2-2021

Insights into *Leptopilina* Spp. Immune-Suppressive Strategies Using Mixed-omics and Molecular Approaches

Brian Wey
The Graduate Center, City University of New York

INSIGHTS INTO LEPTOPILINA SPP. IMMUNE-SUPPRESSIVE STRATEGIES USING MIXED-OMICS AND MOLECULAR APPROACHES

by

BRIAN WEY

A dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2021

INSIGHTS INTO LEPTOPILINA SPP. IMMUNE-SUPPRESSIVE STRATEGIES USING MIXED-OMICS AND MOLECULAR APPROACHES

by

BRIAN WEY

This manuscript has been read and accepted for the Graduate Facility in Biology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy

_____	_____
Date	Sanna Goyert, Ph.D.
	Chair of Examining Committee
_____	_____
Date	Cathy Savage-Dunn, Ph.D.
	Executive Officer

Advisor: Shubha Govind, Ph.D.

Supervisory Committee:

Karen Hubbard, Ph.D.

Shaneen Singh, Ph.D.

Keith Hopper, Ph.D.

The City University of New York

ABSTRACT

INSIGHTS INTO LEPTOPILINA SPP. IMMUNE-SUPPRESSIVE STRATEGIES USING MIXED-OMICS AND MOLECULAR APPROACHES

by BRIAN WEY

Advisor: Shubha Govind, Ph.D.

Host-parasite interactions influence the biology of each over the course of evolution. Parasite success allows for the passage of potent virulence strategies from generation to generation. Host success passes stronger immunity and resistance strategies to the following generations as well. Only by studying both partners within their natural contexts can we begin to understand the relationship between the two and how immune mechanisms and virulence strategies interact as a molecular arms race.

In this work, we focus on a natural host-parasite pair, the *Drosophila-Leptopilina* model. *Leptopilina* species are parasites of several fruit fly species, including *Drosophila melanogaster*. This model offers many advantages, including the well-annotated *Drosophila* genome, the genetic and transgenic tools available for *Drosophila*, and the ease of culturing these insects in the lab.

The wasps *Leptopilina heterotoma*, *Leptopilina boulardi*, and *Leptopilina victoriae* utilize somewhat distinct immune-suppressive strategies to suppress the host defenses. These are mediated in part through virus-like particles (VLPs) that are injected into the host larva along with the wasp egg. VLPs target hosts' blood cells. The primary *Leptopilina* species studied within this work is *L. heterotoma*, a wasp that is successful against a wide range of *Drosophila* species. Its infection leads to the destruction of almost all larval blood cells, blocking encapsulation and antimicrobial peptide production in *D. melanogaster*. *L. heterotoma*'s VLPs contain ~400 proteins; some are predicted to carry out conserved cellular processes or may modulate the immune response. To understand the nature of VLPs and *L. heterotoma*'s unique immune suppression strategy, this work has utilized mixed-omics, transgenic, and RNA interference (RNAi) approaches.

In Chapter 1, we characterized the expanded particle proteome, the whole-body transcriptome, and sequenced the *L. heterotoma* genome to search for evidence of virus-related proteins within the

particle's expanded proteome. We found no viral coat proteins within the expanded particle proteome. In addition, more than 90% of VLP proteins had coding regions within the wasp genome.

In an Addendum to Chapter 1, we compared the recently published *L. boulardi* VLP proteome to that of the *L. heterotoma* VLP/MSEV proteome to better understand the similarities of both particles and their extracellular vesicle character. While both wasps only share approximately 30% of their particle proteins, the overall extracellular vesicle profile and distribution of proteins across classes is similar. A search of highly assembled wasp genomes improved our published assessment of the proportion of *L. heterotoma* MSEV proteins that are encoded by the wasp genome. This observation also held for the *L. boulardi* VLP proteins, reinforcing the extracellular nature of these particles.

In Chapter 2, we focus on the virulence function of a key *L. heterotoma* spike protein, p40. In transgenic expression, a full-length construct of p40 localized to cell membranes while a truncated construct (without the putative transmembrane domain) was found to be secreted, supporting previous structural predictions of p40. The secreted protein prevented encapsulation of eggs of a closely related wasp *L. victoriae*, which normally elicits strong encapsulation in *D. melanogaster*. RNA interference-mediated knockdown of p40 significantly reduced *L. heterotoma*'s ability to suppress encapsulation as a larger proportion of p40RNAi-infected wild type hosts showed strong encapsulation. Together, these results underscore the importance of p40 in *L. heterotoma*'s ability to prevent encapsulation and ensure offspring survival.

Parasitic wasps are keystone species and some species are utilized to control agricultural pests. By providing genomics, transcriptomics and proteomics analyses, this work expands the utility of the *Drosophila-Leptopilina* model; the released data will facilitate studies in novel areas of host-parasite biology. It also illuminates virulence strategies and the evolution of virulence factors of these wasps.

ACKNOWLEDGEMENTS

This dissertation could not have been completed without the love and support of my mother and father. Thank you for continuing to believe that I could make it this far and for not asking when I am expected to graduate after my third year in the program. Thank you to Dustin Zuelke and Zully Santiago for being the best cohort members, friends, and fellow scientists I could ask for. A special thank you to all my non-graduate student friends for reminding me that there is a life outside the lab and living it with me: Audrey, Rich, Kathleen, Victor, and Andrew.

I would like to extend another thank you to my mentor and advisor, Dr. Shubha Govind, for taking a chance on me and teaching me how to find my way on the path to becoming a good scientist. Another extended thank you to my committee members Dr. Sanna Goyert, Dr. Karen Hubbard, Dr. Shaneen Singh, and Dr. Keith Hopper for their wisdom, guidance, and experience.

A great thank you to Dr. Johnny Ramroop and Dr. Mary Ellen Heavner for teaching me about flies, the lab, and for laying the groundwork for the projects within. Another thank you to all the other members of the Govind Lab that I have had the pleasure of working with. Jennifer, Zubi, Joyce, Kushagra, Thuriah, Jackson, Brandon, Alitzel, Lady, Carlo, and Riyami, thank you for making each day interesting, different, and a constant learning experience.

In Chapter 1, Dr. Keith Hopper and Dr. Kameron T. Wittmeyer assisted with many portions including additional genome assembly, KAT analysis used in Figure 2 and Supplemental Figure 1, and helpful discussions. Sequencing of *L. heterotoma* VLP proteins and enrichment analysis was performed by Dr. Mary Ellen Heavner. Initial sequencing of *L. heterotoma* genome was performed by Dr. Thomas Briese. Bioinformatics work was conducted in-house and with the BIOMIX Shared Computing Cluster at Delaware Biotechnology Institute, University of Delaware. This chapter was published in G3: Genes, Genomes, Genetics under the title “Immune suppressive extracellular vesicle proteins of Leptopilina heterotoma are encoded in the wasp genome.”

The *Leptopilina boulardi* VLP protein sequences analyzed in an Addendum to Chapter 1 were provided by Dr. Julien Varaldi (University of Lyon, France). Bioinformatics work was conducted in-house

and with the BIOMIX Shared Computing Cluster at Delaware Biotechnology Institute, University of Delaware.

In Chapter 2, initial cloning of p40 and design of original p40 primers was performed by Dr. Johnny Ramroop. Western blot in Figure 1 was performed by Carlo Sevilla. Gateway destination vectors for transgenic flies were obtained from the Drosophila Genomics Resource Center, supported by NIH grant 2P40OD010949-10A1. Western blot in Figure 1 was performed by Carlo Sevilla. Tumor burden assay and viability assay performed by Carlo Sevilla and Kushagra Vashist. Asif Siddiq helped with the preparation of figures.

All chapters within this work were supported by the following grants from NASA (NNX15AB42G), NSF (IOS-1121817, IOS-2022235), NIH (1F31GM111052-01A1, 5G12MD007603-30, and GM103446).

CONTENTS

LIST OF FIGURES

INTRODUCTION:

CHAPTER 1:

ADDENDUM TO CHAPTER 1:

CHAPTER 2:

LIST OF TABLES

LIST OF ABBREVIATIONS

15.2: Recombinant 15.2

BV: Bracovirus

Cg: *Collagen*

CyO: Curly of Oster

da: *daughterless*

EVE: Endogenous viral element

FM7: First Multiple 7

GFP: Green fluorescent protein

$hop^{Tum\text{-}L}$: $hopscotch^{Tumor\text{-}lethal}$

IV: Ichnovirus

Lh: *Leptopilina heterotoma*

Lb: *Leptopilina boulardi*

LbFV: Leptopilina boulardi Filamentous Virus

Lv: *Leptopilina victoriae*

mCD8: Murine Cluster of Differentiation 8

mCD8-GFP: Murine Cluster of Differentiation 8-green fluorescent protein

mRFP: Monomeric red fluorescent protein

MSEV: Mixed strategy extracellular vesicle

msn: *misshapen*

myr-mRFP: Myristoylated monomeric red fluorescent protein

PBS: Phosphate buffered saline

PDV: Polydnavirus

RFP: Red fluorescent protein

v: *vermillion*

VLP: Virus-like particle

w: *white*

y: *yellow*

INTRODUCTION

Parasitioids and their Hosts

Heterologous animal-animal interactions, and even animal-bacterial interactions, are often studied at many different levels, including cost-benefit. Such studies have revealed examples and mechanisms of not only predation (feeding on prey species) and pathogenicity (process and progress of infection by microbes), but also symbiosis, or how the two parties live together. Symbiosis is further divided into three categories: parasitism – where only one party benefits, mutualism – where both parties benefit equally, and commensalism – where neither party is affected positively or negatively (reviewed by Leung et al. in (*1*)). All three divisions of symbiosis can be observed within the class Insecta.

