroa* types could expand their host preferences and become a new threat to apiculture worldwide. In addition, these findings suggested that the colonization of *A. mellifera* by new haplotypes is dependent on how long these haplotypes are exposed to *A. mellifera*. Possibly, longer exposure to a new host could result in the adaptation of one type of mite to a specific characteristic of a local host. Given a worst case scenario, it can

be predicted that new *Varroa* types, including those reported to be non-reproducing in *A. mellifera* will eventually colonize the European honey bee outside of Asia.

Currently, nothing has been published about *V. jacobsoni* haplotypes reproducing on *A. mellifera*. However, samples of *V. jacobsoni* reproducing on the European honey bee have been collected from Papua New Guinea (Roberts et al. submitted) and will be subjected to RNA sequencing in the project described here.

References

1. Anderson DL & Trueman JW (2000) *Varroa jacobsoni* (Acari: Varroidae) is more than one species. Exp Appl Acarol 24: 165-189.

2. Bailey L (1981) Honey bee pathology. London ; New York : Academic Press, London ; New York, pp. 124.

3. Biasiolo A (1992) Lack of allozyme variability among*Varroa* mite populations. Exp Appl Acarol 16: 287-294.

4. Boecking O & Spivak M (1999) Behavioral defenses of honey bees against *Varroa jacobsoni* Oud. Apidologie 30: 141-158.

5. Bowen-Walker PL, Martin SJ & Gunn A (1999) The transmission of deformed wing virus between honeybees (*Apis mellifera* L.) by the ectoparasitic mite *Varroa jacobsoni* Oud. J Invertebr Pathol 73: 101-106. doi:10.1006/jipa.1998.4807.

6. Chen YP & Siede R (2007) Honey bee viruses. Adv Virus Res 70: 33-80. doi:10.1016/S0065-3527(07)70002-7.

7. Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan PL, Briese T, Hornig M, Geiser DM, Martinson V, vanEngelsdorp D, Kalkstein AL, Drysdale A, Hui J, Zhai JH, Cui LW, Hutchison SK, Simons JF, Egholm M, Pettis JS & Lipkin WI (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. Science 318: 283-287. doi:Doi 10.1126/Science.1146498.

8. Currie RW, Pernal SF & Guzmán-Novoa E (2010) Honey bee colony losses in Canada. Journal of Apicultural Research 49: 104-106.

9. Dahle B (2010) role of *Varroa destructor* for honey bee colony losses in Norway. Journal of Apicultural Research.

10. de Guzman L, Rinderer T & Stelzer J (1997) DNA evidence of the origin of *Varroa jacobsoni* Oudemans in the Americas. Biochem Genet 35: 327-335.

11. de Guzman L & Rinderer TE (1999) Identification and comparison of *Varroa* species infesting honey bees. Apidologie 30: 85-95.

12. de Miranda JR & Genersch E (2010) Deformed wing virus. J Invertebr Pathol 103 Suppl 1: S48-61. doi:10.1016/j.jip.2009.06.012.

13. Elzen P, Baxter J, Spivak M & Wilson W (1999a) Amitraz resistance in *Varroa*: new discovery in North America. American Bee Journal.

14. Elzen P, Eischen FA, Baxter JR, Elzen GW & Wilson WT (1999b) Detection of resistance in US *Varroa jacobsoni* Oud.(Mesostigmata: Varroidae) to the acaricide fluvalinate. Apidologie 30: 13-18.

15. Elzen P & Westervelt D (2002) Detection of coumaphos resistance in *Varroa destructor* in Florida. American Bee Journal 142: 291-292.

16. Francis RM, Nielsen SL & Kryger P (2013a) Patterns of viral infection in honey bee queens. Journal of General Virology 94: 668-676. doi:10.1099/vir.0.047019-0.

17. Francis RM, Nielsen SL & Kryger P (2013b) *Varroa*-virus interaction in collapsing honey bee colonies. PLoS One 8: e57540. doi:10.1371/journal.pone.0057540.

18. Fuchs S & Anderson DL (2000) A scientific note on the genetic distinctness of *Varroa* mites on *Apis mellifera* L. and on *Apis cerena* Fabr. in North Vietnam. Apidologie 31: 459-460.

19. Garrido C, Rosenkranz P, Paxton RJ & Goncalves LS (2003) Temporal changes in *Varroa destructor* fertility and haplotype in Brazil. Apidologie 34: 535-541. doi:Doi 10.1051/Apido:2003041.

20. Genersch E & Aubert M (2010) Emerging and re-emerging viruses of the honey bee (*Apis mellifera* L.). Vet. Res. 41: 54.

21. Guzmán-Novoa E, Eccles L, Calvete Y, McGowan J, Kelly PG & Correa-Benítez A (2010) *Varroa destructor* is the main culprit for the death and reduced populations of overwintered honey bee (*Apis mellifera*) colonies in Ontario, Canada. Apidologie 41: 443-450. doi:10.1051/apido/2009076.

22. Kraus B & Hunt G (1995) Differentiation of Varroa jacobsoni Oud populations by random amplification of polymorphic DNA (RAPD). Apidologie 26: 283-290.

23. Kuster RD, Boncristiani HF & Rueppell O (2014) Immunogene and viral transcript dynamics during parasitic *Varroa destructor* mite infection of developing honey bee (*Apis mellifera*) pupae. J Exp Biol 217: 1710-1718. doi:10.1242/jeb.097766.

24. Martin S (2001) Biology and life-history of Varroa mites. Mites of the honey bee. Dadant & Sons, Hamilton, IL: 131-148.

25. Martin SJ, Highfield AC, Brettell L, Villalobos EM, Budge GE, Powell M, Nikaido S & Schroeder DC (2012) Global honey bee viral landscape altered by a parasitic mite. Science 336: 1304-1306. doi:10.1126/science.1220941.

26. Milani N (1999) The resistance of *Varroa jacobsoni* Oud. Apidologie 30: 229-234.

27. Mondet F, de Miranda JR, Kretzschmar A, Le Conte Y & Mercer AR (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 Pathog 10: e1004323. doi:10.1371/journal.ppat.1004323.

28. Navajas M (2010) Tracking the colonisation history of the invasive species *Varroa destructor*: Trends in Acarology (ed. by MW Sabelis & J Bruin) Springer Netherlands, pp. 375-378.

29. Navajas M, Anderson DL, de Guzman LI, Huang ZY, Clement J, Zhou T & Le Conte Y (2010) New Asian types of *Varroa destructor*: a potential new threat for world apiculture. Apidologie 41: 181-193. doi:10.1051/apido/2009068.

30. Nazzi F, Brown SP, Annoscia D, Del Piccolo F, Di Prisco G, Varricchio P, Della Vedova G, Cattonaro F, Caprio E & Pennacchio F (2012) Synergistic Parasite-Pathogen Interactions Mediated by Host Immunity Can Drive the Collapse of Honeybee Colonies. PLoS Pathog 8: e1002735. doi:10.1371/journal.ppat.1002735.

31. Nguyen BK, Ribiere M, vanEngelsdorp D, Snoeck C, Saegerman C, Kalkstein AL, Schurr F, Brostaux Y, Faucon JP & Haubruge E (2011) Effects of honey bee virus prevalence, *Varroa destructor* load and queen condition on honey bee colony survival over the winter in Belgium. Journal of Apicultural Research 50: 195-202. doi:Doi 10.3896/Ibra.1.50.3.03.

32. Rosenkranz P, Aumeier P & Ziegelmann B (2010) Biology and control of *Varroa destructor*. J Invertebr Pathol 103 Suppl 1: S96-119. doi:10.1016/j.jip.2009.07.016.

33. Ryabov EV, Wood GR, Fannon JM, Moore JD, Bull JC, Chandler D, Mead A, Burroughs N & Evans DJ (2014) A virulent strain of deformed wing virus (DWV) of honeybees (*Apis mellifera*) prevails after *Varroa destructor*-mediated, or in vitro, transmission. PLoS Pathog 10: e1004230. doi:10.1371/journal.ppat.1004230.

34. Skinner A, Tam J & Melin S (2003) Monitoring for fluvalinate resistant varroa mites and coumaphos efficacy trials in honey bee colonies in Ontario. The Sting 21: 23-25.

35. Solignac M, Cornuet JM, Vautrin D, Le Conte Y, Anderson D, Evans J, Cros-Arteil S & Navajas M (2005) The invasive Korea and Japan types of *Varroa destructor*, ectoparasitic mites of the Western honeybee (*Apis mellifera*), are two partly isolated clones. Proc Biol Sci 272: 411-419. doi:10.1098/rspb.2004.2853.

36. Spreafico M, Eordegh F, Bernardinelli I & Colombo M (2001) First detection of strains of *Varroa destructor* resistant to coumaphos. Results of laboratory tests and field trials. Apidologie 32: 49-56.

37. Thompson HM, Brown MA, Ball RF & Bew MH (2002) First report of *Varroa destructor* resistance to pyrethroids in the UK. Apidologie 33: 357-366.

38. vanEngelsdorp D, Hayes J, Underwood RM & Pettis JS (2010) A survey of honey bee colony losses in the United States, fall 2008 to spring 2009. Journal of Apicultural Research 49: 7-14. doi:Doi 10.3896/Ibra.1.49.1.03.

39. Webster TJ & Delaplane KS (2001) Mites of the honey bee. Dadant & Sons.

40. Yang X & Cox-Foster DL (2005) Impact of an ectoparasite on the immunity and pathology of an invertebrate: evidence for host immunosuppression and viral amplification. Proc Natl Acad Sci U S A 102: 7470-7475. doi:10.1073/pnas.0501860102.

41. Zhou T, Anderson DL, Huang ZY, Huang S, Yao J, Ken T & Zhang Q (2004) Identification of Varroa mites (Acari: Varroidae) infesting Apis cerana and *Apis mellifera* in China. Apidologie 35: 645-654.

42. Zioni N, Soroker V & Chejanovsky N (2011) Replication of Varroa destructor virus 1 (VDV-1) and a Varroa destructor virus 1-deformed wing virus recombinant (VDV-1-DWV) in the head of the honey bee. Virology 417: 106-112. doi:10.1016/j.virol.2011.05.009.

Table 1.1 Identification and classification of *Varroa* haplotypes collected on *A. cerana* from Asia based on mtDNA cox 1 gene sequences

V. jacobsoni haplotypes	*V. destructor* haplotypes	Unresolved haplotypes
Ambon (Indonesia)[1]	China 1[1]	Luzon 1 (Philippines)[1]
Bali(Indonesia)[1]	China 2[2]	Luzon 2 (Philippines)[1]
Borneo (Malaysia)[1]	Japan[§][1]	Mindanao (Philippine)[1]
Borneo 2[3]	Korea[§][1]	
Flores (Indonesia)[1]	Nepal[1]	
India[2]	Pakistan[2]	
Java (Indonesia/PNG)[*][1]	Sri Lanka[1]	
Laos[2]	Vietnam[1]	
Lombok (Indonesia)[1]		
Malaysia Peninsular[1]		
North Thailand 1[2]		
North Thailand 2[2]		
Samui 1[2]		
Sumatra (Indonesia)[1]		
Sumbawa (Indonesia)[1]		

* Specimens of this haplotype were the ones first classified as *V. jacobsoni* at the beginning of the last century.

§ Haplotypes that were also found successfully reproducing in *A. mellifera* worldwide.

[1] First described by Anderson and Trueman, (2000).

[2] Described in Zhou et al. (2004).

[3] First described by Koeniger et al. (2002).

Table 1.2 Modified from Navajas et al. (2010). Identity of *V. destructor* (VD) and *V. jacobsoni* (VJ) haplotypes examined from A. cerana (Ac) and *A. mellifera* (Am) collected in Asia. Mites were first assigned to a known (e.g. K1) or new haplotype group based on partial nucleotide sequence (458 bp) of their mtDNA cox1 gene (Anderson and Trueman, 2000). Haplotypes were further classified (e.g. K1-2) based on concatenated nucleotide sequences of fragments of their mtDNA cox1, cox3, atp6 and cytb gene sequences (2700 bp)

Country	Mite species	Bee host	Collection Year	mtDNA *cox1* seq haplotype [1]	Concatenated mtDNA haplotype
China	VD	Ac	2002	C2	C2-1
China	VD	Ac	2002	**C3** (new)	C3-1
China	VD	Am	2002	K1	**K1-4** (new)
China	VD	Ac	2002	V1	V1-2
China	VD	Ac	2001	C1	C1-1
China	VD	Ac	2002	C1	C1-2
China	VD	Ac	2004	K1	K1-3
China	VD	Am	2004	K1	**K1-2** (new)
China	VD	Ac	2004	K1	K1-3
Japan	VD	Ac	1994	J1	J1-2
Japan	VD	Am	1996	J1	J1-6
Japan	VD(3)	Am	2000	K1	K1-1
Japan	VD	Ac	1998	J1	J1-3
Japan	VD	Ac	1996	J1	J1-4
Korea	VD	Am	1996	K1	K1-1
Russia	VD	Am	1995	K1	K1-1
Taiwan	VD(3)	Am	2002	J1	J1-1
Thailand	VD	Ac	2003	V1	V1-4
Thailand	VJ(2)	Ac	2003	L1	L1-1
Thailand	VJ(2)	Ac	2003	L1	L1-2
Thailand	VD	Am	1997	J1	**J1-5** (new)
Thailand	VD	Ac	2003	V1	V1-3
Vietnam	VD	Ac	1996	V1	V1-1
Vietnam	VD	Am	1996	K1	**K1-2** (new)

[1] Haplotype names have been abbreviated: China 1 (C1), China 2 (C2), China 3 (C3), Japan 1(J1), Vietnam 1 (V1), Korea 1 (K1), Laos 1 (L1).

[2] Haplotype determined on basis of 1635 bp fragments of the mtDNA *cox1*, *cox3* and *atp6* genes. These samples were described as *V. jacobsoni* and found in Taiwan where it is known that both species of mites occur sympatrically.

[3] The concatenated sequences of the K1-1 and J1-1 variants described are those of *V. destructor* which had been previously identified using microsatellites as K1 and J1 types by Solignac et al. (2005) and J and K by Anderson and Trueman, (2000).

CHAPTER 2. DIFFERENTIAL GENE EXPRESSION IN VARROA JACOBSONI

MITES FOLLOWING A HOST SHIFT TO EUROPEAN HONEY

BEES (APIS MELLIFERA)

Abstract

Varroa mites are widely considered the biggest honey bee health problem worldwide. Until recently, *Varroa jacobsoni* was known to only live and reproduce in Asian honey bee (*Apis cerana*) colonies while *V. destructor* successfully reproduces in both *A. cerana* and *A. mellifera* colonies. However, we have sampled an island population of *V. jacobsoni* that is highly destructive to *A. mellifera*, the primary species used for pollination and honey production. These recently discovered populations of mites represent an enormous threat to apiculture. Our aim was to investigate differences in gene expression between populations of *V. jacobsoni* reproducing on *A. cerana* and the population reproducing on *A. mellifera*. Our hypothesis is that genetic variation exists among populations of *V. jacobsoni* that influence gene expression and reproductive status. We sequenced and assembled a *de novo* transcriptome of *V. jacobsoni*. We also performed a differential gene expression analysis contrasting biological replicates of *V. jacobsoni* populations that differ in their ability to parasitize *A. mellifera*. Using the edgeR. EBSeq and DESeq R packages for the differential gene expression analysis we found 287 differentially expressed genes (FDR ≤ 0.05), of which 91% were up regulated

 in mites parasitizing *A. mellifera.*, Further mites found parasitizing *A. mellifera* showed substantially more variation in expression among replicates. We searched for orthologous genes in public databases and were able to associate 100 of these 287 differentially expressed genes with a functional description. There is differential gene expression between the two mite groups, with more variation between gene expression among mites parasitizing *A. mellifera*. A small set of genes showed reduced expression in mites on this host, including putative transcription factors and digestive tract developmental genes. The vast majority of differentially expressed genes were up-regulated in this host. This gene set showed enrichment for genes associated with mitochondrial respiratory function and apoptosis, suggesting that mites on this host may be experiencing higher stress and may be less optimally adapted to parasitize it. Some genes involved in reproduction and oogenesis were also overexpressed, which should be further studied in regards to this host shift.

2.1 Background

Honey bees (*Apis mellifera* L.) are the most important insect for pollination of crops and wildflowers (Aizen & Harder, 2009; Ghazoul, 2005; Klein et al., 2007) but they have experienced increasing colony die-offs during the past two decades (Oldroyd, 2007; vanEngelsdorp et al., 2009; Williams et al., 2010). *Varroa destructor* is widely considered the most serious risk factor for honey bee colony mortality worldwide (Currie et al., 2010; Dahle, 2010; Guzmán-Novoa et al., 2010; Vanengelsdorp et al., 2008). These large ectoparasitic mites are associated with a condition known as parasitic mite syndrome, or PMS. When colonies exhibit PMS pathogens, including brood diseases and viruses are present at unusually high levels (Bowen-Walker et al., 1999; Rosenkranz et

al., 2010; Shimanuki & Knox, 1994). The open wound caused during feeding can allow microorganisms to enter and weaken the host (Bailey, 1981), and mites themselves are vectors for viruses and perhaps other bee pathogens. The V*arroa* mite's life cycle consists of two phases, the phoretic phase, during which the adult female mite lives on the adult bee, and the reproductive phase when the female mite reproduces inside the sealed brood cell of the pupating honey bee (Martin, 2001b). After a female mite invades the brood cell, the first egg laid will develop into a haploid male, which will later mate with his sisters to give rise to the next generation. *Varroa* mites feed on the hemolymph of the larva, pupa and adults. The most common Varroa-associated viral infection is deformed wing virus (DWV). The incidence of DWV is closely associated with mite infestation and colony mortality but other viruses such as acute bee paralysis virus have also been identified as part of the "parasitic mite syndrome" (Dainat et al., 2012; Martin, 2001b). Failure to treat colonies with miticides typically results in colony death within 1-3 years.

V. destructor was originally a parasite of the Asian honey bee, *Apis cerana*. At least 60 years ago *V. destructor* made a host switch and now parasitizes several European and African races of *A. mellifera* (Oldroyd, 1999). Population studies indicate that there was a genetic bottleneck associated with the host switch to *A. mellifera* (Navajas, 2010; Navajas et al., 2010; Solignac et al., 2005). These studies revealed that there is a remarkable absence of heterozygosity in the *Varroa* populations of Europe and USA collected on *A. mellifera* (Anderson & Trueman, 2000; Biasiolo, 1992; Kraus & Hunt, 1995). Furthermore, a study using microsatellite markers in 45 different populations of *Varroa*

mites from around the world showed a relative lack of polymorphisms within each of the two *V. destructor* mitochondrial haplotypes, Japan (J) and Korea (K), that successfully infest *A. mellifera* outside of Asia. These results suggested that these two haplotypes, J and K, each correspond to a single host capture event, followed by a rapid spread worldwide. These haplotypes also seem to be almost completely reproductively isolated from each other. Two routes of invasion of *V. destructor* into the Americas and specifically into the USA have been proposed based on the dates and places where they were first detected (Navajas, 2010; Navajas et al., 2010; Solignac et al., 2005). The J haplotype first shifted from *A. cerana* to *A. mellifera* in Japan during the last century following the introduction of *A. mellifera*. From Japan, it spread to Thailand, then to Paraguay in (1971), and then to Brazil in 1972, and later North America in 1987. The K haplotype first shifted from *A. cerana* to *A. mellifera* near Vladivostok (north of the Korean peninsula), following the introduction of *A. mellifera* from Ukraine in the 1950's. Later, it spread from eastern Russia to western Russia, then to Bulgaria in 1972, and Germany in 1977, and then continued spreading around Europe and also the U.S.

A sister species, *V. jacobsoni* is reportedly restricted to *A. cerana* and only reproduces on drone brood in this species. Similarly, a *Euvarroa* sp. infests *A. florea* colonies where it is restricted to reproducing on drone brood (D. Anderson, pers. comm.). Recently, a population of *V. jacobsoni* was found reproducing on *A. mellifera* drone and worker brood and was associated with colony mortality in Papua New Guinea, (Roberts et al. 201, in review). Evidence suggests that only a few such host shift events have occurred.

Varroa mites routinely invade sympatric non-host colonies and enter the worker brood but for some reason do not produce offspring, perhaps as a result of failure to recognize host signals to initiate reproduction. Since *V. destructor* has caused widespread losses wherever it has become established it is important to study the *V. jacobsoni* host switch to *A. mellifera* to determine whether host-parasite signaling may be involved and what those cues may be that are associated with mite reproduction.

We endeavored to study the transcriptome profile of *V. jacobsoni* from colonies where they were reproducing on *A. mellifera* and to compare their gene expression with that of *V. jacobsoni* from the source population that are still restricted to reproduction on *A. cerana*.

2.2 Methods

2.2.1 Sample collection

A total of nine samples of *V. jacobsoni* from Papua New Guinea (PNG) and Solomon Islands (Table 2.1) were collected from drone brood cells during April 2010. When collected, their reproductive status (reproducing or not reproducing) was recorded. *V. jacobsoni* reproducing on *A. mellifera* were collected from Goroka, PNG. Samples collected from St Christobel Island (SC) and Ugi Island in Solomon Islands were mites that were reproducing on *A. cerana*. In addition, single adult females from *A. mellifera* colonies on SC Island and Ugi Island in Solomon Islands were found not reproducing on either worker or drone brood. All samples were collected in RNAlater® and stored at 80 °C until RNA extraction.

2.2.2 RNA extraction and sequencing

Pools of adult female mites from each sample were ground in liquid nitrogen and total RNA from was extracted using the Invitrogen TRIzol® reagent protocol with one exception; the RNA precipitation step was slightly modified by the addition of 250µl of RNA precipitation solution (1.2 M NaCl + 0.8 M Sodium citrate dihydrate) mixed with 250µl of isopropanol to the aqueous phase of the mite homogenate to help precipitate more RNA. Approximately 20 mites per sample were used for extraction except for the non-reproducing mite sample. Only five non-reproducing mites were available for sequencing and RNA from all 5 mites was pooled. Total RNA per sample was then assessed for quality using a NanoDrop 2000/2000c (Thermosceintific) and submitted to the Purdue University Genomics Core Facility (PGCF) for sequencing. Total RNA was further analyzed for quality and concentration using an Agilent Technologies 2100 Bioanalyzer (Agilent Technologies, Inc. Santa Clara. CA). Seven out of nine cDNA libraries were prepared and barcoded by PGCF using the TruSeq™ RNA sample preparation kit (Illumina, Inc. San Diego, CA). These libraries were prepared and sequenced at two different time points (April 2012 and January 2013) using the Illumina platforms HiScanSQ (100 b paired reads in two lanes) and Hiseq2000 (100 bp paired reads in 4 lanes) respectively. Two of the nine cDNA libraries were prepared sequenced using a Hiseq2000 (100 b paired reads in one lane) at the Biomolecular Resource Facility (BRF), Canberra, Australia (February 2014). Raw sequence reads from all 9 samples were then processed.

2.2.3 Read pre-processing

Viral, bacteria, mitochondrial RNA and ribosomal RNA contaminants were removed using DeconSeq v 0.4.3 software (Schmieder & Edwards, 2011). Libraries containing sequences of these contaminants were created by downloading all available sequences in the corresponding categories from the NCBI database for use with DeconSeq. Initial analysis demonstrated an abundance of ribosomal RNA sequences and that these sequences were not fully represented in the library. Library content was readjusted and DeconSeq rerun until ribosomal RNA contaminants were reduced to less than two percent of the reads. Reads were checked for duplicates, and adapters removed using in-house Perl scripts. Sequence quality was assessed using FastQC software (v 0.10.0, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and quality trimming was performed using Trimmomatic v 0.30 (Bolger et al., 2014), trimmomaticSE -phred33 ILLUMINACLIP:adapters.fa:2:35:15 LEADING:7 TRAILING:7 SLIDINGWINDOW:4:13 MINLEN:30 (Table 2.1). Because, the reads were independently quality trimmed, some reads were unpaired after quality trimming and application of the minimum length cut off (30 bases).

2.2.4 Transcriptome assembly

The transcriptome assembly was performed using the paired and unpaired reads. Three different transcriptome assemblies were created using Trinity assembler software (trinity_beta_Jan28_2014) (Grabherr et al., 2011), and the Program to Assemble Spliced Alignments (PASA; v pasa_r20130907) (Haas et al., 2003). First, a *de novo* transcriptome assembly was created using Trinity default parameters (kmer length = 25

and min_contig_length 200 nucleotides). Then we created a genome guided assembly (Trinity --genome_guided_max_intron 11000); using as a reference a draft of the genome of *V. destructor* (Jay Evans personal communication, December 2013). For this assembly an *in silico* normalization of the full data set was performed using Trinity (normalize_by_kmer_coverage.pl --max_cov 50) in order to minimize the CPU running time. Finally, we created a hybrid transcriptome assembly using the *de novo* and genome-guided assemblies using PASA software (default parameters).

2.2.5 Description of a gene according to Trinity

For the *de novo* assembly Trinity reports many predicted transcripts, which are calculated by combining all the splice junctions observed in the data; some of these predicted isoforms are not observed. In the first stage of Trinity reads are clustered according to their sequence overlap into components. As the assembly progresses the components are further divided into subcomponents and predicted isoforms. The concept of a gene most closely matches the component level as determined by BLAST comparisons (data not shown). In the genome guided assembly genes are grouped according to their alignment to the reference genome. Each group then is independently assembled using the *de novo* Trinity assembly process. Again, the initial group of reads (Trinity component) most closely correspond to the concept of a gene. In the PASA hybrid assembly, *de novo* assemblies that do not match to the genome guided assembly are reported as with their original Trinity component IDs, and *de novo* assemblies that match the genome guided assembly are merged with the genome guided assembly and combined into PASA assembly clusters (genes) based on exon overlap. When considering the gene expression

analysis the term gene should be considered to mean a Trinity component or a PASA assembly cluster (gene).

2.2.6 Assessing quality of the assembly

To assess the quality of the final assembled transcripts, all RNAseq cleaned reads were aligned back to the hybrid assembly using bowtie2 (Langmead & Salzberg, 2012) and overall mapping statistics were estimated. In addition, to evaluate the completeness of the transcriptome assembly, the CEGMA (Core Eukaryotic Genes Mapping Approach, (Parra et al., 2007) software was applied to identify the presence of a core protein set consisting of 248 highly conserved proteins that are found in a wide range of eukaryotes.

2.2.7 Assembly Annotation

A comprehensive automated functional annotation of the final assembled transcripts was performed using Trinotate (Transcriptome Functional Annotation and Analysis, (Garber et al., 2011) and all the data derived from the analysis was entered into a user-friendly database. Trinotate makes use of a number of comprehensive annotation databases for functional annotation including homology searches of sequence data (NCBI-BLAST), protein domain identification (HMMER/PFAM), protein signal prediction (singalP/tmHMM), and comparison to other databases (EMBL UniProt/Swissprot eggNOG/GO pathways). To annotate the assembled transcripts, we also conducted a complete Blastx similarity search against the protein database UniProt/Swissprot (538,259 proteins sequences as of June 25, 2014) and against the western orchard

predatory mite *Metaseiulus occidentalis* predicted peptides (11,767 sequences/descriptions as of November 6, 2014), with an E-value cutoff of ≤ 1e-06.

2.2.8 Differential expression analysis

Quantification of the assembled transcripts was performed using standalone RSEM (Li & Dewey, 2011) which evaluates transcript abundances by mapping the RNAseq reads to the assembled transcriptome using the aligner tool bowtie2. Briefly, RSEM is a software that calculates posterior mean estimates, 95% credibility intervals, and maximum likelihood abundance estimates or expected counts (EC) for genes and predicted transcripts.

2.2.9 Identifying consistently differentially expressed mite genes CDEMG

Expected counts per gene per sample were combined into a matrix count, and this matrix was used as input for all downstream expression analyses. These analyses were performed using three different R packages EBSeq, EdgeR and DESeq2 (Leng et al., 2013; Love et al., 2014; Robinson et al., 2010). All differential expressed genes that were common among the three methods, using a False Discovery Rate (FDR) of 0.05, were extracted and used for downstream analyses. We will refer to these genes as consistently differentially expressed mite genes (CDEMG) in the rest of the manuscript.

EdgeR is a Bioconductor R package used to call differentially expressed genes from read counts obtained from RNAseq sequences (Robinson et al., 2010). EdgeR was used to normalize the EC (obtained from RSEM) for relative expression and effective library size

using the Trimmed Mean of M-values (TMM) normalization method. Genes with at least 0.18 counts per million (cpm), which corresponds to 14 read counts per gene in at least three samples were selected as target genes for further differential expression analysis. Furthermore, common and tag wise dispersions were estimated for the target genes. Differentially expressed genes (DEG) at FDR threshold of 0.05 and log fold change (logFC) of two were extracted.

The DESeq2 Bioconductor R package v. 1.0.19 (Love et al., 2014) was also used to call differentially expressed genes. DESeq2 implements a model based on negative binomial distribution and was run under R release 3.0.1. Before performing the DE analysis, DESeq2 automatically performs independent filtering of the genes with low counts (weakly expressed) in order to maximize the number of DEG with adjusted p values less than a critical value of 0.1. For the differential expression analysis gene-wise dispersion were estimated and DEG were extracted using a FDR threshold set to 0.05 and a logFC of two.

EBSeq v 1.4.0 is a Bioconductor R package that uses empirical Bayesian methods to identify differentially expressed genes (Leng et al., 2013). EBSeq estimates a posterior probability of being DE (PPDE). A list of DE genes with a FDR controlled at α was extracted using a PPDE value greater than $1 - \alpha$, where α was set to 0.05. By default EBSEq removes transcripts that have > 75% of the samples with expression less than 10 counts.

2.2.10 Heatmap and genes clustering

To generate the heatmap and gene clustering, we used the R packages EdgeR and heatmap.3 FPKM (Fragments Per Kilobase per Million) values obtained from RSEM for each of the CDEMG were normalized and log2 transformed prior to gene clustering. Afterwards, the CDEMG were clustered according to their patterns of differential expression (correlation distance) and complete linkage clustering. In addition, we used the Trinity script (define_clusters_by_cutting_tree.pl —-Ktree 5) to further differentiate the patterns of expression by partitioning the hierarchically clustered CDEMG tree into five groups. However, based on close observation of the pattern of expression, three of the five previously generated CDEMG clusters were manually regrouped into one, giving a total of 3 gene clusters used for further analysis. Plots of the expression patterns for the CDEMG were generated by modifying the Trinity script (plot_expression_patterns.pl).

2.2.11 GO enrichment analyses of the CDEMG

The assembled target genes (37,661, genes that passed the cpm cutoff used in EdgeR) were further analyzed using Blast2GO (Gotz et al., 2008) to assign gene ontology (GO) terms to each transcript. The xml output from the Blastx search against the NCBI protein database UniProt/Swissprot. We retained the highest hit for each gene with an E-value $\leq$ 1e-06. The xml output was used for Blast2GO and GO enrichment analysis was performed for each of the CDEMG cluster, using the target genes as the reference set and each individual gene cluster was used as the test set. A p-value cutoff of 0.1 was used for GO enrichment test instead of the FDR cutoff.

2.3 Results

2.3.1 *Varroa jacobsoni* assembled transcriptome

V. jacobsoni mite samples were collected from two different honey bee hosts, *A. cerana* and *A. mellifera* which correspond to two different geographic locations, the Solomon Islands and PNG, respectively (Table 2.1). A total of nine RNAseq libraries were constructed and sequenced using two Illumina sequencing platforms (Table 2.1), yielding a total of 2,184,624,960 raw paired reads (1,092,312,480 PE) (Table 2.2).

After filtering, a total of 591,878,383 (27%) contaminant reads were removed from the raw data set. Four contaminant libraries were created for this purpose and used with DeconSeq to remove the contaminants. The viral library contained a total of 30,300 sequences of complete genome viruses. A ribosomal RNA library contained 28,314 sequences including *V. destructor* 18S and 28S ribosomal RNA sequences while a bacterial library contained 2,451,824 complete genomic sequences and a mitochondrial library contained the complete sequence of *V. destructor* mitochondrial genome. Decontaminated reads were then subjected to adapter removal, a total of 26,780,048 (1.22%) reads with adapters were removed. Clean reads (contaminant-adapter free reads) were further checked with fastQC and in-house perl scripts to make sure that less than 2% of contaminant reads were present. Furthermore, a total of 36,132,637 (1.65%) reads with low quality were removed from the clean reads set, leaving a total of 1,529,833,892 (70%) reads that were used for the transcriptome assemblies (Table 2.2).

Three different transcriptome assemblies were created using Trinity/PASA. The first one, a *de novo* assembly, produced a total of 374,530 putative transcripts (252,445 putative

genes, N50 = 3,406 bp). The second one a genome-guided assembly produced a total of 428,912 putative transcripts and 155,121 putative "genes" as defined by Trinity (N50 = 6,266 bp). We used as a reference the draft genome of *V. destructor* (28,777 putative scaffolds, N50: 125,895 bp, unpublished data). Finally, a third hybrid assembly was created using PASA, which combined the assembled transcripts from the *de novo* and the genome-guided assemblies, which produced a total of 319,231 putative transcripts (223,620 putative genes, N50 = 3,549 bp; Figure 2.1).

2.3.2 Assessing the quality of the assembly

The quality and completeness of our hybrid *V. jacobsoni* transcriptome assembly was assessed in three different ways: using CEGMA, by comparison with predicted gene sequences of the predatory mite *M. occidentalis*, and by aligning back the clean reads to the hybrid assembly.

Analysis of our hybrid assembly against the CEGMA protein set identified 246 out of 248 core proteins (99.19%) as complete (defined as >70% alignment length versus the core protein). Furthermore, there was an average of about 3 *V. jacobsoni* assembled transcripts aligning with each core protein, with 221 of those detected having more than 1 alignment (Table 2.3). We compared the hybrid assembly against the UniProt/Swissprot database using Blastx (hits with E-value $\leq$ 1e-06) and we identified 4,957 proteins represented by nearly full-length transcripts, having > 80% alignment coverage, and 8,372 proteins are > 50% alignment coverage. In addition, we compared the hybrid assembly against the *M. occidentalis* predicted petides, using Blastx (E-value $\leq$ 1e-06)

and found 5,362 proteins that are represented by nearly full-length transcripts, having > 80% alignment coverage, and 7,063 proteins are > 50% alignment coverage, which represents 60% of the total *M. occidentalis* predicted peptides. Clean reads ranging from 5.5 to 282.5 millions for each sample were mapped back to the hybrid reference assembly using Bowtie2. Overall 95.63% of reads aligned to the reference indicating that almost all reads are represented in the assembly (Table 2.2).