Commensalism is usually observed in insect-bacterial interactions, such as the association of the intracellular bacterial genus *Wolbachia* with many insects (*2*). Commensal relationships can also be observed between two insect species, for example, species of foraging ants which follow each other's chemical trails to food (*3, 4*). In general, commensalism appears to be a tolerance of interaction without a very large cost or gain on fitness. Other works have also delineated that animal hosts may adapt to parasites such that the burden of parasitism is not as severe (implying a lesser penalty on fitness and survival), eventually leading to what Leung et al. call a "superficial" commensalism (*1, 5*).

Mutualism is observed in the relationship between particular species of aphids and ants (reviewed by Stadler et al. in (*6*)). Aphids are parasites of plant species because they consume sap which directly harms the growth of the plant host; however, this method of feeding produces an excrement called honeydew (*6*). Ants collect honeydew from aphids as food (*6*). In return, depending on the species, ants tend, cultivate, and/or protect aphids from their predators and parasites to ensure their food supply of honeydew from the aphids (*6*). In some cases, aphids will tolerate damage to their wings from ants as part of the relationship (*6*).

As for parasitism, insects can be parasites (as in the blood-feeding mosquito *Aedes aegypti* (*7*)) and/or become parasitized (such as fruit flies in the genus *Drosophila* by Plasmodium, mites, nematodes,

and wasps *(8-10)*). Classical insect parasites feed on their host, but may not directly kill them (*1, 10*)), whereas parasitoids, kill their host.

Parasitoids are insects which generally attack juvenile (embryonic, larval or pupal) stages of the host, although adult hosts can also be attacked and parasitized. They feed on host tissue and consume the entire host (usually in the case of insect hosts) (*11, 12*). The host serves as the primary food source for the parasitoids' larval stages and may be the sole major feeding the parasitoid does in the life cycle (*11, 12*). Insect parasitoids can be further divided into ectoparasitoids (parasitoids that develop outside the host) or endoparasitoids (parasitoids that develop within the host) (*11, 12*). Furthermore, ectoparasitoids tend to be idiobionts, which arrest or stop the growth of the host in order to prevent dislodging or death of the parasite from host movement. Whereas endoparasitoids are koinobionts which do not pause host growth (*12, 13*). Both kinds of parasitoids typically possess a dual life cycle, where the larval stage is the parasitoid, feeding on a host, while the adult stage is free-living (*12*).

Oviposition by Hymenopteran parasitoids involves laying their eggs inside a host by injecting them into the host body with a needle like structure called an ovipositor (*12*). In addition to the egg, the wasps introduce factors to counter the hosts' immune defense mechanisms: venom proteins, virus-like particles (VLPs), or domesticated viruses (*8, 9, 12, 14-31*). Successful development of the parasitoid offspring depends on the success of the adult female's strategy in using the above mechanisms to suppress the host immune response to prevent death of offspring. Further obstacles to parasitization include host behavioral immunity such as methods of self-medication via alcohol-containing substrate feeding, decrease in egg laying and production, and even rolling to avoid oviposition (*32-35*). Understanding this biological arms race-- between host defense and parasite offense, within a natural context, is vital for studies in both immunity and virulence/pathology as it can allow further dissection of how immune and virulence strategies arose. Such studies can also help identify species which may fill similar ecological niches and can be used to reduce the threat of invasive species.

This dissertation focuses on the interactions between *Drosophila* host species and wasp parasitoid species using the Leptopilina-Drosophila host-parasite model.

The Leptopilina-Drosophila Host-Parasite Model

The wasp genus *Leptopilina* is within the family Figitidae and contains at least seven species (*36*). Wasps of this genus parasitize fruit flies of the genus *Drosophila* (*9, 36-38*). Of these, three *Leptopilina* species are well characterized in studies of host-parasite interactions and immune suppression strategies: *Leptopilina heterotoma* (*Lh*) (Fig 1), *Leptopilina boulardi* (*Lb*), and *Leptopilina victoriae* (*Lv*) (*36*). Experiments on host range and success of parasitism (as defined by percentage of wasp offspring emerging out of total number of infections), *Lh* is considered to be a generalist parasite, capable of parasitizing many different *Drosophilia* species with high success, and *Lb* is considered to be a specialist parasite, highly successful only on flies of the melanogaster group (Fig 1) (*9*).

Lh, *Lb*, and *Lv* share similar developmental stages and the most extensive observations on developmental timing were made by Kopleman and Chabora in *Lb* with some observations made in *Lh* by Small et al. and are described in part here (*39, 40*). Usually, females chemically detect that a host has been parasitized; however, superparasitism (multiple instances of oviposition by the same female within one host) can still occur in the wild. In some strains of *Lb*, superparasitism is associated with a behavior-altering virus known as Leptopilina boulardi Filamentous Virus (LbFV) (*41-44*). The early stage wasp larva hatches and feeds on host hemolymph for two to three days until the host pupariates (*39, 40*). In the instance of superparasitism, one wasp larva dominates, while the development of the supernumerary individuals is arrested and they are killed, allowing for development of one parasitoid in each host (*9, 45*). The mechanisms underlying wasp selection, developmental arrest and the elimination of the animals are not understood. Developing *Leptopilina* wasps then feed on the larval hemolymph as endoparasites and eventually on the prepupal and pupal host stages, until the entire host is consumed, at around ten days post oviposition (*39, 40*). Utilizing the host pupal case, one wasp grows, finishes development and ecloses as a free-living adult from the fly's pupal case instead of the adult fly (*39*) (Fig 1).

Anti-parasitoid Immunity in *Drosophila*

The major anti-parasitoid response in Drosophila is the encapsulation of wasp eggs, or of early wasp larvae, by the host blood cells called hemocytes. Wild *D. melanogaster* live on rotting fruit and are subject to infection by microbes (bacteria, viruses, and fungi) and attack by parasitoid wasps.

Immunological responses to both microbes and metazoan parasites are initiated by the two branches of innate immunity, the cellular branch and the humoral branch, each controlled by highly conserved NF-kB signaling (*46-51*). Toll, initially discovered as part of the maternally-encoded embryonic dorsal-ventral patterning pathway, also plays a role in both pathways in response to parasitism (*52*). The cellular branch is executed by three hemocyte types: the macrophage-like plasmatocyte, the lamellocyte involved in encapsulation and melanization, and the crystal cells that carry out melanization and wound healing (*49, 51*). The cellular immunity is responsible for phagocytosis of microbes and formation of capsules around juvenile parasite stages (*29, 40, 47, 48, 53-57*).

Fly hemocytes are produced from three tissues. The initial larval circulating hemocytes are produced from embryonic mesoderm (*49, 51, 58*). Hemocytes produced in response to immune challenge are produced from the anterior lobe of the immune sensing organ called the lymph gland (LG) and from circulating precursors (*50, 58, 59*). The anterior lobe possesses a cortical zone (CZ) of mostly differentiated cells surrounding a medullary zone (MZ) of progenitor cells kept in quiescence by the posterior signaling center (PSC) (*50*). The posterior lobes of the LG contain progenitor cells which will become hemocytes for the adult fly stage (*50*). A third population of hemocytes exists as sessile, or non-circulating, and are largely found in sub-epidermal compartments (*60*). These sessile cells originate from the circulating hemocytes and are maintained in a niche created by the peripheral nervous system (*61*). The Toll/NF-kB pathway is activated in the PSC upon parasitization and react to reactive oxygen species (ROS) produced in the LG, particularly by the PSC niche, to cause cell differentiation (*52, 62*).

Plasmatocytes are the predominant hemocyte population in unchallenged *Drosophila* larvae (*49, 51*) (Fig 2). They circulate in the hemolymph and phagocytose bacteria and dead cells (*49, 51*). Plasmatocytes can also be found in the sessile compartments (*60*). Lamellocytes are not normally seen in circulation in high numbers in *D. melanogaster* larvae until after parasitism or nematode infection and are usually produced in two ways: (a) already differentiated plasmatocytes transdifferentiate into lamellocytes; or (b) and progenitor cells, found in the lymph gland and in circulation, differentiate into lamellocytes (*51, 63-67*) (Fig 2). Some lamellocytes then adhere to parasite eggs and embryos to encapsulate and kill the invading parasite (*19, 63, 65-70*). Crystal cells exist as both circulating, LG, and sessile populations and

possess enzymes involved in melanization and production of reactive oxygen species in order to form melanized clots in response to wounding (*49, 51*).