In summary, these four analyses suggest that our *V. jacobsoni* transcriptome assembly contains a good representation of the CEGs and that the assembly is a fair representation of the mites' gene expression potential.

2.3.3 Annotation of the assembly at transcript and gene level

The hybrid transcriptome assembly of *V. jacobsoni* was used to query entries described in the UniProt/Swissprot protein database, using Blastx (E-value ≤1e-06) and retaining only the most significant query blast hit for each database peptide. At the transcript level we found that 51,025 (~16%) out of 319,231 transcripts have the best match to a protein sequence and 2,870 (6%) of those matches have a sequence identity ≥ 90%.

At the gene level we found 24,128 out of 223,620 putative genes have the best match to a protein sequence and 2,413 (10%) of them have a sequence identity ≥ 90%. Furthermore, when we compared the hybrid transcriptome assembly of *V. jacobsoni* against the *M. occidentalis,* which contains 11,767 unique predicted peptide annotations, we found that 23,779 (10%) of the *V. jacobsoni* genes, had the most significant alignment with a protein

sequence in the predatory mite. However, these 23,779 blast hits covered 8,388 (~71%) of the predicted peptide descriptions of the predatory mite.

2.3.4 Comparing Expression profiling between mites reproducing in *A. cerana* host versus the mites reproducing in the *A. mellifera* host

Quantification of the assembled transcripts was performed using standalone RSEM and the representation of putative genes from each of the 8 mite samples in the assembled hybrid transcriptome was evaluated in terms of expected counts (EC). Transcript abundances were evaluated by mapping the RNAseq clean reads to the assembled hybrid transcriptome using the aligner tool bowtie2. For all samples, we looked at the overall distribution of the gene expression profile. The variation about the median for normalized FPKM values across all samples was uniform, ranging from -0.458 to -2.965 (see Additional file 1; Figure S1, histograms distribution).

To identify differentially expressed genes we used three different R packages. The EC of only eight of out the nine mites samples were used to create the matrix of EC. We chose not to include the Am-non-reproductive mite sample, because we did not have a biological replicate and the reproductive status of the adult females was not the same as the other eight samples. In addition, the amount of reads obtained during sequencing was not as high as the rest of samples (Table 2.2).

2.3.5 Differential expression analysis of the mites reproducing in *A. cerana* host versus

the mites reproducing in the *A. mellifera* host

We used EBESeq, EdgeR and DESeq2, to quantify expression and to identify differentially expressed genes (DEG). The expression analysis was performed at the gene level (223,620 putatives genes in the count matrix). Using EdgeR we removed genes with cpm < 0.18 and selected a total of 37,661 target genes for the expression analysis. EdgeR identified 1013 differentially expressed genes (FDR < 0.05 and absolute logFC ≥ 2), for complete genes list (see Additional file 2; Table S1). In addition, using DESeq2 and EBSeq we identified a total of 586 and 6809 DEG (FDR < 0.05), respectively. For a complete gene list (see Additional file 3; Table S2 and Table S3, respectively). Finally, we combined the results of the three methods, and the 287 CDEMG common to all three methods, were extracted (Figure 2.1). Out the 287 CDEMG, we found a total of 23 down-regulated genes and 264 up-regulated genes in the mites reproducing in the *A. mellifera* host compared to those reproducing in the *A. cerana* host.

We cluster the samples by using correlation distances and complete linkage clustering, which grouped the samples according to the pattern of expression using the 287 CDEMG. Originally five gene clusters were generated, however, after visual inspection of the pattern of expression we manually clustered the CDEMG into three groups (Figure 2.2). Cluster one contains 23 CDEMG that were down-regulated in the *A. mellifera* host. Cluster two contains 208 genes and cluster three contains 56 genes that were all up-regulated in the mites reproducing on the *A. mellifera* host. These 264 genes were separated in two cluster groups, due to clear differences in their pattern of expression. For

example for cluster two we observed that four samples (three *A. cerana* samples plus one *A. mellifera* sample) showed consistent low expression patterns across all 208 genes, while the other four remaining *A. mellifera* samples showed higher expression patterns.

2.3.6 GO terms assignments and Enrichments analysis of the CDEMG

We used Blast2GO to assign GO terms to the 287 CDEMG and to test whether certain biological functions or GO terms are more frequently observed in either of the two mite groups, we used the Fisher's exact test in Blast2GO to compare the GO terms of the CDEMG in each of the gene clusters versus the target genes (reference set of 37,661 genes) each gene was represented by the highest Blastx hit and a p-value 0.1 was used for the Fisher's exact test, see (Additional file 4; Table S4).

2.3.7 Cluster 1 CDEMG down-regulated in *A. mellifera*

We found 23 CDEMG down-regulated in the *A. mellifera* host. However, only two out of these 23 genes had a GO term associated with them. Visual inspection of the GO terms associated with these genes are related to digestive tract development and transcription factors (Table 2.4). For the full report of all the GO terms and the 36 unique GO-ID associated with genes see (Additional file 5: Table S5). Furthermore, only 5 out of the 23 genes had a significant Blastx similarity hit (E-value $\leq$ 1e-06) to the *M. occidentalis* predicted peptideds (see Additional file 6: Table S6) and only 3 out the 23 genes had significant Blastx similarity hits (E-value $\leq$ 1e-06) to the UniProt/Swissprot database. It make sense that mites feeding in a suboptimal host, will show differences in digestive tract development and this might be mediated by transcriptional regulation.

2.3.8 Cluster 2 and 3 CDEMG up-regulated in *A. mellifera*

We found 208 CDEMG up-regulated in *A. mellifera* contained in cluster 2. However, only eight out of these 208 genes had an associated GO term. The top 10 enriched GO terms associated with these genes are primarily involved in either oxidative metabolism and stress (mitochondrial respiratory chain complex, oxidoreductase complex) or in development and reproduction (developmental process involved in reproduction, germ cell development, establishment of endothelial barrier, cis-Golgi network, post-embryonic organ development, cellular process involved in reproduction; Table 2.5). For a full report of all the GO terms and the 45 unique GO-ID associated with genes see (Additional file 7: Table S7). Only 80 out of the 208 genes had a significant Blastx similarity hits (E-value ≤ 1e-06) to the *M. occidentalis* predicted peptides (Additional file 6: Table S8) and only 88 out the 208 genes had significant Blastx similarity hits (E-value ≤ 1e-06) to the UniProt/Swissprot database.

We found 56 CDEMG up-regulated in *A. mellifera* contained in cluster 3. However, only two out of these 56 genes had a GO term associated with them. Visual inspection of the GO terms associated with these genes are related to either apoptosis (Bcl-2 family protein complex and B cell apoptotic process, BH-domain binding), or the following terms: epoxide hydrolase activity, leukotriene metabolic process, ether hydrolase activity and Type I pneumocyte differentiation (Table 2.5). For a full report of all the GO terms and the 51 unique GO-IDs associated with genes see (Additional file 8: Table S9). Furthermore, only 15 out of the 56 genes had a significant Blastx similarity hits (E-value ≤ 1e-06) to the *M. occidentalis* predicted peptides (Additional file 6: Table S10) and only

12 out the 56 genes had significant Blastx similarity hits (E-value $\leq$ 1e-06) to the UniProt/Swissprot database.

2.4 Discussion

2.4.1 Functions of CDEMG genes

An analysis of the functions of the genes that were differentially expressed (DE) between *V. jacbosoni* mites that differed in their ability to parasitize European honey bees revealed several trends. One obvious trend is that most of the DE transcripts were higher expressed in mites using *A. mellifera* as a host. Only 23 mite transcripts were down-regulated for mites on this host. These included genes coding Proteins with RNAII polymerase promoter region specific DNA binding activity, as well as genes involved in digestive tract development. These results suggest a down-regulation of transcription factors and perhaps transcription in general.

The two larger classes of genes that were higher expressed in mites on the *A. mellifera* host contained nuclear encoded mitochondrial genes and genes involved in metabolic regulation and apoptosis. Included in this broad category were genes encoding 8 mitochondrial proteins, a heatshock protein and the conserved NAD+ sensing histone deacetylase SIRT6 that regulates glucose homeostasis in mammals (Zhong et al., 2010). Other higher-expressed genes that have roles in cellular primary metabolism included 6-phosphofructo kinase, a coordinator of glucose metabolism and cell cycle, phospholipase A2 activating protein involved in calcium/CaMKII signaling, and phosphodiesterase 8A homolog, a regulator of cyclic AMP levels (Conti et al., 2014; Doroudi et al., 2014; Yalcin et al., 2014). Up-regulation of genes in involved in primary metabolism may be a

result of stress induced in the mites from existing on a suboptimal host. Stress induced by a number of treatments in *Drosophila* results in increased expression of mitochondrial and heat shock genes (Brown et al., 2014). Interestingly genes putatively involved in reproductive development and growth were also higher expressed on this host. For example a transcript with highly significant alignment to *Drosophila* Src64 was over expressed. This gene encodes a tyrosine kinase that is required for Drosophila oogenesis and affects insulin signaling through interactions with the transcription factor dFOXO (Bulow et al., 2014; Djagaeva et al., 2005).

Our samples come from populations that differ in their ability to parasitize *A. mellifera* but not only were they exposed to different host colony environments, they also were geographically separated, perhaps confounding our differential expression analyses. However, other analyses indicate that our samples of mites parasitizing *A. mellifera* are likely derived from the same source population as the mites we collected parasitizing *A. cerana* (Roberts et al., submitted) and colony environments are buffered from external climactic conditions. These differential expression analyses provide a valuable resource for future studies into the mechanisms involved in this singular host shift to European honey bees. Discovering why mated mites fail to lay eggs upon entering brood cells of different honey bee host species is critical to our understanding of this devastating pest species, and for predicting the ability of *Varroa* mites to successfully make a host switch to *A. mellifera*.

References

1. Aizen MA & Harder LD (2009) The global stock of domesticated honey bees is growing slower than agricultural demand for pollination. Current biology : CB 19: 915. doi:10.1016/j.cub.2009.03.071.

2. Anderson DL & Trueman JW (2000) *Varroa jacobsoni* (Acari: Varroidae) is more than one species. Exp Appl Acarol 24: 165-189.

3. Bailey L (1981) Honey bee pathology. London ; New York : Academic Press, London ; New York, pp. 124.

4. Biasiolo A (1992) Lack of allozyme variability amongVarroa mite populations. Exp Appl Acarol 16: 287-294.

5. Bolger AM, Lohse M & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120. doi:10.1093/bioinformatics/btu170.

6. Bowen-Walker PL, Martin SJ & Gunn A (1999) The transmission of deformed wing virus between honeybees (*Apis mellifera* L.) by the ectoparasitic mite *Varroa jacobsoni* Oud. J Invertebr Pathol 73: 101-106. doi:10.1006/jipa.1998.4807.

7. Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, Booth BW, Wen J, Park S, Suzuki AM, Wan KH, Yu C, Zhang D, Carlson JW, Cherbas L, Eads BD, Miller D, Mockaitis K, Roberts J, Davis CA, Frise E, Hammonds AS, Olson S, Shenker S, Sturgill D, Samsonova AA, Weiszmann R, Robinson G, Hernandez J, Andrews J, Bickel PJ, Carninci P, Cherbas P, Gingeras TR, Hoskins RA, Kaufman TC, Lai EC, Oliver B, Perrimon N, Graveley BR & Celniker SE (2014) Diversity and dynamics of the Drosophila transcriptome. Nature. doi:10.1038/nature12962.

8. Bulow MH, Bulow TR, Hoch M, Pankratz MJ & Junger MA (2014) Src tyrosine kinase signaling antagonizes nuclear localization of FOXO and inhibits its transcription factor activity. Scintific Reports 4. doi:10.1038/srep04048.

9. Conti M, Mika D & Richter W (2014) Cyclic AMP compartments and signaling specificity: Role of cyclic nucleotide phosphodiesterases. J. Gen. Physiol. 143: 29-38. doi:10.1085/jgp.201311083.

10. Currie RW, Pernal SF & Guzmán-Novoa E (2010) Honey bee colony losses in Canada. Journal of Apicultural Research 49: 104-106.

11. Dahle B (2010) role of *Varroa destructor* for honey bee colony losses in Norway. Journal of Apicultural Research.

12. Dainat B, Evans JD, Chen YP, Gauthier L & Neumann P (2012) Dead or alive: deformed wing virus and *Varroa destructor* reduce the life span of winter honeybees. Appl Environ Microbiol 78: 981-987. doi:10.1128/AEM.06537-11.

13. Djagaeva I, Doronkin S & Beckendorf SK (2005) Src64 is involved in fusome development and karyosome formation during Drosophila oogenesis. Developmental Biology 284: 143-156. doi:10.1016/j.ydbio.2005.05.012.

14. Doroudi M, Plaisance MC, Boyan BD & Schwartz Z (2014) Membrane actions of 1α, 25(OH) 2D3 are mediated by Ca2+/ calmodulin- dependent protein kinase II in bone and cartilage cells. Journal of Steroid Biochemistry and Molecular Biology 145: 65-74. doi:10.1016/j.jsbmb.2014.09.019.

15. Garber M, Grabherr MG, Guttman M & Trapnell C (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods 8: 469-477. doi:10.1038/nmeth.1613.

16. Ghazoul J (2005) Buzziness as usual? Questioning the global pollination crisis. Trends Ecol Evol 20: 367-373. doi:10.1016/j.tree.2005.04.026.

17. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J & Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36: 3420-3435. doi:10.1093/nar/gkn176.

18. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N & Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29: 644-652. doi:10.1038/nbt.1883.

19. Guzmán-Novoa E, Eccles L, Calvete Y, McGowan J, Kelly PG & Correa-Benítez A (2010) *Varroa destructor* is the main culprit for the death and reduced populations of overwintered honey bee (*Apis mellifera*) colonies in Ontario, Canada. Apidologie 41: 443-450. doi:10.1051/apido/2009076.

20. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, Salzberg SL & White O (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res 31: 5654-5666. doi:Doi 10.1093/Nar/Gkg770.

21. Klein AM, Vaissiere BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C & Tscharntke T (2007) Importance of pollinators in changing landscapes for world crops. Proc Biol Sci 274: 303-313. doi:10.1098/rspb.2006.3721.

22. Kraus B & Hunt G (1995) Differentiation of *Varroa jacobsoni* Oud populations by random amplification of polymorphic DNA (RAPD). Apidologie 26: 283-290.

23. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-359. doi:10.1038/nmeth.1923.

24. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BMG, Haag JD, Gould MN, Stewart RM & Kendziorski C (2013) EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics 29: 1035-1043. doi:10.1093/bioinformatics/btt087.

25. Li B & Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 323. doi:10.1186/1471-2105-12-323.

26. Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2.

27. Martin SJ (2001) The role of Varroa and viral pathogens in the collapse of honeybee colonies: a modelling approach. Journal of Applied Ecology 38: 1082-1093. doi:Doi 10.1046/J.1365-2664.2001.00662.X.

28. Navajas M (2010) Tracking the colonisation history of the invasive species *Varroa destructor*: Trends in Acarology (ed. by MW Sabelis & J Bruin) Springer Netherlands, pp. 375-378.

29. Navajas M, Anderson DL, de Guzman LI, Huang ZY, Clement J, Zhou T & Le Conte Y (2010) New Asian types of *Varroa destructor*: a potential new threat for world apiculture. Apidologie 41: 181-193. doi:10.1051/apido/2009068.

30. Oldroyd BP (1999) Coevolution while you wait: *Varroa jacobsoni*, a new parasite of western honeybees. Trends Ecol Evol 14: 312-315.

31. Oldroyd BP (2007) What's killing American honey bees?, Vol. 5, pp. 1195-1199.

32. Parra G, Bradnam K & Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23: 1061-1067. doi:10.1093/bioinformatics/btm071.

33. Robinson MD, McCarthy DJ & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140. doi:10.1093/bioinformatics/btp616.

34. Rosenkranz P, Aumeier P & Ziegelmann B (2010) Biology and control of *Varroa destructor*. J Invertebr Pathol 103 Suppl 1: S96-119. doi:10.1016/j.jip.2009.07.016.

35. Schmieder R & Edwards R (2011) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One 6: e17288. doi:10.1371/journal.pone.0017288.

36. Shimanuki H & Knox DA (1994) Parasitic mite syndrome: the symptoms. American Bee Journal 134: 827-828.

37. Solignac M, Cornuet JM, Vautrin D, Le Conte Y, Anderson D, Evans J, Cros-Arteil S & Navajas M (2005) The invasive Korea and Japan types of *Varroa destructor*, ectoparasitic mites of the Western honeybee (*Apis mellifera*), are two partly isolated clones. Proc Biol Sci 272: 411-419. doi:10.1098/rspb.2004.2853.

38. vanEngelsdorp D, Evans JD, Saegerman C, Mullin C, Haubruge E, Nguyen BK, Frazier M, Frazier J, Cox-Foster D, Chen YP, Underwood R, Tarpy DR & Pettis JS (2009) Colony Collapse Disorder: A Descriptive Study. PLoS One 4. doi:ARTN e6481. DOI 10.1371/journal.pone.0006481.

39. Vanengelsdorp D, Hayes J, Underwood RM & Pettis J (2008) A Survey of Honey Bee Colony Losses in the US, Fall 2007 to Spring 2008. PLoS One 3. doi:ARTN e4071. DOI 10.1371/journal.pone.0004071.

40. Williams GR, Tarpy DR, Vanengelsdorp D, Chauzat MP, Cox-Foster DL, Delaplane KS, Neumann P, Pettis JS, Rogers REL & Shutler D (2010) Colony Collapse Disorder in context. Bioessays 32: 845-846. doi:10.1002/bies.201000075.

41. Yalcin A, Clem BF, Imbert-Fernandez Y, Ozcan SC, Peker S, Neal J, Klarer AC, Clem AL, Telang T & Chesney J (2014) 6- Phosphofructo- 2- kinase (PFKFB3) promotes cell cycle progression and suppresses apoptosis via Cdk1- mediated phosphorylation of p27. Cell Death and Disease 5. doi:10.1038/cddis.2014.292.

42. Zhong L, Urso A, Toiber D, Sebastian C, Henry RE, Vadysirisack DD, Guimaraes A, Marinelli B, Wikstrom JD, Nir T, Clish CB, Vaitheesvaran B, Iliopoulos O, Kurland I, Dor Y, Weissleder R, Shirihai OS, Ellisen LW, Espinosa JM & Mostoslavsky R (2010) The Histone Deacetylase Sirt6 Regulates Glucose Homeostasis via Hif1α. Cell 140: 280-293. doi:10.1016/j.cell.2009.12.041
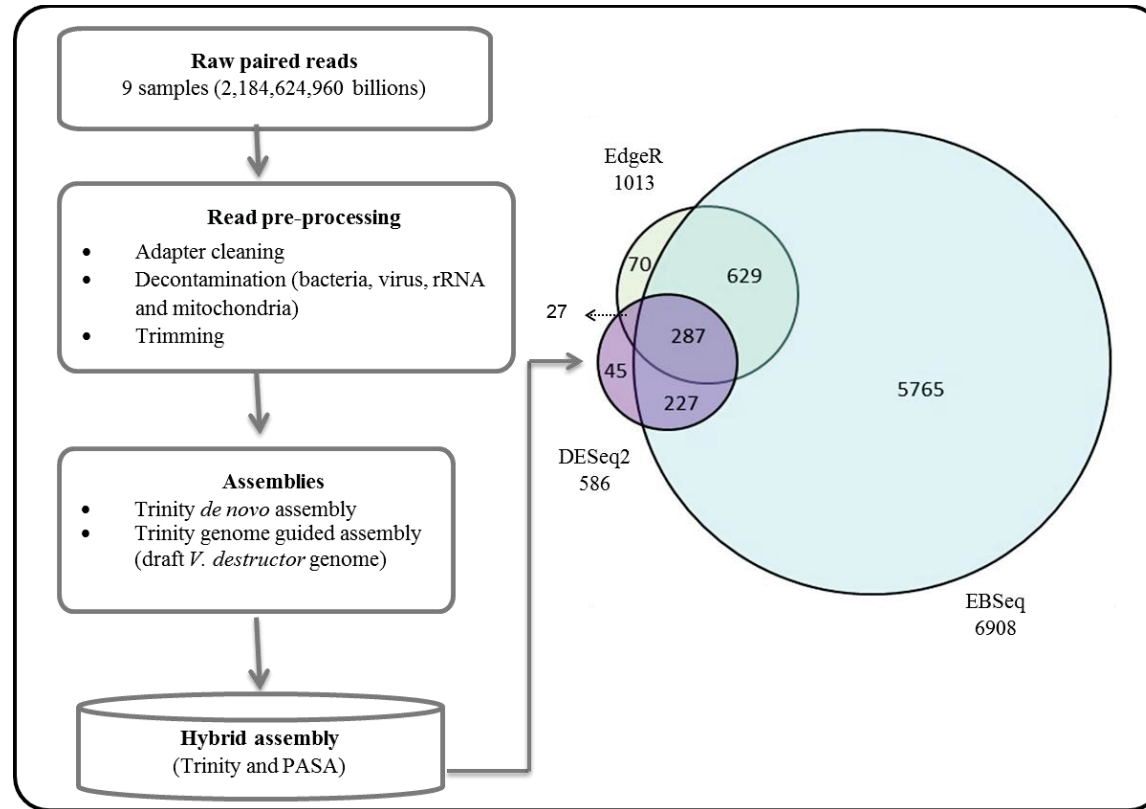
Figure 2.1 Transcriptome assembly and differentially expressed genes

Pipeline steps followed to build the assembly and expression profiles using 3 different R packages. Flow chart shows the steps implemented from raw reads to the selection of the final assembly and the selection of the consistently differentially expressed mite genes (CDEMG).
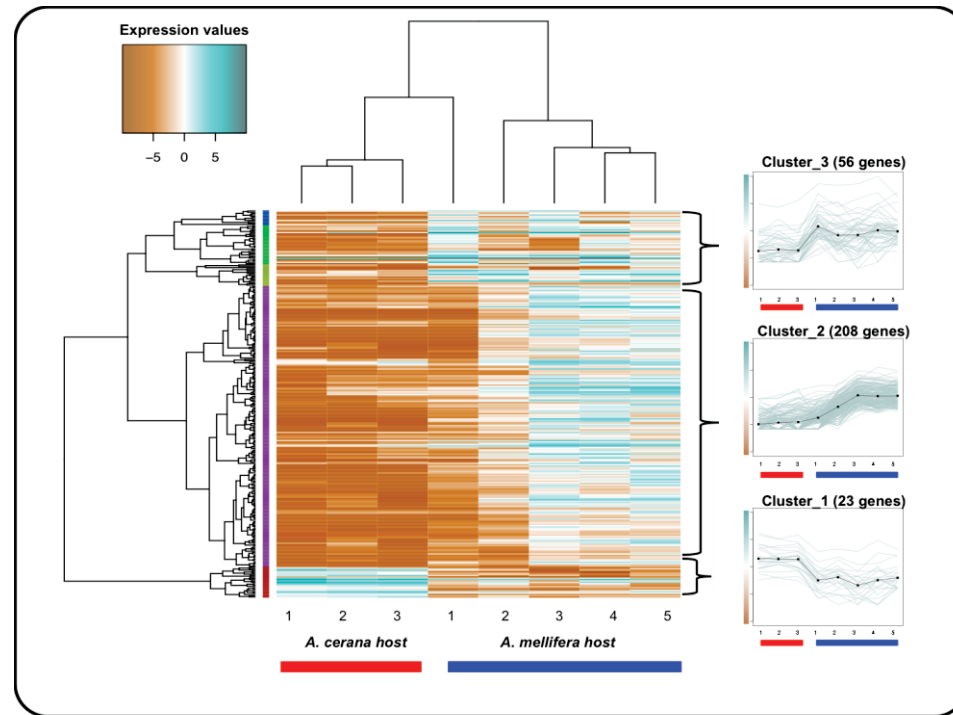
Figure 2.2 Heatmap and gene clusters of CDEMG genes for *V. jacobsoni* mites

Heatmap of expression values (log2 transformed normalized FPKM) of the CDEMG adult female *V. jacobsoni* mites reproducing in *A. cerana* and *A. mellifera*. Orange and turquoise blue indicate higher and lower expression values, respectively. Red and blue tick bars indicate the *A. cerana* host and *A. mellifera* host respectively.

Table 2.1 Description of *V. jacobsoni* RNA samples

Bee host	Reproductive status	Collection	Collection sites	Year of sequencing
A. cerana	Reproducing	Drone cells	SC*, Solomon Islands	Apr 2012 (HiScanSQ)
A. cerana	Reproducing	Drone cells	Ugi, Solomon Islands	Jan 2013 (Hiseq2000)
A. cerana	Reproducing	Drone cells	Guadalcanal, Solomon Islands	Jan 2013 (Hiseq2000)
*A. mellifera***	Non-reproducing	Drone and worker cells	SC and Ugi (Solomon Islands)	Apr 2012 (HiScanSQ)
A. mellifera	Reproducing	Drone cells	Goroka, Papua New Guinea	Apr 2012 (HiScanSQ)
A. mellifera	Reproducing	Drone cells	Goroka, Papua New Guinea	Jan 2013 (Hiseq2000)
A. mellifera	Reproducing	Drone cells	Goroka, Papua New Guinea	Jan 2013 (Hiseq2000)
A. mellifera	Reproducing	Drone cells	Goroka, Papua New Guinea	Jan 2014 (Hiseq2000)
A. mellifera	Reproducing	Drone cells	Goroka, Papua New Guinea	Jan 2014 (Hiseq2000)

*SC = San Cristobel, Salomon Islands

** Non-reproducing, individual adult females were pooled together expecting to get more RNA for sequencing.

Table 2.2 Sequencing reads and mapping summary

Sample-ID	Raw reads	Contaminants	Adapters	Trimmed reads	Clean reads	Mapped reads
Ac-reproductive	154854698	31372885	5868855	2684238	114928720	109400649 (95.19%)
Ac-reproductive	376336622	96168948	3863069	5785505	270519100	259373713 (95.88%)
Ac-reproductive	460610232	167624944	3759060	6705901	282520327	269524392 (95.40%)
Am-not reproductive	10427368	2019087	2294687	535882	5577712	5306077 (95.13%)
Am-reproductive	146287844	27746078	871943	1942313	115727510	111179419 (96.07%)
Am-reproductive	203052598	30479539	1337684	3330200	167905175	161793427 (96.36%)
Am-reproductive	209363152	44563144	1102128	2502475	161195405	153861014 (95.45%)
Am-reproductive	264092696	91851166	4377311	2797800	165066419	157324804 (95.31%)
Am-reproductive	303036016	79679461	2046179	4996895	216313481	206579374 (95.50%)
Undetermined*	56563734	20373131	1259132	4851428	30080043	28693353 (95.39%)
Total reads	**2184624960**	**591878383**	**26780048**	**36132637**	**1529833892**	**1463036222**

*Reads where the barcode could not be decoded. The order of the sample-ID is the same as in figure 2.1.

Table 2.3 Completeness of the *V. jacobsoni* transcriptome based on 248 CEGs

	# Prots*	% Completeness**	# Total***	Average[§]	% Ortho[§§]
Complete[δ]	246	99.19	807	3.28	89.84
Group 1	66	100.00	230	3.48	90.91
Group 2	56	100.00	196	3.5	91.07
Group 3	60	98.36	182	3.03	85.00
Group 4	64	98.46	199	3.11	92.19
Partial[ξ]	248	100.00	967	3.9	98.39
Group 1	66	100.00	271	4.11	96.97
Group 2	56	100.00	229	4.09	100.00
Group 3	61	100.00	221	3.62	98.36
Group 4	65	100.00	246	3.78	98.46

These results are based on the set of genes selected by Genis Parra

[*] Prots = number of 248 ultra-conserved CEGs present in genome.

[**] %Completeness = percentage of 248 ultra-conserved CEGs present.

[***] Total = total number of CEGs present including putative orthologs.

[§] Average = average number of orthologs per CEG.

[§§] %Ortho = percentage of detected CEGS that have more than 1 ortholog.

[δ]Complete = refers to those predicted proteins in the set of 248 CEGs that when aligned to the HMM for the KOG for that protein-family, give an alignment length that is 70% of the protein length.

[ξ] Partial = If a protein is not complete, but if it still exceeds a pre-computed minimum alignment score.

Table 2.4 Most specific GO terms related to mite genes that are down-regulated in the *A. mellifera* host, cluster 1

# Genes	GO-ID	Term	Category	P-Value	Am-down seq. count*	Ref seq. count**
1	GO:0060237	regulation of fungal-type cell wall organization	P	0.000611	1	1
2	GO:0000978	RNA polymerase II core promoter proximal region sequence-specific DNA binding	F	0.000102	2	56
	GO:0000987	core promoter proximal region sequence-specific DNA binding	F	0.000181	2	75
	GO:0001159	core promoter proximal region DNA binding	F	0.000195	2	78
	GO:0048546	digestive tract morphogenesis	P	0.000440	2	118
	GO:0003705	RNA polymerase II distal enhancer sequence-specific DNA binding transcription factor activity	F	0.000549	2	132
	GO:0000977	RNA polymerase II regulatory region sequence-specific DNA binding	F	0.000697	2	149
	GO:0001012	RNA polymerase II regulatory region DNA binding	F	0.000782	2	158
	GO:0048565	digestive tract development	P	0.001380	2	211
	GO:0055123	digestive system development	P	0.001610	2	228

Fisher's exact test showing enriched GO terms in mite genes that are down-regulated in *A. mellifera* host (cluster 1). For a complete list see (Additional file 5: Table S5). * 23 genes in test set ** number of times the GO was identified in reference set of 37,661 genes.

Table 2.5 Most specific GO in mite genes that are up-regulated in the *A. mellifera* host, cluster 2 and 3

		Cluster 2 (208 CDEMG)				
# Genes	GO-ID	Term	Category	P-Value	Am-Up seq. count*	Ref seq. count**
4	GO:0005746	mitochondrial respiratory chain	C	0.00198	4	61
2	GO:0016272	prefoldin complex	C	0.00182	2	6
24	GO:0003006	developmental process involved in reproduction	P	0.00016	24	1332
2	GO:0010029	regulation of seed germination	P	0.00233	2	7
14	GO:0007281	germ cell development	P	0.00119	14	665
5	GO:1990204	oxidoreductase complex	C	0.000855	5	85
3	GO:0061028	establishment of endothelial barrier	P	0.000476	3	16
3	GO:0005801	cis-Golgi network	C	0.00169	3	26
11	GO:0048569	post-embryonic organ development	P	0.00254	11	489
18	GO:0048610	cellular process involved in reproduction	P	0.00267	18	1065
		Cluster 3 (56 CDEMG)				
1	GO:0097136	Bcl-2 family protein complex	C	0.0016	1	1
	GO:0051400	BH domain binding	F	0.0057	1	6
	GO:0001783	B cell apoptotic process	P	0.0089	1	10
1	GO:0004301	epoxide hydrolase activity	F	0.0033	1	3
	GO:0004463	leukotriene-A4 hydrolase activity	F	0.0041	1	4
	GO:0060509	Type I pneumocyte differentiation	P	0.0049	1	5
	GO:0019370	leukotriene biosynthetic process	P	0.0057	1	6
	GO:0016803	ether hydrolase activity	F	0.0073	1	8
	GO:0016801	hydrolase activity, acting on ether bonds	F	0.0097	1	11
	GO:0006691	leukotriene metabolic process	P	0.0138	1	16

Fisher's exact test showing enriched GO terms in mite genes that are up-regulated in *A. mellifera* host (cluster 2 and 3). For a complete list see (Additional file 7: Table S7; Additional file 8: Table S9). * 208 and 56 genes in each test set, respectively.

** number of times the GO was identified in reference set of 37,661 genes.

CHAPTER 3. SURVEY OF VIRUSES IN

VARROA JACOBSONI MITES

Abstract

Mite infestations of colonies are associated with a number of brood diseases, but viruses are the pathogens most often associated with colony mortality. We have assembled a virus transcriptome of *Varroa jacobsoni* to provide the first survey of pathogens in this species. Among the list of putative viruses are Deformed wing virus, Dragonfly cyclovirus 1, Farmington virus, Formica exsecta virus 2, Halyomorpha halys virus, Heliconius erato iflavirus, Kakugo virus, Kirsten murine sarcoma virus, Sacbrood virus, Spodoptera exigua iflavirus 1. A search against a honey bee associated microbe database revealed the likely presence of Macula-like virus (Tymoviridae), a microsporidan and a spiroplasma previously reported from *A. mellifera*. This is the first report of deformed wing virus in *V. jacobsoni.* Simlar viruses were found in mites reproducing on *A. mellifera* host that were collected from Goroka, PNG and also found in mite samples reproducing on *A. cerana* collected from Solomon Islands.

3.1 Background

The honey bee *Apis mellifera* is the most important pollinator of agricultural crops worldwide, but there has been an alarming global increase in annual colony mortality over the past few decades. The parasitic mite, *Varroa destructor,* is perhaps the prime

suspect as an important contributor to these colony losses (Le Conte et al., 2010; Neumann & Carreck, 2010). The mite's life cycle consists of two phases, the phoretic phase, during which the adult female mite lives on the adult bee, and the reproductive phase when the female mite reproduces inside the sealed brood cell of the pupating honey bee (Martin, 2001b). After a female mite invades the brood cell, the first egg laid will develop into a haploid male, which will later mate with his sisters to give rise to the next generation.

Mite infestations of colonies are associated with a number of brood diseases, but viruses are the pathogens most often associated with colony mortality (Francis et al., 2013b; vanEngelsdorp et al., 2009). Approximately twenty honey bee viruses have been discovered and it was not until the 1980's when *V. destructor* become widely spread that the presence of viruses stopped being considered as relatively harmless. *V. destructor* mites have been described to act as a physical and or biological vector (Kevan et al., 2006). Viruses associated with Varroa infestation include two species complexes; Israeli acute paralysis virus (IAPV), acute bee paralysis virus (ABPV) and Kashmir bee virus (KBV) constitute one complex (de Miranda et al., 2010) and deformed wing virus (DWV) and Varroa destructor virus-1 (VDV-1) constitute the other complex (de Miranda & Genersch, 2010). In addition, black queen cell virus (BQCV) and sacbrood virus (SBV) are also commonly found in parasitized bees (Cox-Foster et al., 2007a). High populations of *V. destructor* within the colonies often result in a high incidence of viral infections among the bees and possibly a compromised immune system, which can result in increased susceptibility to other brood diseases. This condition is known as parasitic

mite syndrome (PMS) (Shimanuki & Knox, 1994) and usually kills colonies within months to years after mite infestation (Dainat et al., 2012; Martin, 2001b).