The encapsulation reaction responsible for killing wasp eggs and larvae is similar to granulomas in mammalian responses to bacterial species that that evade phagocytosis (for example *Mycobacterium*) and has been observed in multiple insects and even invertebrates in general in response to pathogens and parasites that cannot be phagocytosed (*56, 57, 71-74*). As species of *Drosophila*, and many stem-boring Lepidopteran species, are considered agricultural pest species, research into the encapsulation response is extensive as the search for the ideal biocontrol for these pests continues (*32, 53, 75-80*). The process of encapsulation requires the following steps: recognition (usually after a damage signal is released), adhesion, spreading (hemocytes flatten themselves along the foreign body), and degranulation (release of enzymes involved in melanization in order to kill the foreign body) (*72*).

Typically, the humoral arm of *Drosophila* immunity responds to bacterial infection by producing antimicrobial peptides from the fat body through the Toll and/or Imd pathways (Gram positive and negative bacteria respectively) (*46-48, 81*). *D. melanogaster* possesses nine different Tolls and the various receptors respond to different immune challenges (*82-84*). For example, Toll-7 responds to viruses (*85*) and Toll-8 is seen to modulate the immune response in trachea (*86*). However, the humoral response is also activated upon wasp parasitism (*52, 87*). Production of Spätzle by cells in the LG or circulating hemocytes occurs without immune challenge (*87, 88*). Production of Spätzle processing enzyme in response to wasp parasitism, activates Toll signaling, causing the release of Dorsal and Dorsal-related immune factor (the fly NF-kappa B proteins) from I-kappa B cognate protein, Cactus. This activation step promotes hemocyte proliferation, egg recognition and encapsulation (*52, 87*). Given the central roles of this conserved pathway in both the cellular and humoral responses and inflammation, modification of this pathway through various means or depletion of the hemocytes themselves, inflammatory cells, thus appear to be crucial to parasitoids to ensure offspring survival.

Immune suppression by *Leptopilina*

Research into the mechanisms used by *Lh* to successfully suppress the host immune response started with the Rizkis in 1984 by examining the effects of venom gland components on hemocytes (*89*).

Lh, *Lb*, and *Lv* each possess a distinct venom gland composed of a long gland, connecting duct, and reservoir (*90-92*). All three wasps produce venom proteins within the long gland as well as discrete particles that are secreted in an immature form via actin-lined canals into the long gland lumen. The particles mature and develop their spikes, required for interaction with host cells, as they move to the reservoir (*29, 90, 92-95*). These particles were called “virus-like” particles (VLPs) and are linked to parasite success (*29, 93*). However, the morphology and variability of size and structure of mature VLPs, and their step-wise biogenesis and maturation suggest that the nature of these particles is non-viral. If the immune suppression caused by the VLPs is not successful, the host immune response activates. Lamellocytes aggregate to encapsulate (formation of a physical barrier/sheath around a foreign body/object) and melanize (generation and deposition of melanin as part of the process to produce reactive oxygen species) wasp eggs or larvae in the larval hosts’ body cavity to kill it (*63, 72*) (Fig 1). Because VLPs are linked to parasite success, differences in the proteins within them might indicate somewhat different suppression strategies (*96-98*). In addition to immune evasion, *Lb* suppress host fly immunity either by influencing blood cell shape or function (*8, 54, 64, 94*). However, the continued presence of hemocytes appears to allow for induction of humoral response (*9, 87*). *Lb* eggs also tend to adhere to host tissue, allowing for a potential gap in the capsule from which the wasp larva may escape from (*8*). In contrast, *Lh* parasitism causes cell death in hemocytes and prevents activation of the host humoral response (*8, 9, 29, 54, 89, 90, 93, 94, 99*). Lamellocytes die from entry of *Lh* VLPs through an unknown mechanism, eventually inducing lysis (*29, 30*). Phagocytosis of *Lh* VLPs by plasmatocytes/macrophages results in terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL)-positive cell death (*29, 30, 100*). These two distinct cell-killing effects of *Lh* VLPs on lamellocytes and macrophages are not observed after *Lb* attack of *D. melanogaster* (*8, 9, 64, 94*). Although *Lv* appears to be much more closely related to *Lh*, *Lv* parasitism still causes capsule formation through potential failure to manipulate host immune signaling (*87, 90, 91*).

The Biotic Nature of VLPs and their Potential Origins from Domesticated Insect Viruses

The strategy of using VLPs to suppress the host immune response is not unique to *Leptopilina spp*. and the origins and nature of VLPs and functionally related secretion products is a rapidly developing

topic in the study of parasitoid wasps. Further enriching this discussion are wasps that have domesticated viruses for use in their immune suppression strategies.

Parasitic wasps in the superfamily Ichneumonoidea diverged more than 20 million years ago from Figitid wasps (like *Leptopilina*) (see Fig 3 for phylogenetic tree). Wasps of this superfamily parasitize a diverse range of host arthropods, and have symbiotic relationships with insect viruses under the family Polydnaviridae (*14, 15, 17, 101-107*). These wasps fall into two subfamilies: Braconidae and Ichneumonidae (*101, 108-110*). Accordingly, Polydnaviruses (PDVs) are divided into two genera, Bracovirus (BV) and Ichnovirus (IV) (*101, 111, 112*). BVs are descended from insect nudiviruses (a family of large rod-shaped, enveloped viruses with large circular genomes) and IVs are thought to have descended from large cytoplasmic DNA viruses (potentially in the family ascoviridae) (*15, 103-106, 108, 110, 113-117*). BV and IV genomes are integrated into the wasp genome and particles are produced in calyx cells found around and supporting the wasp ovary. PDVs contain circular dsDNA, and are released from calyx cells via lysis (BV) or budding (IV) (*101, 112, 117, 118*). The PDV particles are injected into the host during egg laying where they infect host cells and express viral genes to suppress host immunity (*14, 101, 102, 117, 119, 120*).

Although BV and IV genomes are integrated into these wasp lineages and reliant on wasp success for continued survival, this integration is not always perfect, and some genes may be lost over evolutionary history. Some lineages have either lost the viral capsid or produce incomplete capsids. These species then use other viral genes to produce a viral envelope for transmission of other effector proteins (*109*). For example, *Fopius arisanus,* produces incomplete capsids that fail to properly package DNA but still include virulence proteins (*110*). Another wasp, *Venturia canescenes*, produces unencapsidated immune-suppressive VLPs surrounded by a lipid envelope. These *V. canescenes* particles are thought to be of alphanudiviridae origin as 21 nudiviral genes encode the viral envelope components and proteins for VLP biogenesis (*109*).

Viral integration does not appear to be unique to ichnoviruses and bracoviruses. Using predictive methods trained on published protein data from several species of Ichneumonid and Braconid wasps and their associated viruses, Burke et al. (*121*) and located sets of the “ancient core” genes for IV, BV, and

even for the *Leptopilina boulardi* Filamentous Virus (LbFV) in several Ichneumonid wasp genomes (*121*). LbFV manipulates wasp behavior but is not known to be immune suppressive (*42, 122, 123*).

To understand the biological nature of *Leptopilina* VLPs, their mechanisms of action, and determine their possible relation to PDVs, Heavner et al. characterized VLP proteomes from two *Lh* strains (Lh14 and LhNY) and identified a shared set of 161 proteins in three classes: Class 1 – conserved eukaryotic, Class 2 – infection and immunity related, and Class 3 – unannotated proteins (*96*). They also examined if these *Lh* proteins were expected to be present in *Lb* and a more distantly related wasp *Ganaspis hookeri.* Several similar proteins in Class 1 and 2 proteins were found in both *Lb a*nd *G. hookeri*, but very few proteins in Class 3 were shared by all three wasps (*96*). The proteome also did not contain known viral coat proteins but contains several vesicular transport and endomembrane system proteins (considered under Class 1) (*96*).

Heavner et al. also found several interesting gene families in the *Lh* VLP proteome such as a family of prokaryotic-like GTPases and a family of RhoGAPs (*96*). These results suggested that *Lh* VLPs may not be “virus-like”, but instead are more akin to specialized eukaryotic vesicles for transport of virulence factors from parasite to host. Their variable morphologies and the stepwise biogenesis process of VLPs also support their organelle nature (*96*). More interesting, these GTPases and a highly produced protein called ‘p40’ are absent from *Lb* VLPs, implying a difference in strategy may be due to differences in proteomic profile (*96*).

In addition, in collaboration with Dr. T. Briese (Columbia Univ.) our lab characterized total nucleic acids from gradient-purified *Lh* VLPs. This analysis did not support the presence of an independent VLP “genome” (Govind and Briese labs, unpublished). Thus, Heavner et al. proposed the renaming of *Lh* VLPs to Mixed-Strategy Extracellular Vesicles (MSEVs) (*96*).