The viruses that are most closely connected to *Varroa* parasitism are those that can replicate within the mites and be transmitted by mites, such as the IAPV/ABPV/KBV and DWV/VDV-1 species complexes. At least for these viruses the mites can act as a reservoir and a host for the virus, and their feeding behavior can inject the virus into the hemolymph of their hosts. The distribution and abundance of DWV is strongly correlated with the presence and spread of *V. destructor*. DWV was rare or undetectable in Europe, New Zealand and the island of Hawaii prior to the introduction of *Varrroa* mites and spread to virtually 100% of colonies with establishment of the *Varrroa* population (Martin et al., 2012; Mondet et al., 2014a; Rosenkranz et al., 2010). In New Zealand levels of BQCV and KBV tracked the spread of Varroa infestations but DWV titers continued to increase with the duration of infestation. This was interpreted as a result of increased DWV titers in the mites themselves (Mondet et al., 2014a). DWV is believed to be the primary cause of colony collapse, rather than the mites themselves (de Miranda & Genersch, 2010; Genersch et al., 2010; Highfield et al., 2009). However in some cases ABPV has been more closely associated with colony losses (Nguyen et al., 2010).

It is possible that viral replication, and or mite feeding wounds in developing pupae could suppress the host immune system. Whether the direct feeding of *Varroa* mites suppresses the honey bee immune system is controversial. A recent study that used quantitative PCR of selected immune transcripts suggested that artificial wounding of developing pupae

increased immune gene expression and also the DWV titers, but mite feeding although it increased DWV titers had little effect on immune gene expression ((Kuster et al., 2014). But another recent study using RNAseq of the transcriptome of bees from naturally infested colonies found that infestation was correlated with down-regulation of a number of immune response genes, particularly the transcription factor *dorsal 1-A* and that experimental silencing of this gene resulted in increased DWV titers (Nazzi et al., 2012). In other words, it is not yet clear whether the increased titers of viruses associated with *V. destructor* are a benefit to the mite by overcoming host defenses or just an unfortunate result of Varroa feeding that suppresses immune response. Even though *V. destructor* is the most important mite affecting *A. mellifera,* there is a sister species *V. jacobsoni* that is gaining the attention of beekeepers in Papua New Guinea (PNG). This mite was first found parasitizing the Asian honey bee *A. cerana*, in Java, Indonesia, at the beginning of the 19th century and has currently been found living and reproducing on *A. mellifera* in PNG. As far as we know there is no information about viruses associated with this mite. Although a lot has been learned about Varroa-virus interactions in *V. destructor* it is not yet clear whether virus plays a role in allowing the mite to colonize the host, or whether these viruses are also present in *V. jacobsoni*.

We have collected nine samples of *V. jaconsoni*; three samples were collected in *A. cerana* hosts and five samples were collected form *A. mellifera* hosts. RNAseq was used to perform the first analysis of *V. jacobson* viral titers in these two populations. Results were used to analyze which viruses were present and whether their abundances differed between mites on the two honey bee hosts.

3.2 Materials and methods

3.2.1 Sample collection

Nine samples of *V. jacobsoni* from Papua New Guinea (PNG) and the Solomon Islands (Table 3.1) were collected from brood cells during April 2010. When collected, their reproductive status (reproducing or not reproducing) was recorded. *V. jacobsoni* reproducing on *A. mellifera* drone brood were collected from Goroka, PNG. Samples collected from St. Christobel Island (SC) and Ugi Island in Solomon Islands were reproducing on their usual host, *A. cerana*, and were only reproducing in drone brood cells, which is the expected situation on this host. In addition, single adult *V. jacobsoni* females were collected from *A. mellifera* colonies on SC Island and Ugi Island in the Solomon Islands that were not reproducing on either worker or drone brood. All samples were collected in RNAlater® and stored at -80 °C until RNA extraction.

3.2.2 RNA extraction and sequencing

Pools of adult female mites from each sample were ground in liquid nitrogen and total RNA from was extracted using the Invitrogen TRIzol® reagent protocol. All the steps for extraction were followed according to the TRIzol® protocol with one exception; the RNA precipitation step was slightly modified by the addition of 250μl of RNA precipitation solution (1.2 M NaCl + 0.8 M Sodium citrate dihydrate) mixed with 250μl of isopropanol to the aqueous phase of the mite homogenate to help precipitate more RNA. Approximately 20 mites per sample were used for extraction except for the non-reproducing mite sample. A total of five non-reproducing individual mites were used and the RNA was pooled. Total RNA per sample was assessed for quality using a NanoDrop

2000/2000c (Thermosceintific) and submitted to Purdue University Genomics Core Facility (PGCF) for sequencing. Total RNA was further analyzed for quality and concentration using an Agilent Technologies 2100 Bioanalyzer (Agilent Technologies, Inc. Santa Clara. CA). Seven out of nine cDNA libraries were prepared and barcoded by PGCF using TruSeq^{TM} RNA sample preparation kit (Illumina, Inc. San Diego, CA). These libraries were prepared and sequenced on April 2012 and January 2013 using the Illumina platforms HiScanSQ (100 bp paired reads in 2 lanes) and Hiseq2000 (100 bp paired reads in 4 lanes) respectively. Two of the nine cDNA libraries were prepared and sequenced using Hiseq2000 (100 bp paired reads in 1 lane) at the Biomolecular Resource Facility (BRF), Canberra, Australia (February 2014). Raw sequence reads from all 9 samples were then processed.

3.2.3 Read pre-processing

We extracted all the sequence reads that were a match to a virus or a virus-like sequence from the complete set of raw reads. We created a virus database by downloading all available complete virus genome sequences form the NCBI database with the exception of the human immunodeficiency viruses. In addition, all sequences that were a match to a virus (from Blastx of the non-redundant NCBI database) found in the transcriptome assembly of *V. jacobsoni* (data not shown) were also added to the database. Next, we used DeconSeq v 0.4.3 software (Schmieder & Edwards, 2011) to remove the virus contaminant reads. Finally, adapters were removed from the virus contaminant reads using in-house Perl scripts and quality trimming was performed using Trimmomatic v 0.30 (Bolger et al., 2014), trimmomaticSE -phred33

ILLUMINACLIP:adapters.fa:2:35:15   LEADING:7   TRAILING:7

SLIDINGWINDOW:4:13 MINLEN:30 (Table 3.2). Since the reads were independently

quality trimmed, some reads were unpaired after trimming, theses unpaired reads were

pooled together and added to the R1 reads set, for assembly purposes. Furthermore, we

added all paired and unpaired unmapped reads that were thrown out during the creation

of the *V. jacobsoni* virus assembly, in order to capture any reads that were a match to

virus that was not present in our database (Figure 3.1).

3.2.4 Creating the *V. jacobsoni* virus transcriptome assembly

The transcriptome assembly was performed using the paired and unpaired reads. A *de novo* transcriptome assembly was created using Trinity default parameters (kmer length = 25 and min_contig_length 200 nucleotides, Figure 3.1). In addition a second virus assembly was created with not only the virus-like reads extracted using DeconSeq, but also from the all the raw reads data and this was used to search specifically for deformed wing virus.

3.2.4.1 Description of a gene according to Trinity

For the *de novo* assembly Trinity reports many predicted transcripts, which are calculated by combining all the splice junctions predicted in the data; some of these predicted isoforms are not observed. In the first stage, Trinity reads are clustered according to their sequence overlap into components. As the assembly progresses the components are further divided into subcomponents and predicted isoforms.

The idea of a gene most closely matches the component level as determined by blast comparisons (data not shown). Therefore we refer to a Trinity component (gene/sequence) as a gene in our expression analyses.

3.2.4.2 Assessing the quality of the assembly

To assess the quality of the final assembled transcripts, all the virus RNAseq cleaned reads were aligned back to the *de novo* assembly using botwie2 (Langmead & Salzberg, 2012) and overall mapping statistics were estimated. To annotate and survey the assembled transcripts, we also conducted a complete Blastx similarity (E-value ≤ 1e-06) search against the non-redundant NCBI database. Furthermore, tblastx was performed for each virus genome sequence identified in the previous step using an E-value ≤ 1e-06.

3.2.5 Surveying the presence and absence of viruses

We first quantified the assembled transcripts using the Trinity align_and_estimate_abundance.pl script (Grabherr et al., 2011). This script uses RSEM (Li & Dewey, 2011) to estimate transcript abundances by mapping the RNAseq reads to the assembled transcriptome using the aligner tool bowtie2. Briefly, RSEM estimates the expected (true) counts (EC) for genes and predicted transcripts. The EC for each gene were further analyzed by ranking the EC according to their abundance. The most abundant genes, those covering at least 90% of EC in each sample, were selected for further analysis. The gene lists for each sample were compared with the nr database using Blastx, using an E-value ≤ 1e-06, and genes corresponding to the same viral sequence

identified. EC and FPKM values for viral sequences found to be abundant in any sample were extracted for analysis.

3.2.5.1 Surveying for the presence and absence of other pathogens

Besides viruses, other microbes and parasites associated with *V. jacobsoni* may be present. Therefore we used the assembly to perform a Blastn query of a curated public database containing about 60 mega-base pairs of public sequence from microbes associated with honey bees, and requiring an E-value < 0.00001. Positive hits were also queried against the non-redundant NCBI database.

3.2.6 Performing differential expression analysis

Expected counts per gene per sample were combined into a matrix count, and this matrix was used as input for all downstream expression analyses. This analysis was performed using DESeq2 R package v. 1.0.19 (Love et al., 2014). All differentially expressed genes with a False Discovery Rate (FDR) of 0.2 were extracted and used for downstream analysis; we will refer to these genes as differentially expressed mite virus genes (DEMVG) in the rest of the manuscript. Briefly, DESeq2 implements a model based on negative binomial distribution and was run under R release 3.0.1. Before performing the DE analysis, DESeq2 automatically performs independent filtering of the genes with low counts (weakly expressed) in order to maximize the number of DEMVG with adjusted p values less than a critical value of 0.1.

3.2.6.1 Heatmap of differentially expressed mites genes

To generate the heatmap and gene clustering, we used Edger and heatmap.3, R packages. FPKM (Fragments Per Kilobase per Million) values obtained from RSEM for each of the DEMVG were normalized and log2 transformed prior to gene clustering. Genes were clustered according to their pattern of expression.

3.2.7 Looking for evidence of deformed wing virus presence

We performed a Blastx to align the putative virus assembly against the non-redundant NCBI database and extracted the best-hit sequences that matched deformed wing virus (DWV) with an E-value $\leq$ 1e-06. Furthermore, we performed a local Blastn of the virus assembly against the best DWV isolate match found in the previous Blastx hits.

3.3. Results

3.3.1 *V. jacobsoni* virus transcriptome assembly

Using the Trinity assembler we created two *de novo* assemblies, the first one was created using all virus-like reads and all unmapped reads. This assembly produced a total of 11,877 trinity putative transcripts (10,205 trinity putative genes, N50 = 1,153 bp). The second assembly was created using only the virus-like reads that were extracted using DeconSeq. This assembly produced a total 943 trinity putative transcripts (454 trinity putative genes/sequences, N50 = 3,096 bp). *V. jacobsoni* mite samples were collected from two different honey bee hosts, *A. cerana* and *A. mellifera* which correspond to two different geographic locations, the Solomon Islands and PNG, respectively. A total of nine RNAseq libraries were constructed and sequenced using two Illumina sequencing

platforms (Table 3.1). Using DeconSeq we extracted total of 31,112,591 raw paired reads (15,556,295 PE) that were a match to a virus or a virus-like from the complete sequenced reads set of 2,184,624,960 billions reads. The viral library we created contained a total of 84,150 sequences of complete genome viruses (including the 1198 sequences that hast a match to a virus extracted from the *V. jacobsoni* assembly). Virus reads were then subjected to adapter removal, a total of 235,759 (<1%) reads with adapters were removed. Furthermore, a total of 617,561 (1.98%) reads with low quality were removed from the adapter clean reads set, leaving a total of 25,740,108 (82.73%) paired virus reads (cleaned and trimmed) and/or a total of 30,259,271 (97.25%) paired plus unpaired. A total of 543,73,224 paired unmapped reads that did not mapped to the *V. jacobsoni* assembly were extracted using bowtie2. Finally, a total of (84,620,808) clean virus reads plus unmapped (pair and unpair) reads were combined to create the transcriptome assemblies (Table 3.2).

3.3.1.1 Assessing the quality of the assembly

Clean reads ranging from 323,022 to 15,655,164 millions for each sample were mapped back to the *de novo* reference assembly using Bowtie2. Overall 84.3% of reads aligned to the reference indicating that almost all reads are represented in the assembly (Table 3.2).

3.3.2 What viruses are present in *V. jacobsoni*?

 By ranking the EC per each sample according to their accumulative values and extracting the genes with at least 90% of virus counts per each sample, we found a total of 27 sequences had a match to a virus annotation in the non-redundant NCBI database

(E-value ≤ 1e-06) database. Each one of these 27 sequences was found at least once in one of the samples. We found a total of 13 different viruses were represented in these sequences among them, Brevicoryne brassicae picorna-like virus, Deformed wing virus, Dragonfly cyclovirus 1, Farmington virus, Formica exsecta virus 2, Halyomorpha halys virus, Heliconius erato iflavirus, Kakugo virus, Kirsten murine sarcoma virus, Sacbrood virus, Spodoptera exigua iflavirus 1, and Tomato mosaic virus/Ngewotan virus which is a mite transmitted plant virus, see Table 3.3 for a detailed description of absence and presence of virus per sample. Among the most important viruses found in this survey was deformed wing virus. As far as we know this is the first report that DWV is present in *V. jacobsoni*.

Furthermore we evaluated the most abundant viruses across all samples and found a total of 12 sequences aligned to the non-redundant NCBI database with an E-value ≤ 1e-06. A total of 7 different virus descriptions are represented by these sequences among them Brevicoryne brassicae picorna-like virus, Dragonfly cyclovirus 1, Farmington virus, Halyomorpha halys virus, Heliconius erato iflavirus, Kirsten murine sarcoma virus and tomato mosaic virus. Table 3.4 shows the list of these viruses color coded for their abundances (FPKM values) across all nine samples of *V. jacobsoni*. The most abundant virus found, represented in this assembly is the Brevicoryne brassicae picorna-like virus but the second best hit for this sequence was to DWV so we investigated further as to its identity. Interestingly we also found evidence for the presence of Dragonfly cyclovirus 1, this virus has been previously found in birds and mammalian feces (Dayaram et al., 2013). It is a small, single-stranded circular DNA virus. Many isolates were found and

described in the Dragonfly (Rosario et al., 2011). Therefore its host range is not well defined. The top Blastx hit to dragonfly cyclovirus is a sequence (c37904) that is 932 bp long. A pairwise Blastx alignment of the replication-associate peptide sequence covers almost all the query with 49% identity. (Dayaram et al., 2013) reported that the replication-associated protein of the Dragonfly cyclovirus 1 isolates have at least 48% identity. In addition, when we translated the nucleotide sequence using ExPasy we found that this sequence it is one open reading frame. The capsid part of these viruses share only 29% identity. We did a Blastn query of the capsid sequence against the assembly but we were unable to find it.

3.3.2.1 Survey of honey bee microbes present in *V. jacobsoni*

Blastn of the virus sequences against the honey bee microbe/parasite database (p < 0.00001) revealed relatively few hits. However a few 'novel' microbes, were identified with the help of subsequent alignments to the non-redundant NCBI database. Among them were a few bacterial taxa, such a transcript with 94% identity to Spriroplasma apis 23S sequence. We also found one likely microsporidian (100% identity) and a few fungi. Microsporidians are to obligate parasites, which may present some control options if these parasites also infect *V. destructor*. We also found a solid match to a Tymovirus (Macula-like; Table 3.5). However, no hits to this virus were found in our previous Blastx search against non-redundant NCBI database.

3.3.3 Virus expression profile in *A. cerana* vs. *A. mellifera* hosts

We used DESeq2, to quantify expression and identify the differentially expressed mite virus genes (DEMVG). The expression analysis was performed at the gene level (10,205) putative genes in the count matrix) and a total of 25 DEMVG were extracted with a FDR < 0.2 (Figure 3.2). Of those DEMVG we found only four of these sequences were a match to a virus; 3 of them were higher expressed and one lower expressed in the mites reproducing in the *A. mellifera* host. Among the four viruses we found Cyclovirus PK5222, Farmington virus, Sacbrood virus, Spodoptera exigua iflavirus (Figure 3.2). It is worth clarifying that these virus descriptions are found multiple times across the survey virus list (Table 3.3) and the 12 most abundant virus list (Table 3.4). However, their Blastx ids and the sequence assembly ids are different, indicating that there are multiple sequences across the assembly that match the same virus description. We also observed that the pattern of expression levels across these four DEMVG showed a clear geographical separation between the samples collected in the Solomon Islands and PNG. To further investigate this we performed a principal component analysis (PCA) and the results from this analysis provided further evidence of this regional separation (Figure 3.3). According to the DEVMG the samples that were collected on the SI and found reproducing in the *A. cerana* host plus the sample not-reproducing on the *A. mellifera* host also collected in SI clustered together. The samples reproducing in *A. mellifera* that were collected in PNG clustered together in a separate group (Figure 3.3).

3.3.4 Evidence of Deformed wing virus in *V. jacobsoni*

To look for evidence of deformed wing virus we used the trinity assembly that produced only 943 transcripts (545 putatives genes/sequences). Since this assembly was less fragmented and had larger average length contigs, we decided it would be more suitable for looking for the deformed wing virus sequences. We performed a local Blastn of this assembly against the DWV- Pennsylvania isolate genome (10,166 bp) and we found a total of eight unique sequences were a match to DWV-PA isolate with percent identity > 97% using and E-value ≤ 1e-06 detailed Blastn results given in (Additional file 9: Table S11). Furthermore, we manually assembled these eight sequences and we observed they covered nearly the entire genome (Figure 3.4).

3.4 Discussion

We have assembled two *de novo* virus transcriptomes of *V. jacobsoni*. It is clear that there are some differences in the total amount of transcripts that each assembly produced. One assembly is more fragmented than the other one. We speculate that the reasons behind these differences are first, the total amount of reads used to create each assembly and second, the types of reads that were used. For the larger assembly with 10,206 sequences we used all virus-like reads that were a match to the custom viral database that we created plus all the unmapped reads that were not used by Trinity to create the transcriptome of *V. jacobsoni*. Usually the reads that do not assemble are either low quality or contain to many repetitive sequences (which is typical of contaminant sequences) that are difficult for Trinity to assemble, but we included unmapped reads in this assembly to try to capture any novel microbe sequences that were not present in the

viral library. The assembly with fewer and larger contigs was able to capture only the viruses that were present in our custom viral database. Our mapping results of 83.44% overall alignment of the viral reads against the virus reference assembly we consider to have a good representation of all the reads in the assembly.

With regards to the list of virus sequences that we find represented in this transcriptome it is surprising that we did not find many viruses known to infect honey bees, with the exception of DWV, sacbrood virus, and slow bee paralysis virus. Blackened queen cell virus and Varroa destructor virus-1 were not observed even though these viruses are very common in Europe and North America and associated with *V. destructor* infestations. We were interested to determine not just which viruses were represented in this data but also whether or not there were any differences in the types of viruses present between Varroa parasitizing the two bee species, that might indicate an association with the success of these mites in reproducing and living on a particular host or reflect acquisition of virus from a host.

Our findings suggest that overall all the mite samples had similar viruses, with slight differences in their abundances for some sequences. Twelve most abundant non-mite sequences had a match to a virus description, however, in some cases multiple sequences had a match to the same virus resulting in seven unique virus descriptions represented; among them; Brevicoryne brassicae picorna-like virus, Dragonfly cyclovirus 1, Farmington virus, Halyomorpha halys virus, Heliconius erato iflavirus, Kirsten murine sarcoma virus, tomato mosaic virus (Table 3.4). Interestingly, we found that the sequence

(c38419) is the sequence with the highest count across seven out of the nine samples and this sequence had a best Blastx match to the Brevicoryne brassicae picorna-like virus, however, the second best Blastx hit for this sequence was to a DWV and given our other analyses we believe this sequence is actually a DWV sequence. We also found another interesting virus, the Dragonfly cyclovirus 1. Further research is needed to determine whether this is a closely related species since we were unable to find the capsid sequence for this virus. So far only dragonflies appear to be a confirmed host of this species (Rosario et al., 2011). In addition, when we performed a Blatsn of the virus assembly against the complete genome of these two viruses we found not matches at all.

The expression analysis suggested that there are four different viruses that are differentially expressed between mites on *A. cerana* and *A. mellifera* hosts, three of them were found up-regulated on the *A. mellifera* host. Among those three sequences we found a match to dragonfly cyclovirus PK5222. Even though this exact sequence was not found among the most abundant sequences, this cyclovirus appears to be one of the most abundant viruses that were common to all samples. This analysis shows a clear geographical clustering of the samples according to the expression patterns. Samples collected in SI clustered together and the sample collected in PNG also clustered together in their own group. The lack of clear expression differences between the two hosts suggests that the viruses are not critical for host acquisition or overcoming host defenses.

We have solid evidence that deformed wing virus infects *V. jacobsoni* and it is surprising that it is most closely related to an isolate from North America. As far as we know this is

the first time that DWV has been reported in *V. jaconsoni* and indeed the virus pathogens of this mite had not previously been determined. We have clear evidence that a good match to the DWV-PA isolate is almost fully contained in eight sequences of our assembly at about 97% overall sequence identity and E-value ranging from 1E-113 to 0. It seems likely that the DWV-PA isolate originated in Asia and spread to the rest of the world using *V. destructor* as a vector. *V. jacobsoni* also harbors this virus.

3.5 Conclusions

We have assembled a virus transcriptome of *V. jacobsoni* to provide the first survey of pathogens in this species. Among the list of putative viruses found we have Brevicoryne brassicae picorna-like virus, Deformed wing virus, Dragonfly cyclovirus 1, Farmington virus, Formica exsecta virus 2, Halyomorpha halys virus, Heliconius erato iflavirus, Kakugo virus, Kirsten murine sarcoma virus, Sacbrood virus, Spodoptera exigua iflavirus 1, and Tomato mosaic virus. Additional Blastn search of a honey bee associated microbe database revealed the likely presence of Macula-like virus (Tymoviridae), a microsporidan and a spiroplasma previously reported from *A. mellifera*. We are reporting for the first time the presence of deformed wing virus in *V. jacobsoni* samples reproducing on *A. mellifera* host that were collected from Goroka, PNG and also found in mite samples reproducing on *A. cerana* collected from Solomon Islands.

References

1. Bolger AM, Lohse M & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120. doi:10.1093/bioinformatics/btu170.

2. Cox-Foster D, Conlan S, Holmes E, Palacios G, Evans J, Moran N, Quan P, Briese T, Hornig M & Geiser D (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. Science 318: 283 - 287.

3. Dainat B, Evans JD, Chen YP, Gauthier L & Neumann P (2012) Dead or alive: deformed wing virus and *Varroa destructor* reduce the life span of winter honeybees. Appl Environ Microbiol 78: 981-987. doi:10.1128/AEM.06537-11.

4. Dayaram A, Potter KA, Moline AB, Rosenstein DD, Marinov M, Thomas JE, Breitbart M, Rosario K, Argüello-Astorga GR & Varsani A (2013) High global diversity of cycloviruses amongst dragonflies. Journal of General Virology 94: 1827-1840.

5. de Miranda JR, Cordoni G & Budge G (2010) The Acute bee paralysis virus–Kashmir bee virus–Israeli acute paralysis virus complex. J Invertebr Pathol 103, Supplement: S30-S47. doi:http://dx.doi.org/10.1016/j.jip.2009.06.014.

6. de Miranda JR & Genersch E (2010) Deformed wing virus. J Invertebr Pathol 103 Suppl 1: S48-61. doi:10.1016/j.jip.2009.06.012.

7. Francis RM, Nielsen SL & Kryger P (2013) Varroa-Virus Interaction in Collapsing Honey Bee Colonies. PLoS One 8: e57540. doi:10.1371/journal.pone.0057540.

8. Genersch E, von der Ohe W, Kaatz H, Schroeder A, Otten C, Buchler R, Berg S, Ritter W, Muhlen W, Gisder S, Meixner M, Liebig G & Rosenkranz P (2010) The German bee monitoring project: a long term study to understand periodically high winter losses of honey bee colonies. Apidologie 41: 332-352. doi:Doi 10.1051/Apido/2010014.

9. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N & Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29: 644-652. doi:10.1038/nbt.1883.

10. Highfield AC, El Nagar A, Mackinder LCM, Noël LM-LJ, Hall MJ, Martin SJ & Schroeder DC (2009) Deformed Wing Virus Implicated in Overwintering Honeybee Colony Losses. Appl Environ Microbiol 75: 7212-7220. doi:10.1128/aem.02227-09.

11. Kevan PG, Hannan M, Ostiguy N & Guzman-Novoa E (2006) A summary of the varroa-virus disease complex in honey bees.

12. Kuster RD, Boncristiani HF & Rueppell O (2014) Immunogene and viral transcript dynamics during parasitic *Varroa destructor* mite infection of developing honey bee (*Apis mellifera*) pupae. J Exp Biol 217: 1710-1718. doi:10.1242/jeb.097766.

13. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-359. doi:10.1038/nmeth.1923.

14. Le Conte Y, Ellis M & Ritter W (2010) Varroa mites and honey bee health: can Varroa explain part of the colony losses?*. Apidologie 41: 353-363.

15. Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2.

16. Martin SJ (2001) The role of Varroa and viral pathogens in the collapse of honeybee colonies: a modelling approach. Journal of Applied Ecology 38: 1082-1093. doi:Doi 10.1046/J.1365-2664.2001.00662.X.

17. Martin SJ, Highfield AC, Brettell L, Villalobos EM, Budge GE, Powell M, Nikaido S & Schroeder DC (2012) Global honey bee viral landscape altered by a parasitic mite. Science 336: 1304-1306. doi:10.1126/science.1220941.

18. Mondet F, de Miranda JR, Kretzschmar A, Le Conte Y & Mercer AR (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 Pathog 10: e1004323. doi:10.1371/journal.ppat.1004323.

19. Nazzi F, Brown SP, Annoscia D, Del Piccolo F, Di Prisco G, Varricchio P, Della Vedova G, Cattonaro F, Caprio E & Pennacchio F (2012) Synergistic Parasite-Pathogen Interactions Mediated by Host Immunity Can Drive the Collapse of Honeybee Colonies. PLoS Pathog 8: e1002735. doi:10.1371/journal.ppat.1002735.

20. Neumann P & Carreck NL (2010) Honey bee colony losses. Journal of Apicultural Research 49: 1-6.

21. Nguyen BK, Mignon J, Laget D, de Graaf DC, Jacobs FJ, vanEngelsdorp D, Brostaux Y, Saegerman C & Haubruge E (2010) Honey bee colony losses in Belgium during the 2008-9 winter. Journal of Apicultural Research 49: 337-339. doi:Doi 10.3896/Ibra.1.49.4.07.

22. Rosario K, Marinov M, Stainton D, Kraberger S, Wiltshire EJ, Collings DA, Walters M, Martin DP, Breitbart M & Varsani A (2011) Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). Journal of General Virology 92: 1302-1308. doi:10.1099/vir.0.030338-0.

23. Rosenkranz P, Aumeier P & Ziegelmann B (2010) Biology and control of *Varroa destructor*. J Invertebr Pathol 103 Suppl 1: S96-119. doi:10.1016/j.jip.2009.07.016.

24. Schmieder R & Edwards R (2011) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One 6: e17288. doi:10.1371/journal.pone.0017288.

25. Shimanuki H & Knox DA (1994) Parasitic mite syndrome: the symptoms. American Bee Journal 134: 827-828.

26. vanEngelsdorp D, Evans JD, Saegerman C, Mullin C, Haubruge E, Nguyen BK, Frazier M, Frazier J, Cox-Foster D, Chen YP, Underwood R, Tarpy DR & Pettis JS (2009) Colony Collapse Disorder: A Descriptive Study. PLoS One 4. doi:ARTN e6481. doi 10.1371/journal.pone.0006481.

Table 3.1 Description of *V. jacobsoni* samples submitted for RNA sequencing

Bee host	Reproductive status	Collection	Collection sites	Year of sequencing
A. cerana	Reproducing	Drone cells	SC*, **Solomon Islands**	Apr 2012 (HiScanSQ)
A. cerana	Reproducing	Drone cells	UGI, **Solomon Islands**	Jan 2013 (Hiseq2000)
A. cerana	Reproducing	Drone cells	Guadalcanal, **Solomon Islands**	Jan 2013 (Hiseq2000)
*A. mellifera***	Non-reproducing	Drone and worker cells	SC and UGI (**Solomon Islands**)	Apr 2012 (HiScanSQ)
A. mellifera	Reproducing	Drone cells	Goroka**, PNG**	Apr 2012 (HiScanSQ)
A. mellifera	Reproducing	Drone cells	Goroka**, PNG**	Jan 2013 (Hiseq2000)
A. mellifera	Reproducing	Drone cells	Goroka**, PNG**	Jan 2013 (Hiseq2000)
A. mellifera	Reproducing	Drone cells	Goroka**, PNG**	Jan 2014 (Hiseq2000)
A. mellifera	Reproducing	Drone cells	Goroka**, PNG**	Jan 2014 (Hiseq2000)

*SC = San Cristobel, Salomon Islands

PNG = Papua New Guinea

**Non-reproducing, individual adult females were pooled together expecting to get more RNA for sequencing.

Table 3.2 Summary of sequencing, processing, and mapping of *V. jacobsoni* virus reads

Sample-ID	Raw reads	w/Adapters	w/o-Adapters	Trimmed reads	Clean pairs	Unmapped pairs	All pairs (%align)
Ac-reproductive	2344084	14968	2329116	22889	2098000	4572544	6670544 (86.45%)
Ac-reproductive	5030223	41384	4988839	125532	4337846	8992860	13330706 (84.69%)
Ac-reproductive	7838360	53156	7785204	164120	5562598	10092566	15655164 (82.30%)
Am-not reproductive	117713	2077	115636	1787	101486	221536	323022 (82.75%)
Am-reproductive	2277776	9640	2268136	18380	1960224	3645812	5606036 (83.62%)
Am-reproductive	2703034	17740	2685294	48169	2364678	5233234	7597912 (84.90%)
Am-reproductive	3196890	52283	3144607	74535	2752184	6747124	9499308 (85.12%)
Am-reproductive	3499734	19934	3479800	102555	2905228	7548686	10453914 (84.00%)
Am-reproductive	4104777	24577	4080200	59594	3657864	6311402	9969266 (85.10%)
Undetermined	-	-	-	-	-	1007460	-
Total	**31112591**	**235759**	**30876832**	**617561**	**25740108**	**54373224**	**80113332 (84.33%)**

Summary	
Total pairs	80113332
R1 (Left)	40056666
R2 (Right)	40056666
Total reads for Trinity assembly	
R1 plus un-mapped unpaired	**44564142**
R2	**40056666**
Total	**84620808**

*Reads where the barcode could not be decoded.

Table 3.3 Survey of viruses that were abundant (at least 90% of virus counts) in at least one of the *V. jacobsoni* samples

Gene id	Blastx id	Description	Am_1	Am_2	Am_3	Am_4	Am_5	Am_6	Ac1	Ac2	Ac3	E-value
c37570	gi|297598943|gb|ADI48251.1|	Bat cyclovirus GF-4c		X	X	X	X	X				1E-30
c38383	gi|290783614|gb|ADD62453.1|	Cyclovirus PK5034	X	X			X				X	1E-26
c37988	gi|290783653|gb|ADD62479.1|	Cyclovirus TN18	X	X	X	X	X	X	X	X	X	1E-24
c33258			X	X	X	X	X	X	X	X	X	1E-41
c36526							X					1E-20
c37904	gi|480306442|gb|AGJ74756.1|	Dragonfly cyclovirus 1	X	X	X	X	X	X	X	X	X	5E-68
c38220				X			X					1E-47
c38333			X	X	X	X	X	X	X	X	X	3E-64
c38413	gi|511649168|gb|AGN91191.1|		X	X	X	X	X	X	X	X	X	7E-124
c38420		Farmington virus	X	X	X	X	X	X	X	X	X	3E-21
c38421			X	X	X	X	X	X	X	X	X	0.000
c38410	gi|31540604|gb|AAP49283.1|	Deformed wing virus	X				X	X				0.000
c38406	gi|571026785|ref|YP_008888537.1|	Formica exsecta virus 2 (VDV-1 1E-09, DWV 2E-09)		X			X					3E-22
c38409	gi|612400723|ref|YP_009026409.1|	Heliconius erato iflavirus (DWV-1 9E-49)	X	X	X	X	X	X	X	X	X	1E-53
c38419	gi|148717938|ref|YP_001285409.1|	Brevicoryne brassicae picorna-like virus (DWV 3E-57,VDV-1 2E-16)	X	X	X	X	X	X	X	X	X	4E-62
c38383	gi|114842247|dbj|BAF32608.1|	Kakugo virus (DWV 1E-09, and 2E-07)	X	X			X					9E-33

(continued)

c38412	gi|555928176|ref|YP_008719809.1|	Halyomorpha halys virus (SBV, 4E-15)	▓	▓	▓	▓	▓	▓	▓	▓	▓	1E-19
c38403	gi|402749284|gb|AFQ95416.1|	Sacbrood virus (SBPV, 2E-14)	▓	□	▓	□	▓	□	□	▓		4E-31
c38405	gi|296005647|ref|YP_003622540.1|	Slow bee paralysis virus	▓	□	□	▓	□					5E-11
c38416	gi|297578409|gb|ADI46683.1|	Slow bee paralysis virus	□	▓	□	▓	□					2E-11
c38321	gi|357580074|ref|YP_004935365.1|	Spodoptera exigua iflavirus 1 (SBPV, 1E-06)	▓	▓	□	▓	□					2E-15
c38397			▓	□	□	▓	▓	▓	□	□	▓	9E-11
c38417	gi|241911792|gb|ACS71757.1|	Tomato mosaic virus (Ngewotan virus 2E-31, mite transmitted plant virus)	▓	▓	▓	▓	▓	▓	▓	▓		7E-80
c38418	gi|612400723|ref|YP_009026409.1|	Heliconius erato iflavirus	▓	□	□	□	□					2E-70
c37658	gi|584595298|gb|AHI42034.2|	Human cyclovirus	▓	□	□	□	□					1E-24
c38152	gi|939930|emb|CAA80675.1|	Kirsten murine sarcoma virus	▓	▓	▓	▓	▓	▓	▓	▓	▓	4E-89

[*]Descriptions in parenthesis are matches to the second best Blastx hits, DVW = Deformed wing virus, VDV = *Varroa destructor Virus*, SBPV = Slow Bee Paralysis virus and SBV = Sackbrood Virus.