These initial results of (a) similar organizations of their venom gland morphologies (*92*), (b) homologies between proteins and transcripts between *Lh* and *Lb* (*96*), and (c) morphological similarities in mature MSEV/VLP particle suggested that *Lb* VLPs share the same biotic nature as *Lh* MSEVs. Recent studies in *Lb* VLPs have approached this topic in two different directions. Wan et al. (*95*) have published data that supports the idea of *Lb* VLPs as extracellular vesicles due to their biogenesis and have dubbed them “venosomes”. Published proteomic analysis of *Lb* VLPs also indicates several proteins

that are expected to be highly involved in immune suppression (serpins, superoxide dismutase, and calreticulin), as well as the important virulence protein LbGAP (*95, 98*). However, the idea that these structures have a viral origin cannot be ruled out.

LbFVs and their proposed connection to VLPs/MSEVs

The idea that *Lb* VLPs have derived from insect viruses has been proposed by Di Giovanni et al. (*98*). Following previous studies of the behavior-altering virus LbFV (known to induce superparasitism in *Lb*), Di Giovanni et al. found 13 viral ORFs related to LbFV's genes integrated into the *Lb* genome (*41, 98, 124*). These ORFs were also found to be integrated into the *Lh* and *L. claivpes* genomes (*98*). Evidence of at least one ORF (called ORF 96) was found in three other *Leptopilina* species (*98*). Molecular experiments indicate expansion of copies and increased transcription of these viral ORFs in the *Lb* venom gland as wasp larvae mature into adulthood (*98*). However, there is no evidence that these ORFs are required for *Lb* VLP biogenesis. After a proteomic analysis, Di Giovanni et al. also report that at least two of these viral proteins are found in the *Lb* VLP proteome (*98*), whereas the *Lh* MSEVs proteome published by Heavner et al. (*96*) contains a homolog of one of the two LbFV ORFs.(*98*). The viral ORF found within *Lh* MSEVs and *Lb* VLPs (LbFV ORF 85) is annotated as Ac81, expected to be involved in nucleocapsid envelopment for *Autographa californica* nuclear polyhedrosis virus (*125*). Even though we identified the same 13 ORFs integrated into the *Lh* genome in our studies, our search criteria did not identify significant similarity to any of the LbFV proteins in the expanded proteome analysis (*97*). Thus, the links between LbFV ORF-derived proteins and VLP/MSEV production or function are tenuous and further work is needed to examine the functions of these viral ORFs in the wasps' venom.

p40, a putative spike protein of Lh VLPs

In previous studies of the *Lh* venom and MSEVs, our lab identified a highly expressed protein, called 'p40' (*91*). Immuno-gold electron microscopy revealed that p40 to the surface and their spikes and p40. Notably, a p40-like protein is also found in the sister wasp species *Lv* (*91*). Microsequencing of the p40 protein and comparing its peptide sequences to the *Lh* MSEV transcriptome helped identify the coding sequence of p40 (*22, 96*). At the time of its publication, there were no proteins with high primary

sequence similarity to p40 (*22, 96*). Further bioinformatics analyses predicted that p40 possesses a signal peptide, a central domain, and a transmembrane domain (*96*).

The central domain of p40 was modeled and is expected to fold similarly to needle tip proteins, IpaD and SipD of bacterial Type Three Secretion Systems (T3SS) (*96*). T3SS are used by Gram negative bacteria like *Shigella* and *Salmonella* to evade or infiltrate host intestinal cells and macrophages and modify host cell behavior and viability (*126-128*). The similarities in p40 and IpaD/SipD locations and structures may indicate a similarity in function (*96, 97*). A comparison of IpaD (*Shigella spp*. (*129, 130*)), SipD (*Salmonella spp*. (*131, 132*)), and BipD (*Burkholderia spp*. (*133*)) activities with p40 can be found in Table 1. These comparisons will help define the function of p40 in the *Lh* infection strategy in more detail.

Advances in Molecular and Genetic Approaches to Parasitoid Research

To better understand the immune suppression strategies of *Leptopilina spp*., our and other labs have begun genomic analyses and are developing genetic tools for research. Work in other Hymenopteran genomes has revealed much about venom genes as well as the integration of PDV sequences into wasp genomes (*20, 22, 110, 134-138*). Thus, comparative genomics tools can further our understanding of the composition of VLPs/MSEVs and their and relationship with their wasps. Previous studies consisted of transcriptomic studies and estimations on genome size and karyotypes (*22, 24, 25, 139*). The genomes of *Lh* and *Lb* have been recently published, including one from our work on the *Lh* genome, where we reported genes coding for more than 90% of MSEV proteins (*97, 98, 140*).

The sequencing of these *Leptopilina* genomes now allows us to develop better genetic tools for work in this model system. In species where libraries of mutants (for ex. Keio mutant collection in *Escherichia coli* (*141*)) or genetic tools (such as the UAS-GAL4 system, FLP-FRT recombination knock-out system, and P-element insertion mutation libraries in Drosophila (*142-146*)), are not readily available, RNA interference (RNAi) as a means of reverse genetic analysis is a powerful tool. RNAi has already been used in insects other than *Drosophila*, such as the beetle *Tribolium* and multiple Hymenopteran models such as *Apis mellifera* and *Nasonia vitripennis* (*26, 147-150*). Previous studies have successfully used RNAi in *Lb* as well (*151*). However, the difficulty in administration of RNAi in such small endoparasitoid wasps has been survival post injection of dsRNA. Colinet et al. obtained modest survival

rates of 30% after injection (*151*). Feeding of dsRNA to larval insect stages is a possible alternative and has been shown to work in other insect species and is a well-characterized method of administration for *Caenorhabditis elegans* research (*152-158*). Casual observations have seen extended lifespan of wasp adults when provided with honey. Experiments in food preference and foraging have found that adult females with low fat content will actively search out food and seem to prefer honey (*159*). Thus, mixing dsRNA with honey substrate after starvation is a potential avenue of dsRNA administration.

Development of CRISPR-Cas9 in Hymenoptera is also a promising method of gene editing and study. Work has been done in *N. vitripennis* to show that CRISPR does work within the system and can be used to develop stable mutant lines (*160-162*). Continued development of genetic tools in Hymenoptera can only aid in the development and utilization of *Lh* as tools and model organisms.

Description of Work

Within this work, we aim to address two key questions: what is the nature of the MSEVs *L. heterotoma* use to evade the host immune response, and what is the function of the abundant venom protein p40? To address these questions, we sequenced and assembled the genome of *L. heterotoma* and determined if the proteins found in MSEVs are encoded by the genome and not an endosymbiont or part of an unincorporated extrachromosomal DNA. We also expanded our analysis of the MSEV proteome from the Lh14 strain. This analysis did not reveal any viral proteins that may have been missed in the earlier analysis of proteomes shared between Lh14 and LhNY (*96*). These new data were published in *Genes, Genomes, and Genetics* (G3) (*97*) and is presented as Chapter 1.

As part of the rapid development in the field, new genomic and proteomic data has been made available in the last year. We analyzed the *Lb* VLP proteome (as published by Di Giovanni et al. (*98*)) and a compared it directly to the *Lh* MSEV. This analysis shows that *Lb* VLPs and *Lh* MSEVs are similar in enrichment profiles despite only a third of particle proteins being highly similar. We also utilized newly released genomic data to confirm the results of our MSEV protein search within the *Lh* genome. This comparative analysis is added as an Addendum to Chapter 1.

We then experimentally confirmed the prediction that p40 is a membrane protein and determined the importance of p40 in parasite success through use of RNA interference and infection studies using transgenic hosts. This required us to develop protocols for RNAi in *L. heterotoma*. The data and developed methods are presented here in Chapter 2.

Significance

Natural host-parasite interactions are excellent models of study, especially as a means of studying evolution of virulence factors and development of immunity. Studies in *Nasonia spp*. have paved the way in parasitoid research; being larger than *Leptopilina* and of an ectoparasitic life history, makes injections, dissections and molecular analysis easier (*163*). However, *Nasonia* are not as easily cultured due to the requirement of blowfly larval hosts which are grown on rotting meat (*20, 148, 160-162*). In contrast, several species of both Leptopilina and Drosophila are easy to raise, which opens up new possibilities for comparative parasitology studies (*9, 40*). The well-studied *Drosophila* system with numerous genetic lines and tools can be used to better understand the virulence mechanisms of *Leptopilina (142, 143)*.

Understanding the functions of wasp-specific proteins unique to a species and to family members will help understand how *Lh* may have gained a foothold and can parasitize on many *Drosophila spp*. A molecular understanding of species-specificity between the *L. heterotoma-Drosophila* parasite-host partners provides a window on the virulence mechanisms of other wasps of this genus, other parasitic wasps, and even other parasites in general. In addition, developing tools to genetically manipulate these wasps will allow us to begin dissecting the biology of the parasite in more complex ways when used in conjunction with the genetic tools available to *Drosophila*.

Parasitic wasps are of immense agricultural importance in terms of ecosystem impact and more recent invasive pest species, such as *Drosophila suzukii* (*80, 164-166*). In particular, *Lh* and *Ganaspis brasiliensis* are expected to be effective against *D. suzukii (80)*. Furthermore, many other pest species are also parasitized by their own wasp species which are also in consideration for pest control (*75, 78*). Thus, our work will provide a broad understanding of this class of host-parasite interactions and indicate the impact of usage of these wasps as biological pest control.

Figure 1: Flowchart indicating potential outcomes of *L. heterotoma* parasitism of *D. melanogaster*. A female D. melanogaster lays eggs which develop into larvae. (Path A) If the larva is parasitized and manages to activate an immune response, the wasp egg or larva is encapsulated (white arrowhead) and killed. The fly larva then continues development into adulthood and the melanized capsule can still be observed in the adult abdomen (white arrowhead). (Path B) If the wasp is successful in suppressing the immune response, the wasp larvae develops and begins feeding on host hemolymph, moving on to tissue, and eventually emerges as an adult from the host pupal case. Full grown adult *Drosophila* are typically 3 mm long. Fully grown adult *Leptopilina* females are usually less than 2 mm long in comparison. Images are not to scale.

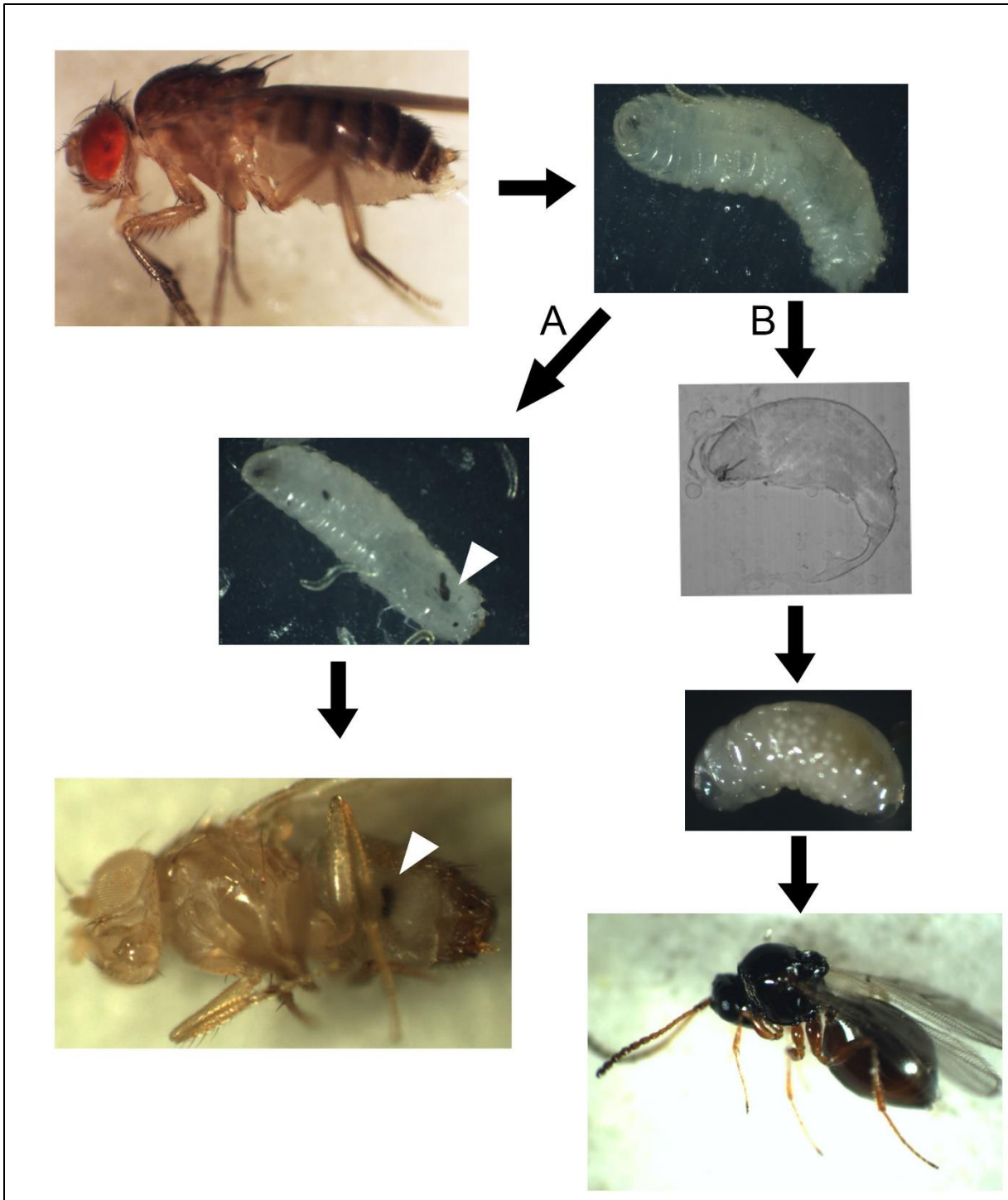

A
B

Figure 2: Representative image of plasmatocytes (yellow arrowheads, smaller macrophage like cells) and lamellocytes (white arrows, large, flat, discoidal cells) from a smear of hemolymph from mutant *Drosophila* larvae (genotype: *y v hop$^{Tum\text{-}L}$/+; msnf9-GFP/Cg-GAL4, UAS-myr-mRFP*). Nuclei are stained with Hoechst (blue); actin is stained with Alexa Fluor 488 (green). Red signal is expression of myristoylated-RFP. Image taken at 40x magnification.

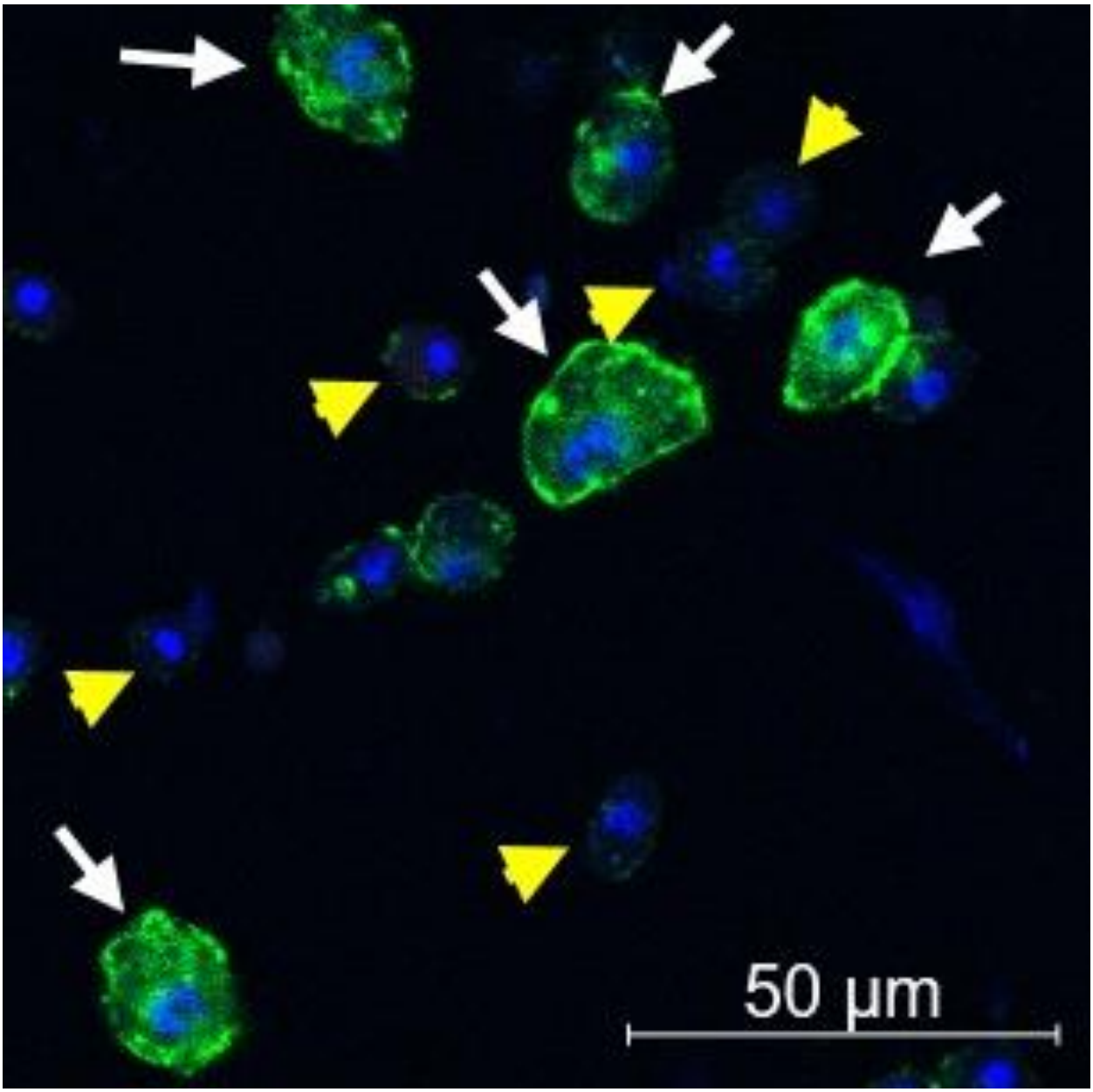


Figure 3: Phylogenetic tree of parasitic wasps and *A. melifera* (latter as outgroup) using cytochrome oxidase 1 protein sequences. Protein sequences were obtained from NCBI. Tree was made using phylogeny.fr workflow containing Mr Coffee, Gblocks, PhyML, and TreeDraw (*167-173*). Bootstrapping was run 100 times. Orange box indicates Figitid wasps (with VLPs) and yellow box indicates Ichneumonid wasps (with polydnaviruses).