Am = *A. mellifera* host, Ac = *A. cerana* host. Virus present = gray box, virus sequence absent = white box

Table 3.4 Most abundant viruses found across all *V jacobsoni* samples

Gene id	Blastx id	Am_1	Am_2	Am_3	Am_4	Am_5	Am_6	Ac1	Ac2	Ac3	Description	E-value
c38421	gi|511649168|gb|AGN91191.1|										Farmington virus	0.000
c37988	gi|290783653|gb|ADD62479.1|										Cyclovirus TN18	1.00E-24
c38413	gi|511649168|gb|AGN91191.1|										Farmington virus	7.00E-124
c33258	gi|480306442|gb|AGJ74756.1|										Dragonfly cyclovirus 1	1.00E-41
c38152	gi|939930|emb|CAA80675.1|										Kirsten murine sarcoma virus	4.00E-89
c38333	gi|480306442|gb|AGJ74756.1|										Dragonfly cyclovirus 1	3.00E-64
c38420	gi|511649168|gb|AGN91191.1|										Farmington virus	3.00E-21
c37904	gi|480306442|gb|AGJ74756.1|										Dragonfly cyclovirus 1	5.00E-68
c38417	gi|241911792|gb|ACS71757.1|										Tomato mosaic virus (Ngewotan virus 2E-31, mite transmitted plant virus)	7.00E-80
c38412	gi|555928176|ref|YP_008719809.1|										Halyomorpha halys virus (SBV, 4E-15)	1.00E-19
c38409	gi|612400723|ref|YP_009026409.1|										Heliconius erato iflavirus (VDV-1 9E-49)	1.00E-53
c38419	gi|148717938|ref|YP_001285409.1|										Brevicoryne brassicae picorna-like virus (DWV 3E-57,VDV-1 2E-16)	4.00E-62

Descriptions in parenthesis are matches to the second best Blastx hits, DVW = Deformed wing virus, VDV = *Varroa destructor* virus, SBPV = Slow Bee Paralysis virus and SBV = Sackbrood Virus. Am = *A. mellifera* host, Ac = *A. cerana* host. Color sacle boxes indicate FPKM values, red = lowest PFKM values to blue = highest PFKM values.

Table 3. 5 Honey bee related microbes and parasites found in *V. jacobsoni* virus assembly

| Gene id	Total Hits	BLASTN or MegaBLAST – NR (% identity)
c4823	1	**Kinase, Spraguea (microsporidia) plus snRNA U6 100% small noncoding**
c39412	1	Cand. Saccharibacteria 16S (94%)
c44280	1	Penicillium marneffei actin (93%)
c3550	4	Corynebacterium 23S (97%)
c37800	5	**Macula-like virus (Tymoviridae) (95%)**
c37050	5	Pseudomonas 23S (99%)
c37899	5	Psychrobacter 23S (gammaprot) (92%)
c43501	6	DWV_Canada2 (plus others) (98%)
c39827	7	DWV_PA (99%)
c31436	8	DWV_PA isolate (98%)
c41977	9	DWV_PA (99%)
c37060	10	DWV_Warwick (98%)
c38410	11	DWV_PA isolate (99%)
c38204	16	DWV_PA capsid (99%)
c35372	16	Neisseria 23S (gamma-proteo bact) (100%)
c38315	70	**23S Spiroplasma (94%)**
c38151	2018	16S Enterococcus faecalis (100%)
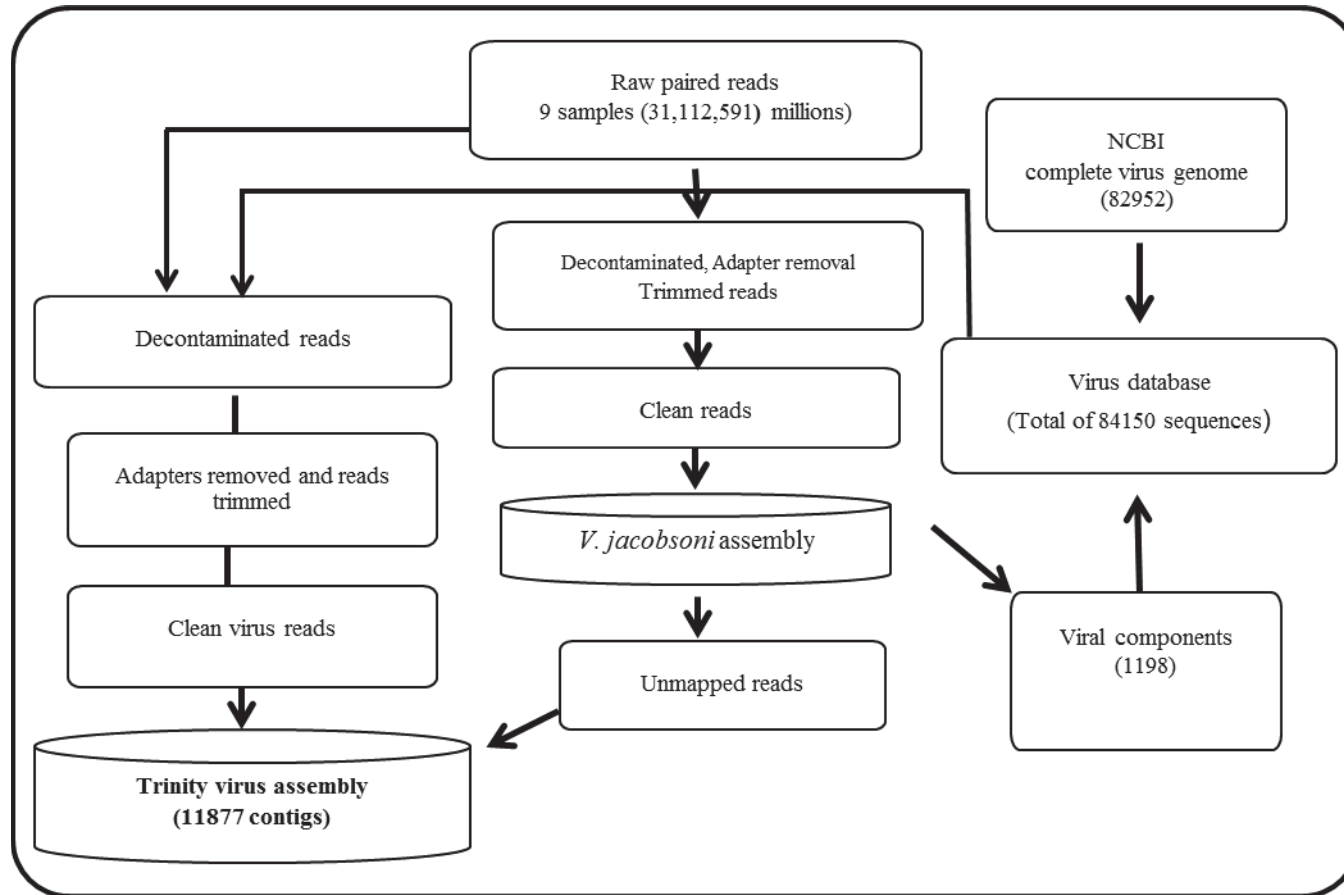
Figure 3.1 Pipeline steps followed to build the *V. jacobsoni* virus assembly

Flow chart shows the steps implemented from raw reads to the creation of the final virus assembly
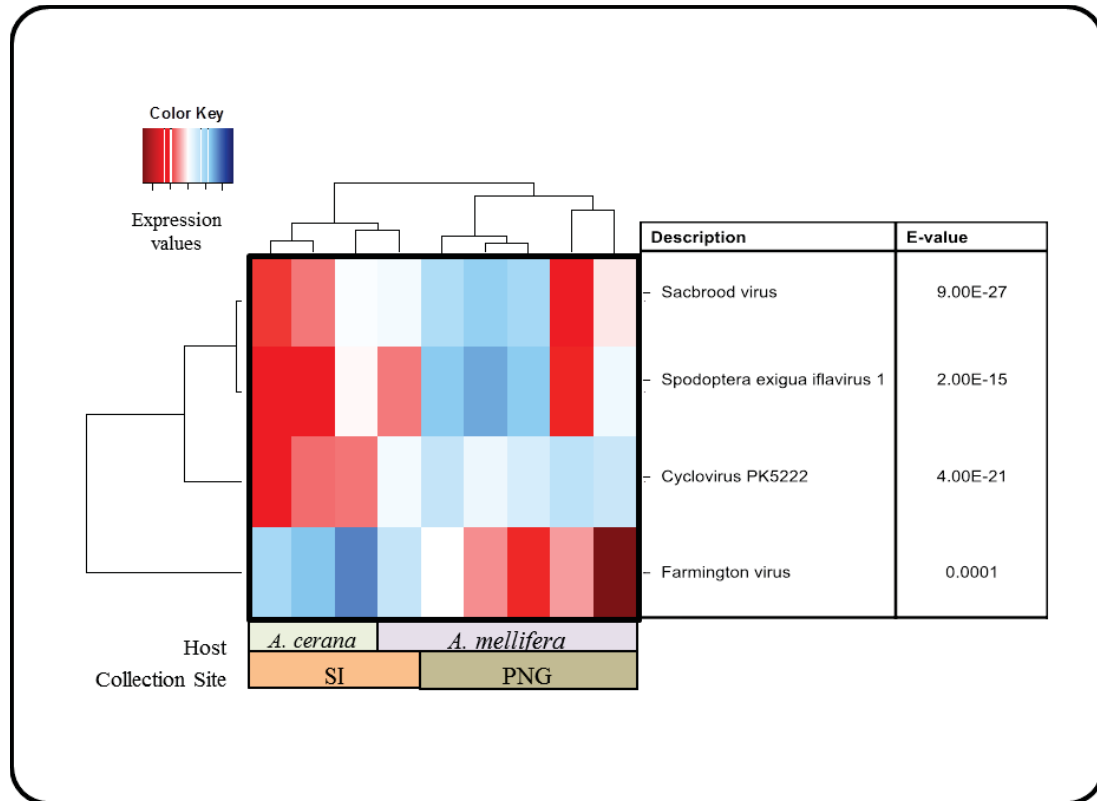
Figure 3.2 Heatmap of differentially expressed virus sequences of *V. jacobsoni* mites (FDR < 0.2)

Heatmap of expression values (log2 transformed normalized FPKM) of the differentially expressed virus genes of adult female *V. jacobsoni* mites reproducing on *A. cerana* or *A. mellifera*. Dark red and dark blue indicate higher and lower expression values, respectively. Light yellow and light purple bars indicate *A. cerana* host and *A. mellifera* host, respectively. Light orange and khaki bars indicate collection site, Solomon Islands and Papua New Guniea, respectively.
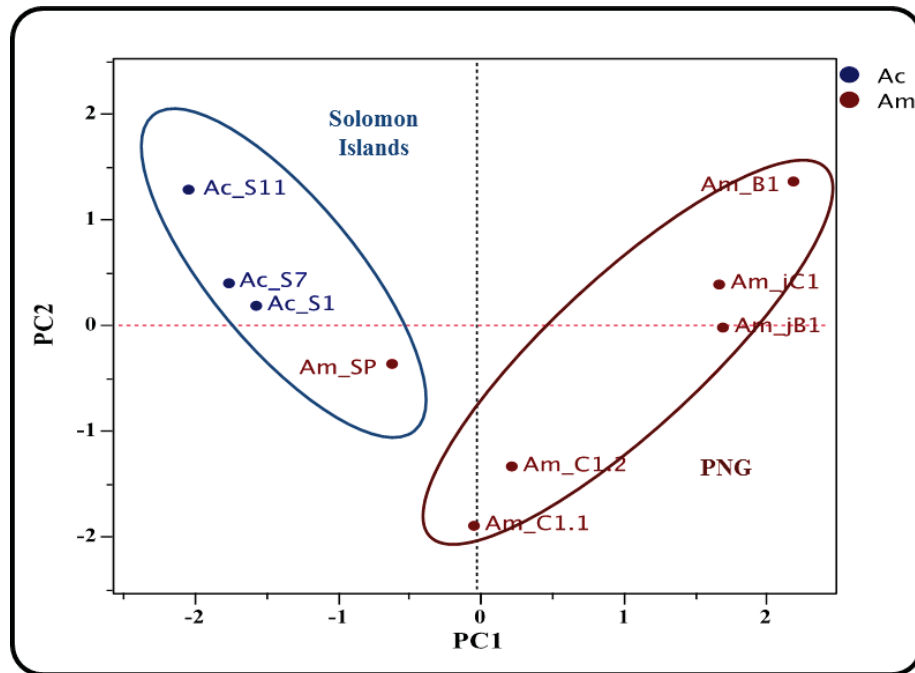
Figure 3.3 Principal component analysis of global expression profiles of *V. jacobsoni*

mite viruses per each sample

Samples names highlighted in the dark blue circle were collected from the Solomon

Island and sample names in dark red circle were collected from Papua New Guinea
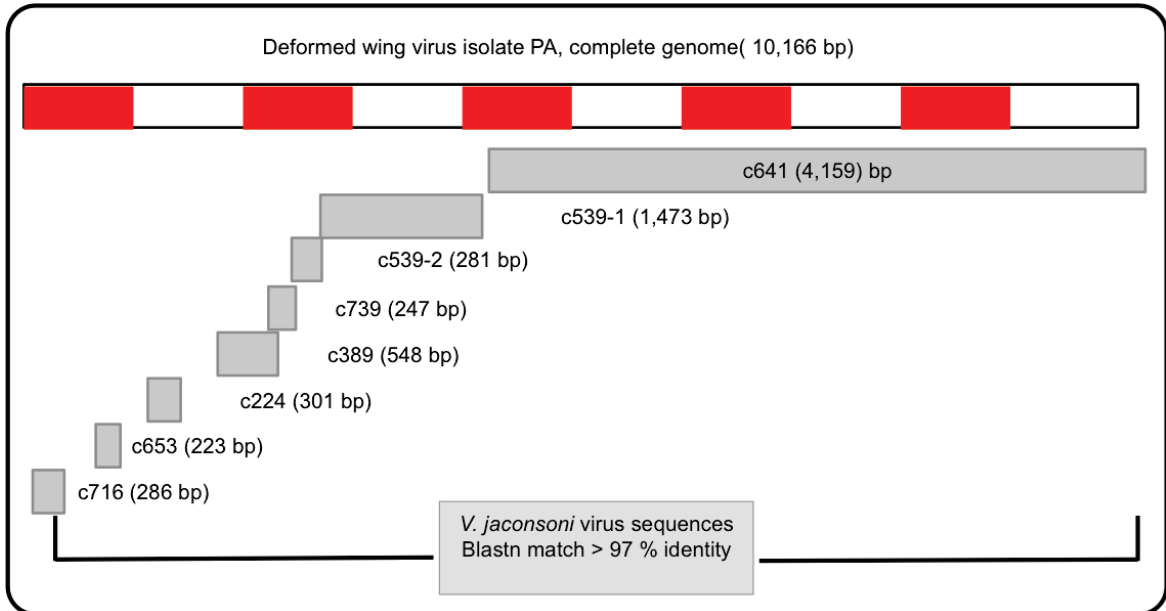
(PNG).

Figure 3.4 DWV-PA isolate sequence coverage by *V. jacobsoni* virus sequences

Red, white bars are spaced every 1,000 bases along the genome DWV-PA as a reference.

VITA

VITA

EDUCATIONAL BACKGROUND

- PhD. in Entomology, Purdue University, West Lafayette, IN. (December 2014).

  Dissertation: "Changes in gene expression and viral titer in *Varroa jacobsoni* mites after a host shift from Asian to European honey bees", GPA: 3.86.

- Bachelor of Science in Agriculture. (December 2002).

  Pan-American School of Agriculture Zamorano, Valle del Yeguare, Francisco Morazán, Honduras. Thesis: "The Transmission Mechanisms of Little Leaf Disease in *Gliricidia sepium* caused by a Phytoplasma"

WORK EXPERIENCE

- Teaching Assistant

  Genetics, Agronomy 32000, Department of Agronomy, Purdue University, West Lafayette, IN. (Spring 2011).

- Assisted the professor with class activities and responsibilities. Provided help to students with the course during office hours and review sessions. I was responsible for grading student work, quizzes and exams as well as preparing review sessions.