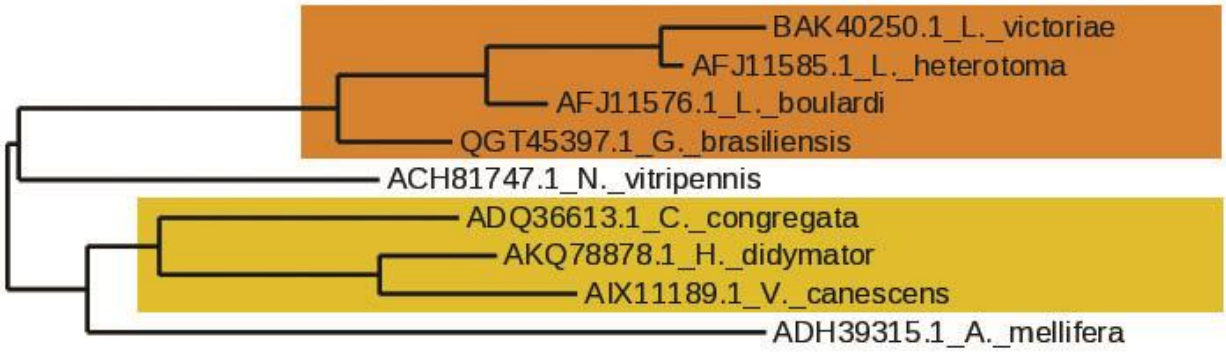


Table 1: Comparison of T3SS needle tip proteins IpaD, SipD, and BipD to wasp protein p40. Citations indicate primary literature where these activities or traits were described.

Table 1: Comparison of T3SS Needle Tip Proteins and p40				
	IpaD	SipD	BipD	p40
Virulence protein in	*Shigella (129, 130, 174)*	*Salmonella (131, 132)*	*Burkholderia (133, 175)*	*L. heterotoma* and *L. victoriae (91)*
Gene location	Plasmid (*129, 130, 174*)	Plasmid (*131*)	Genome (*133, 175*)	Genome (*97*)
Gene has introns	No (*129, 130, 174*)	No (*131*)	No (*133, 175*)	Yes (*97*)
Localizes to tip of needle assembly	Yes (*129, 176*)	Yes (*131*)	Yes (*177, 178*)	MSEV spike tip (localization parallel) (*91*)
Localizes to cell/bacterial membrane	No (*129, 176*)	No (*131*)	No (*177, 178*)	This work, Yes
Key residues required for function	L99, N196, N273, Q277, K291, N292, T296, Q299, Y104 (*179*)	S327, M137 (*180*)	Not known	Not known
Protein is inserted into host cell membrane	No (*176*)	No (*131, 132, 181*)	No (*133*)	This work, Unclear
Needed for insertion of pore proteins into host cell membrane	Yes (*176*)	Yes (*182-184*)	Yes (*133, 175, 177*)	Not known
Needed for injection/secretion of effector proteins	Yes (*176, 179*)	Yes (*132*)	Yes (*177, 178*)	Not known
Causes apoptosis of macrophages	Yes (*128*)	No, SipB does (*185, 186*)	No, BipB does (*187*)	This work, Unclear
Causes changes in cytoskeleton	Has similarity to actin binding proteins (*178*)	No, SipA and SipC (*188*)	Has similarity to actin binding proteins (*178*)	This work, Unclear

REFERENCES

1. T. Leung, R. Poulin, Parasitism, commensalism, and mutualism: exploring the many shades of symbioses. *Vie et Milieu* **58**, 107 (2008).
2. J. H. Werren, L. Baldo, M. E. Clark, Wolbachia: master manipulators of invertebrate biology. *Nature reviews. Microbiology* **6**, 741-751 (2008).
3. B. Gobin, C. Peeters, J. Billen, E. D. Morgan, Interspecific trail following and commensalism between the ponerine ant Gnamptogenys menadensis and the formicine ant Polyrhachis rufipes. *Journal of Insect Behavior* **11**, 361-369 (1998).
4. S. B. Heard, Pitcher-plant midges and mosquitoes: a processing chain commensalism. *Ecology* **75**, 1647-1660 (1994).
5. M. R. Miller, A. White, M. Boots, The evolution of parasites in response to tolerance in their hosts: the good, the bad, and apparent commensalism. *Evolution* **60**, 945-956 (2006).
6. B. Stadler, A. F. G. Dixon, Ecology and Evolution of Aphid-Ant Interactions. *Annual Review of Ecology, Evolution, and Systematics* **36**, 345-372 (2005).
7. S. R. Christophers, *Aedes aegypti: the yellow fever mosquito*. (CUP Archive, 1960).
8. T. M. Rizki, R. M. Rizki, Y. Carton, Leptopilina heterotoma and L. boulardi: strategies to avoid cellular defense responses of Drosophila melanogaster. *Exp Parasitol* **70**, 466-475 (1990).
9. T. A. Schlenke, J. Morales, S. Govind, A. G. Clark, Contrasting infection strategies in generalist and specialist wasp parasitoids of Drosophila melanogaster. *PLoS Pathog* **3**, 1486-1501 (2007).
10. A. Perez-Leanos, M. R. Loustalot-Laclette, N. Nazario-Yepiz, T. A. Markow, Ectoparasitic mites and their Drosophila hosts. *Fly* **11**, 10-18 (2017).
11. P. Eggleton, K. J. Gaston, "Parasitoid" Species and Assemblages: Convenient Definitions or Misleading Compromises? *Oikos* **59**, 417-421 (1990).
12. A. M. C. Santos, D. L. J. Quicke, Large-scale diversity patterns of parasitoid insects. *Entomological Science* **14**, 371-382 (2011).
13. R. Askew, M. R. Shaw, in *Waage, J and Greathead, D (eds), Insect Parasitoids, 13th Symposium of Royal Entomological Society of London*. (Academic Press, London: now Elsevier, 1986), pp. 225-264.
14. G. R. Burke, M. R. Strand, Polydnaviruses of Parasitic Wasps: Domestication of Viruses To Act as Gene Delivery Vectors. *Insects* **3**, 91-119 (2012).
15. E. A. Herniou *et al.*, When parasitic wasps hijacked viruses: genomic and functional evolution of polydnaviruses. *Philos Trans R Soc Lond B Biol Sci* **368**, 20130051 (2013).
16. J. Gauthier, J. M. Drezen, E. A. Herniou, The recurrent domestication of viruses: major evolutionary transitions in parasitic wasps. *Parasitology* **145**, 713-723 (2018).
17. F. Pennacchio, M. R. Strand, EVOLUTION OF DEVELOPMENTAL STRATEGIES IN PARASITIC HYMENOPTERA. *Annual Review of Entomology* **51**, 233-258 (2006).
18. Y. Li, J. F. LU, C. J. FENG, X. Ke, W. J. FU, Role of venom and ovarian proteins in immune suppression of Ostrinia furnacalis (Lepidoptera: Pyralidae) larvae parasitized by Macrocentrus cingulum (Hymenoptera: Braconidae), a polyembryonic parasitoid. *Insect Science* **14**, 93-100 (2007).
19. M. J. Lee *et al.*, Virulence factors and strategies of Leptopilina spp.: selective responses in Drosophila hosts. *Adv Parasitol* **70**, 123-145 (2009).
20. J. H. Werren *et al.*, Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science* **327**, 343-348 (2010).
21. D. Colinet *et al.*, Extensive inter- and intraspecific venom variation in closely related parasites targeting the same host: the case of Leptopilina parasitoids of Drosophila. *Insect Biochem Mol Biol* **43**, 601-611 (2013).
22. J. Goecks *et al.*, Integrative approach reveals composition of endoparasitoid wasp venoms. *PloS one* **8**, e64125 (2013).
23. N. T. Mortimer *et al.*, Parasitoid wasp venom SERCA regulates Drosophila calcium levels and inhibits cellular immunity. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9427-9432 (2013).
24. M. E. Heavner *et al.*, Partial venom gland transcriptome of a Drosophila parasitoid wasp, Leptopilina heterotoma, reveals novel and shared bioactive profiles with stinging Hymenoptera. *Gene* **526**, 195-204 (2013).

25. D. Colinet *et al.*, Extensive inter- and intraspecific venom variation in closely related parasites targeting the same host: The case of Leptopilina parasitoids of Drosophila. *Insect Biochem Molec* **43**, 601-611 (2013).
26. A. L. Siebert, D. Wheeler, J. H. Werren, A new approach for investigating venom function applied to venom calreticulin in a parasitoid wasp. *Toxicon*, (2015).
27. E. O. Martinson, V. G. Martinson, R. Edwards, Mrinalini, J. H. Werren, Laterally Transferred Gene Recruited as a Venom in Parasitoid Wasps. *Mol Biol Evol* **33**, 1042-1052 (2016).
28. E. O. Martinson, Mrinalini, Y. D. Kelkar, C. H. Chang, J. H. Werren, The Evolution of Venom by Co-option of Single-Copy Genes. *Current biology : CB* **27**, 2007-2013.e2008 (2017).
29. R. M. Rizki, T. M. Rizki, Parasitoid virus-like particles destroy Drosophila cellular immunity. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 8388-8392 (1990).
30. R. M. Rizki, T. M. Rizki, Effects of lamellolysin from a parasitoid wasp on Drosophila blood cells in vitro. *J Exp Zool* **257**, 236-244 (1991).
31. T. M. Rizki, R. M. Rizki, Parasitoid-induced cellular immune deficiency in Drosophila. *Ann N Y Acad Sci* **712**, 178-194 (1994).
32. B. Z. Kacsoh, T. A. Schlenke, High hemocyte load is associated with increased resistance against parasitoids in Drosophila suzukii, a relative of D. melanogaster. *PloS one* **7**, e34721 (2012).
33. Neil F. Milan, Balint Z. Kacsoh, Todd A. Schlenke, Alcohol Consumption as Self-Medication against Blood-Borne Parasites in the Fruit Fly. *Current Biology* **22**, 488-493 (2012).
34. T. Lefèvre, J. C. de Roode, B. Z. Kacsoh, T. A. Schlenke, Defence strategies against a parasitoid wasp in Drosophila: fight or flight? *Biology Letters* **8**, 230-233 (2012).
35. R. Y. Hwang *et al.*, Nociceptive Neurons Protect Drosophila Larvae from Parasitoid Wasps. *Current Biology* **17**, 2105-2116 (2007).
36. C.-H. Lue, A. C. Driskell, J. Leips, M. L. Buffington, Review of the genus Leptopilina (Hymenoptera, Cynipoidea, Figitidae, Eucoilinae) from the Eastern United States, including three newly described species. *Journal of Hymenoptera Research* **53**, 35 (2016).
37. M. Schilthuizen, G. Nordlander, R. Stouthamer, J. Van Alphen, Morphological and molecular phylogenetics in the genus Leptopilina (Hymenoptera: Cynipoidea: Eucoilidae). (1998).
38. Y. Carton, M. Bouletreau, J. Alphen, J. van Lenteren. (1986), vol. 3e, pp. 347-394.
39. A. H. Kopelman, P. C. Chabora, Immature Stages of Leptopilina boulardi (Hymenoptera: Eucoilidae), a Protelean Parasite of Drosophila spp. (Diptera: Drosophilidae). *Annals of the Entomological Society of America* **77**, 264-269 (1984).
40. C. Small, I. Paddibhatla, R. Rajwani, S. Govind, An introduction to parasitic wasps of Drosophila and the antiparasite immune response. *J Vis Exp*, e3347 (2012).
41. J. Varaldi, S. Petit, M. Bouletreau, F. Fleury, The virus infecting the parasitoid Leptopilina boulardi exerts a specific action on superparasitism behaviour. *Parasitology* **132**, 747-756 (2006).
42. J. Martinez, F. Fleury, J. Varaldi, Competitive outcome of multiple infections in a behavior-manipulating virus/wasp interaction. *Ecol Evol* **5**, 5934-5945 (2015).
43. G. Salt, J. Gray, Experimental studies in insect parasitism. II.―Superparasitism. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* **114**, 455-476 (1934).
44. J. J. Van Alphen, M. E. Visser, Superparasitism as an adaptive strategy for insect parasitoids. *Annual review of entomology* **35**, 59-79 (1990).
45. M. Silvers, A. Nappi, In vitro study of physiological suppression of supernumerary parasites by the endoparasitic wasp Leptopilina heterotoma. *The Journal of parasitology*, 405-409 (1986).
46. J. A. Hoffmann, The immune response of Drosophila. *Nature* **426**, 33-38 (2003).
47. S. Govind, Innate immunity in Drosophila: Pathogens and pathways. *Insect Sci* **15**, 29-43 (2008).
48. N. Buchon, N. Silverman, S. Cherry, Immunity in Drosophila melanogaster--from microbial recognition to whole-organism physiology. *Nat Rev Immunol* **14**, 796-810 (2014).
49. M. Crozatier, M. Meister, Drosophila haematopoiesis. *Cell Microbiol* **9**, 1117-1126 (2007).
50. S. H. Jung, C. J. Evans, C. Uemura, U. Banerjee, The Drosophila lymph gland as a developmental model of hematopoiesis. *Development* **132**, 2521-2533 (2005).
51. U. Banerjee, J. R. Girard, L. M. Goins, C. M. Spratford, Drosophila as a Genetic Model for Hematopoiesis. *Genetics* **211**, 367-417 (2019).

52. G. Gueguen, M. E. Kalamarz, J. Ramroop, J. Uribe, S. Govind, Polydnaviral ankyrin proteins aid parasitic wasp survival by coordinate and selective inhibition of hematopoietic and immune NF-kappa B signaling in insect hosts. *PLoS Pathog* **9**, e1003580 (2013).
53. L. L. Pech, M. R. Strand, Granular cells are required for encapsulation of foreign targets by insect haemocytes. *J Cell Sci* **109 (Pt 8)**, 2053-2060 (1996).
54. J. Russo, S. Dupas, F. Frey, Y. Carton, M. Brehelin, Insect immunity: early events in the encapsulation process of parasitoid (Leptopilina boulardi) eggs in resistant and susceptible strains of Drosophila. *Parasitology* **112 (Pt 1)**, 135-142 (1996).
55. B. Lemaitre, J. Hoffmann, The host defense of Drosophila melanogaster. *Annu Rev Immunol* **25**, 697-743 (2007).
56. P. Vilmos, É. Kurucz, Insect immunity: evolutionary roots of the mammalian innate immune system. *Immunology Letters* **62**, 59-66 (1998).
57. J. F. Hillyer, Insect immunology and hematopoiesis. *Dev Comp Immunol* **58**, 102-118 (2016).
58. A. Holz, B. Bossinger, T. Strasser, W. Janning, R. Klapper, The two origins of hemocytes in Drosophila. *Development* **130**, 4955-4962 (2003).
59. R. Lanot, D. Zachary, F. Holder, M. Meister, Postembryonic hematopoiesis in Drosophila. *Developmental biology* **230**, 243-257 (2001).
60. R. Márkus *et al.*, Sessile hemocytes as a hematopoietic compartment in <em>Drosophila melanogaster</em&gt. *Proceedings of the National Academy of Sciences* **106**, 4805 (2009).
61. K. Makhijani, B. Alexander, T. Tanaka, E. Rulifson, K. Brückner, The peripheral nervous system supports blood cell homing and survival in the Drosophila larva. *Development* **138**, 5379-5391 (2011).
62. I. Louradour *et al.*, Reactive oxygen species-dependent Toll/NF-κB activation in the Drosophila hematopoietic niche confers resistance to wasp parasitism. *eLife* **6**, e25496 (2017).
63. T. M. Rizki, R. M. Rizki, Lamellocyte differentiation in Drosophila larvae parasitized by Leptopilina. *Dev Comp Immunol* **16**, 103-110 (1992).
64. C. Labrosse, P. Eslin, G. Doury, J. M. Drezen, M. Poirié, Haemocyte changes in D. Melanogaster in response to long gland components of the parasitoid wasp Leptopilina boulardi: a Rho-GAP protein as an important factor. *J Insect Physiol* **51**, 161-170 (2005).
65. R. P. Sorrentino, Y. Carton, S. Govind, Cellular immune response to parasite infection in the Drosophila lymph gland is developmentally regulated. *Developmental biology* **243**, 65-80 (2002).
66. M. Meister, Blood cells of Drosophila: cell lineages and role in host defence. *Curr Opin Immunol* **16**, 10-15 (2004).
67. I. Anderl *et al.*, Transdifferentiation and Proliferation in Two Distinct Hemocyte Lineages in Drosophila melanogaster Larvae after Wasp Infection. *PLoS Pathog* **12**, e1005746 (2016).
68. O. Schmidt, U. Theopold, M. Strand, Innate immunity and its evasion and suppression by hymenopteran endoparasitoids. *Bioessays* **23**, 344-351 (2001).
69. S. Cherry, N. Silverman, Host-pathogen interactions in drosophila: new tricks from an old friend. *Nat Immunol* **7**, 911-917 (2006).
70. E. S. Keebaugh, T. A. Schlenke, Insights from natural host-parasite interactions: The Drosophila model. *Developmental and Comparative Immunology* **42**, 111-123 (2014).
71. A. Nappi, Parasite encapsulation in insects. *Invertebrate immunity*, 293-326 (1975).
72. I. M. Dubovskiy, N. A. Kryukova, V. V. Glupov, N. A. Ratcliffe, Encapsulation and nodulation in insects. *Invertebrate Survival Journal*, 229-246%V 213 (2016).
73. P. Peyron *et al.*, Foamy macrophages from tuberculous patients' granulomas constitute a nutrient-rich reservoir for M. tuberculosis persistence. *PLoS Pathog* **4**, e1000204 (2008).
74. D. Blumberg, Parasitoid Encapsulation as a Defense Mechanism in the Coccoidea (Homoptera) and Its Importance in Biological Control. *Biological Control* **8**, 225-236 (1997).
75. M. Alleyne, R. N. Wiedenmann, Encapsulation and hemocyte numbers in three lepidopteran stemborers parasitized by Cotesia flavipes-complex endoparasitoids. *Entomologia experimentalis et applicata* **100**, 279-293 (2001).
76. J. Hu, X.-X. Zhu, W.-J. Fu, Passive evasion of encapsulation in Macrocentrus cingulum Brischke (Hymenoptera: Braconidae), a polyembryonic parasitoid of Ostrinia furnacalis Guenée (Lepidoptera: Pyralidae). *Journal of Insect Physiology* **49**, 367-375 (2003).