- Apiculture Research Assistant

  Department of Entomology, Purdue University, West Lafayette, IN. (March 2009 - May 2009).

- Involved in the study of grooming behavior in honey bees as a mechanism of resistance to parasitic mites, colony management and queen rearing.

- Visiting Scholar Researcher

  Department of Entomology, Purdue University, West Lafayette, IN. (June 2006 - December 2008).

- Assisted in the analysis of soybean aphid populations and aphid predator populations (insects field sampling and collection).

- Involved in the investigation of the genomics and genetics of Hessian fly.

  o Rearing of hessian flies.

  o Collected insects from genetic mapping populations, extracted DNA, performed polymerase chain reaction (PCR), amplified fragment length polymorphism (AFLP) and analyzed the products of the reactions.

- Involved in the genetic analysis of spruce budworm for a forest service research project using AFLP's techniques.

- Involved in the honey bee breeding project for resistance to parasitic mites, beekeeping, and QTL's analysis.

- Official of Agricultural Quarantine

  International Regional Organization for Plant and Animal Health (OIRSA)/ National Service for Health Agriculture (SENASA) Tegucigalpa, Honduras. (May 2003 - April 2006).

  - Monitoring and identification of the important quarantine insect pests of Honduras.

- National consulting of pest quarantine and manager of the laboratory of distance diagnostics through digital imaging (DDDI).

- Agricultural Extension Specialist

Intibucana of Projects INPRO-PROSOC, Intibucá, Honduras. (March - May 2003).

  - Advised and assisted small farmers on agricultural activities through technical support and field demonstrations.

SYNERGISTIC ACTIVITIES

- Elected President, Association of Zamorano Alumni (AZA) at Purdue (2013).

  - Lead organizer of the 2nd Symposium of Zamoranos in the U.S. "The building bricks for the future of agriculture" and co-author of the proposal that acquired the funding for the event, July 20th, 2013. Kurz Purdue Technology Center, West Lafayette, IN. (January - July 2013).

  - Organized and lead monthly meetings and coordinated association activities.

- Assistant, Pest Management Conference Purdue Pest Management Conference, Department of Entomology, Purdue University, West Lafayette, IN. (2011-2013).

  - Assisted exhibit hall management: facilitating information and answering questions to vendors and attendees as well mediating any issues between vendors.

  - Paging: assisted conference speakers with any demonstrations during their talk and handed out comment cards to any audience members for questions to the moderator.

  - Information booth assistant: assisted attendees answering questions related to the activities of the event.

- Intern, Plant Pathology and Molecular Laboratory of Plant Pathology, Zamorano University, Honduras. (January - April 2002).

    - Assisted and performed lab set up for undergraduate students involved in the laboratory module of Biotechnology "Molecular diagnostics of pathology of plants".

    - Assisted in the diagnostic laboratory with the reception and analysis of samples (microscopic, *In Vivo* and *In Vitro* analysis) of sick plants.

- Spanish Teacher, provided Spanish language lessons to Purdue graduate students

SEMINARS, SHORT COURSES AND WORKSHOPS TAKEN

- 18th Summer Institute in Statistical Genetics (SISG 2013) University of Washington, Seattle, WA July 8, 2013 - July 24, 2013.

    - Module 3: Introduction to R (20 hours).

    - Module 6: Population Genetic data Analysis (20 hours).

    - Module 9: Gene expression Profiling (20 hours).

- VectorBase Hands-on Workshop 2013, University of Notre Dame, Notre Dame, IN.

- Programming for Biologists, Cold Spring Harbor Laboratory. Cold Spring Harbor, NY, 2012 (130 hours).

- Introduction to R for Non-Statisticians, 8th International Purdue Symposium on Statistics, Purdue University, West Lafayette IN, 2012 (4 hours).

- i5K Community Workshop 2012, Kansas City, MO, 2012 (16 hours).

- iPlant Tools and Services Workshop. Purdue University, West Lafayette IN, 2012 (16 hours).

- Beginning UNIX for Bioinformatics: A Hands-on Workshop, Purdue University, West Lafayette, IN. (May 10th, 2012).

- Seminar: Beginning Unix for Bioinformatics. Purdue University, West Lafayette, IN. (April 12th, 2012).

- Entomology Department Graduate Student Seminar Series. Purdue University. (Spring-Fall, 2007 - Present).

- Workshop: Utilizing "Omic" Technologies to Accelerate Crop Improvement. Purdue University, West Lafayette, IN, 2011 (8 hours).

- Insect Pathology, NCR-125 Midwest Institute for Biological Control, Urbana-Champaign University, 2010 (40 hours).

- Use and Handling of Aluminum Phosphide. Detia-Bayer-OIRSA. San Pedro Sula, Honduras, 2005 (8hours).

- Introduction to Acarology General-Case Study in *Stenotarsonemus spinki* Smile (1966). Institute of Investigation of Plant Health (INISAV), Havana, Cuba, 2005 (40 hours).

- Advance Course ¨Agricultural Quarantine¨. International Regional Organization for Plant and Animal Health (OIRSA). San Pedro Sula, Honduras, 2004 (40 hours).

- Sensitization on the National Frame on the Security of the Biotechnology¨. San Pedro Sula, Honduras, 2004 (16 hours).

- "Dangerous Merchandise". Port center of Marine Training (CENCAMAPORT) and National Port Company (ENP), 2004 (40 hours).

- IX International Congress of Integrated Pest Management, San Salvador, El Salvador, November 2004 (72 hours).

- Basic Course ¨Agricultural Quarantine¨. International Regional Organization for Plant and Animal Health (OIRSA) San Pedro Sula, Honduras, 2003 (40 hours).

- Agricultural Science Workshop Series, Zamorano University. Fall 2002 (4-16 hours).

- Urban Landscaping. Zamorano University and Cortés, Honduras, 2001 (40 hours).

- Agricultural Science Workshop Series, Zamorano University. Fall 2001 (8-24 hours).

PRESENTATIONS GIVEN

- Andino, G.K., Gribskov, M., Anderson, D., Hunt, G. J. *Varroa jacobsoni* mites that differ in their reproductive success on the European honey bee (*Apis mellifera*) display differential gene expression. Presentation at the Workshop: Understanding the Apis-Varroa interactions; insights for improving bee health. Cairns, Australia, July 11-12, 2014.

- Andino, G.K., Gribskov, M., Anderson, D., Hunt, G. J. *Varroa jacobsoni* haplotypes that differ in their reproductive success on the European honey bee (*Apis mellifera*) display differential gene expression. Poster presentation at the Sigma Xi Chapter: Graduate Student and Post-Doctoral Research Poster Award Competition. Purdue University, West Lafayette, IN. (February 2014).

- Andino, G.K., Gribskov, M., Anderson, D., Hunt, G. J. *Varroa jacobsoni* haplotypes that differ in their reproductive success on the European honey bee (*Apis mellifera*) display differential gene expression. Poster competition at the 2nd Symposium of Zamoranos in the U.S. "The building bricks for the future of agriculture". Kurz Purdue Technology Center, West Lafayette, IN. (July 2013).

- Andino, G.K. Overview of the Association of Zamorano Alumni (AZA) At Purdue. Speaker invited at the 2nd Symposium of Zamoranos in the U.S. "The building bricks for the future of agriculture". Kurz Purdue Technology Center, West Lafayette, IN. (July 2013).

- Andino, G.K., Gribskov, M., Anderson, D., Hunt, G. J. *Varroa jacobsoni* haplotypes that differ in their reproductive success on the European honey bee (*Apis mellifera*) display

differential gene expression. Graduate student poster competition at the NCB-ESA meeting in Rapid City, South Dakota 2013. (Winner of the 3rd place, June 2013).

- Andino, G.K., Gribskov, M., Anderson, D., Hunt, G. J. *Varroa jacobsoni* haplotypes that differ in their reproductive success on the European honey bee (*Apis mellifera*) display differential gene expression. Poster presentation at the 7th Annual Arthropod Genomics Symposium and VectorBase Workshop 2013, University of Notre, Dame Notre Dame, IN. (June 2013).

- Andino, G.K., San Miguel, P., Westerman, R., Anderson, D., Hunt, G.J. Transcriptome analysis of *Varroa jacobsoni* that differ in their reproductive success on *Apis mellifera*. Poster presentation at 6th Annual Arthropod Genomics Symposium: Arthropod Genomics 2012: Taking Center Stage and i5k Community Workshop, Kansas City, MO. (May 30 – June 2, 2012).

- Andino, G.K. and Hunt, G.J. A new assay to measure mite-grooming behavior in honeybees. Graduate student poster competition at the First Symposium of Zamoranos in the U.S. OARDC, Wooster, OH. May 2011. (Winner of the 1st place).

- Andino, G.K. Genetic analysis of honey bee resistance to *Nosema ceranae*, Methods seminar presentation at Department of Entomology, Purdue University, West Lafayette IN. (April 2011).

- Andino, G.K. and Hunt, G.J. A new assay to measure mite-grooming behavior in honey bees. Graduate student poster competition at the NCB-ESA meeting in Minneapolis, MN. March 2011. (Winner of the 2nd place).

- Andino, G.K. and Hunt, G.J. A new assay to measure mite-grooming behavior. Poster presentation at Sigma Xi Chapter, Purdue University, West Lafayette, IN. (February 2011).

- Andino, G.K. A new assay to measure mite-grooming behavior. Presentation at the American Bee Research Conference (ABRC), Orlando, FL. (January 2010).

- Ammons, A., Andino, G.K., Arechavaleta, M., Emore, E., Guzmaán, E., Hunt, G.J. and Schlipalius, D. The genetics of the Honey Bee Colony Defense. Poster presentation at the Pulse Retreat 2008, Purdue University, West Lafayette, IN. (August 2008).

- Ammons, A., Andino, G.K., Arechavaleta, M., Emore, E., Guzmaán, E., Hunt, G.J. and Schlipalius, D. The genetics of the Honey Bee Colony Defense. Poster presentation at the Second Annual Arthropod Genomics Symposium, Kansas City, MO. (April 2008).

- Study of a New Disease of *Gliricidia sepium* In Honduras. Presentation at the 48th Annual Meeting of the Interamerican Society for Tropical Horticulture. Tegucigalpa, Honduras. (October 2002).

OUTREACH WORKSHOPS AND PRESENTATIONS GIVEN

- Andino, G.K. Grooming behavior assays to select for *Varroa* mite resistance. Workshop Presentation at the Indiana State Beekeepers Association (ISBA), Summer meeting and Nuc Day, Purdue University, West Lafayette, IN. (June 23rd, 2012).

- Andino, G.K. Diseases and mite resistant bees: testing for hygienic and mite-grooming behavior. Workshop presentation at the Heartland Apicultural Society Conference (HAS), St. Vincennes, IN. (July 2011).

- Andino, G.K. Extraction, preparation and identification of *Nosema ceranae* in honey bees. Workshop presentation at the Indiana State Beekeepers Association (ISBA), Purdue University, West Lafayette, IN. (June 2009 and 2010).

- Andino, G.K. A new assay to measure mite-grooming behavior. Presentation at the Indiana State Beekeepers Association (ISBA), Danville, IN. (October 2009).

- Andino, G.K. Looking for resistance and susceptibility response to *Nosema cerana* in honey bees *Apis mellifera*. Presentation at the Indiana State Beekeepers Association (ISBA), Danville, IN. (October 2009).

PUBLICATIONS

Journal Articles

- Krupke, C.H., Hunt, G.J., Eitzer, B.D., Andino, G.K. and Given, K. (2012). Multiple routes of pesticide exposure for honey bees living near agricultural fields. PLoS One 7(1). doi: 10.1371/journal.pone.0029268

- Andino, G.K. and Hunt, G.J. (2011). A scientific note on a new assay to measure honeybee mite-grooming behavior. Apidologie 42: 481-484. doi: 10.1007/s13592-011-0004-1

Abstracts

- Andino, G.K., Caceres, V.A., Giraldo-Calderón G.I., Prado J.K., Raje, K.R. and Van Zee, J.P. (2010). 2008 Student debate: The use of genetically modified organism in entomology. Nielsen, A.L. and Burrus, R.G. (eds.), Am. Entomol. 56 (2): 107-108.

- Andino, G.K. and Hunt, G.J. (2010). A new assay to measure mite-grooming behavior. *In* Proceedings of the 2010 American Bee Research Conference (ABRC-23st). American Bee Journal 150 (5): 497-511.

Undergraduate Thesis

- Andino, G., Doyle, M., and Rueda, A. (2002). Estudio del modo de transmisión de la enfermedad de la Hoja Pequeña de Gliricidia sepium causada por un fitoplasma (The Transmission Mechanisms of Little Leaf Disease in *Gliricidia sepium* caused by a Phytoplasma) http://bdigital.zamorano.edu/handle/11036/2213#sthash.VHnniBnz.dpuf)

SCHOLARSHIPS AND HONOR AWARDS

- Oser Family Scholarship 2014 in the amount of $1500, Entomology Department, Purdue University, West Lafayette IN.

- Travel expenses to present research results at the Workshop: Understanding the Apis-Varroa interactions; insights for improving bee health. Cairns, Australia, July 11-12, 2014.

- NCB-ESA Student travel award to attend the 2014 North Central Branch ESA meeting, Des Moines, IA, March 9-12, 2014. (Award declined).

- Gerald Leeb Scholarship 2013 in the amount of $1000, Entomology Department, Purdue University, West Lafayette IN.

- Tuition scholarship and travel award to attend 18th Summer Institute in Statistical Genetics (SISG 2013). University of Washington, Seattle, WA July 15, 2013 - July 26.

- Scholarship granted by the i5K fellowship committee to attend and present research work at the i5K Community Workshop. Kansas City, MO. (May - June, 2012).

- Certificate of appreciation for service to the Department of Entomology at Purdue University. (December 2011).

- Scholarship granted by the Foundation for the Preservation of Honey Bees to attend and present research work at the 2010 North American Beekeeping Conference, a joint meeting of the American Beekeeping Federation, the Canadian Honey Council, the Apiary Inspectors of America, the Canadian Association of Professional Apiculturists, and the American Association of Professional Apiculturists. (2009).

- Recognition for participating in the Diagnostics of Pests of Quarantine Importance in Honduras and the Region of OIRSA given by the Secretary of Agriculture and Cattle (SAG) and the National Service for Health Agriculture (SENASA) 2005.

- Recognition for most outstanding performance during annual trainings granted by the International Regional Organization for Plant and Animal Health (OIRSA) 2004.

- Scholarship granted by the International Science Foundation for thesis project "The Transmission Mechanisms of Little Leaf Disease in *Gliricidia sepium* caused by a Phytoplasm" (2002).

- Studies in Zamorano: Scholarship granted by the Secretary of Agriculture of the United States, through the program Food For Progress. (1999 - 2002).

- Studies in Zamorano: Scholarship granted by The Secretary of Agriculture and Cattle (SAG), 1999-2002.

MEMBERSHIPS

- Studies in Zamorano: Scholarship granted by The Secretary of Agriculture and Cattle (SAG), 1999-2002.

- American Association for the Advancement of Science (AAAS) 2012-2014.

- American Association for the Advancement of Science (AAAS) 2012-2014

- American Genetic Association and Journal of Heredity (2012).

- Entomology Graduate student Association (EGO) 2009 – Present.

- Entomology Graduate student Association (EGO) 2009 – Present

- Association of Zamorano Alumni (AZA) at Purdue University, 2009 – present.

  - President (2013)

  - Treasurer (2011)

- Children International Organization (2008- Present).

VOLUNTEER WORK

- Winterization project. I assisted in preparing the yards and homes of the elderly and disabled in Tippecanoe County for winter, in teams of 8 to 10 people. (November 2010, 2011 and 2013).

- Springification project. I assisted in preparing the yards and homes of the elderly and disabled in Tippecanoe County for spring, in teams of 8 to 10 people. (April 2013).

- Purdue Spring Fest Bug Bowl. I assisted the Department of Entomology at Purdue University on honey bees specimens and equipment demonstration and honey tasting event. (Spring 2009, 2010, 2011 and 2013).

- Professional oral presentation judge at the 10th annual Undergraduate Research Poster Symposium. Purdue University, West Lafayette IN. (April 2012).

- Professional oral presentation judge at the 2011 Summer Undergraduate Research Fellowships (SURF) Symposium. Purdue University, West Lafayette IN. (September 2011).

- Tippecanoe County Butterfly Encounter. I assisted with the identification of the common butterflies of Indiana and answering questions that the attendees had. (July 2011 and 2012).

- Department of Entomology Seminar Series Committee Chair. I assisted preparing the Department Seminar speakers schedules and setting student lunch with guest speaker. (2010 to present).

- Purdue Entomology Insectaganza. I assisted the Department of Entomology at Purdue University giving a demonstration to fifth grades visiting campus, how to dissect a grasshopper. (2008).

EXTRACURRICULAR ACTIVITIES

- Attend Upper Room Fellowship Christian Church, West Lafayette, Indiana

- Gardening

- Knitting and Crocheting

- Sewing

LANGUAGES

- -Spanish-native language.

- English-writing and speaking fluency.

SOTFWARE AND PROGRAMMING LANGUAGUES

- Working Knowledge

    - Perl, Unix, R-bioconductor

    - Next Generation Sequencing Software (Trinity, RSEM, Bowtie, Samtools, GATK, BWA, and IGV)

- Basic Knowledge

    - MySQL