77. A. Iacovone, N. Ris, M. Poirié, J. L. Gatti, Time-course analysis of Drosophila suzukii interaction with endoparasitoid wasps evidences a delayed encapsulation response compared to D. melanogaster. *PloS one* **13**, e0201573 (2018).
78. M. Alleyne, R. N. Wiedenmann, Suitability of lepidopteran stemborers for parasitization by novel-association endoparasitoids. *BioControl* **46**, 1-23 (2001).
79. M. B. Mochiah, A. J. Ngi‐Song, W. A. Overholt, R. Stouthamer, Variation in encapsulation sensitivity of Cotesia sesamiae biotypes to Busseola fusca. *Entomologia experimentalis et applicata* **105**, 111-118 (2002).
80. X. Wang, A. Biondi, K. M. Daane, Functional Responses of Three Candidate Asian Larval Parasitoids Evaluated for Classical Biological Control of Drosophila suzukii (Diptera: Drosophilidae). *Journal of Economic Entomology* **113**, 73-80 (2020).
81. M. A. Hanson, B. Lemaitre, New insights on Drosophila antimicrobial peptide function in host defense and beyond. *Curr Opin Immunol* **62**, 22-30 (2020).
82. D. S. Schneider, K. L. Hudson, T. Y. Lin, K. V. Anderson, Dominant and recessive mutations define functional domains of Toll, a transmembrane protein required for dorsal-ventral polarity in the Drosophila embryo. *Genes Dev* **5**, 797-807 (1991).
83. S. Tauszig, E. Jouanguy, J. A. Hoffmann, J. L. Imler, Toll-related receptors and the control of antimicrobial peptide expression in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 10520-10525 (2000).
84. S. Valanne, J. H. Wang, M. Ramet, The Drosophila Toll signaling pathway. *J Immunol* **186**, 649-656 (2011).
85. M. Nakamoto *et al.*, Virus recognition by Toll-7 activates antiviral autophagy in Drosophila. *Immunity* **36**, 658-667 (2012).
86. I. Akhouayri, C. Turc, J. Royet, B. Charroux, Toll-8/Tollo negatively regulates antimicrobial response in the Drosophila respiratory epithelium. *PLoS Pathog* **7**, e1002319 (2011).
87. I. Paddibhatla, M. J. Lee, M. E. Kalamarz, R. Ferrarese, S. Govind, Role for sumoylation in systemic inflammation and immune homeostasis in Drosophila larvae. *PLoS Pathog* **6**, e1001234 (2010).
88. A. K. Shia *et al.*, Toll-dependent antimicrobial responses in Drosophila larval fat body require Spatzle secreted by haemocytes. *J Cell Sci* **122**, 4505-4515 (2009).
89. R. M. Rizki, T. M. Rizki, Selective destruction of a host blood cell type by a parasitoid wasp. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 6154-6158 (1984).
90. J. Morales *et al.*, Biogenesis, structure, and immune-suppressive effects of virus-like particles of a Drosophila parasitoid, Leptopilina victoriae. *J Insect Physiol* **51**, 181-195 (2005).
91. H. Chiu, J. Morales, S. Govind, Identification and immuno-electron microscopy localization of p40, a protein component of immunosuppressive virus-like particles from Leptopilina heterotoma, a virulent parasitoid wasp of Drosophila. *J Gen Virol* **87**, 461-470 (2006).
92. R. Ferrarese, J. Morales, D. Fimiarz, B. A. Webb, S. Govind, A supracellular system of actin-lined canals controls biogenesis and release of virulence factors in parasitoid venom glands. *J Exp Biol* **212**, 2261-2268 (2009).
93. S. Dupas, M. Brehelin, F. Frey, Y. Carton, Immune suppressive virus-like particles in a Drosophila parasitoid: significance of their intraspecific morphological variations. *Parasitology* **113 (Pt 3)**, 207-212 (1996).
94. C. Labrosse, Y. Carton, A. Dubuffet, J. M. Drezen, M. Poirie, Active suppression of D. melanogaster immune response by long gland products of the parasitic wasp Leptopilina boulardi. *J Insect Physiol* **49**, 513-522 (2003).
95. B. Wan *et al.*, Venom Atypical Extracellular Vesicles as Interspecies Vehicles of Virulence Factors Involved in Host Specificity: The Case of a Drosophila Parasitoid Wasp. *Frontiers in Immunology* **10**, (2019).
96. M. E. Heavner *et al.*, Novel Organelles with Elements of Bacterial and Eukaryotic Secretion Systems Weaponize Parasites of Drosophila. *Current biology : CB* **27**, 2869-2877.e2866 (2017).
97. B. Wey *et al.*, Immune Suppressive Extracellular Vesicle Proteins of Leptopilina heterotoma Are Encoded in the Wasp Genome. *G3 (Bethesda)* **10**, 1-12 (2020).
98. D. Di Giovanni *et al.*, A behavior-manipulating virus relative as a source of adaptive genes for Drosophila parasitoids. *Mol Biol Evol*, (2020).

99. G. Gueguen, R. Rajwani, I. Paddibhatla, J. Morales, S. Govind, VLPs of Leptopilina boulardi share biogenesis and overall stellate morphology with VLPs of the heterotoma clade. *Virus Res* **160**, 159-165 (2011).
100. H. Chiu, S. Govind, Natural infection of D. melanogaster by virulent parasitic wasps induces apoptotic depletion of hematopoietic precursors. *Cell death and differentiation* **9**, 1379-1381 (2002).
101. M. R. Strand, G. R. Burke, Polydnaviruses: From discovery to current insights. *Virology* **479-480**, 393-402 (2015).
102. B. A. Webb *et al.*, Polydnavirus genomes reflect their dual roles as mutualists and pathogens. *Virology* **347**, 160-174 (2006).
103. Y. Bigot, S. Samain, C. Auge-Gouillou, B. A. Federici, Molecular evidence for the evolution of ichnoviruses from ascoviruses by symbiogenesis. *BMC Evol Biol* **8**, 253 (2008).
104. A. Bézier, J. Herbinière, B. Lanzrein, J. M. Drezen, Polydnavirus hidden face: the genes producing virus particles of parasitic wasps. *J Invertebr Pathol* **101**, 194-203 (2009).
105. A. Bézier *et al.*, Polydnaviruses of braconid wasps derive from an ancestral nudivirus. *Science* **323**, 926-930 (2009).
106. Y. Bigot *et al.*, Symbiotic Virus at the Evolutionary Intersection of Three Types of Large DNA Viruses; Iridoviruses, Ascoviruses, and Ichnoviruses. *PloS one* **4**, (2009).
107. B. Misof *et al.*, Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763-767 (2014).
108. A. N. Volkoff *et al.*, Analysis of virion structural components reveals vestiges of the ancestral ichnovirus genome. *PLoS Pathog* **6**, e1000923 (2010).
109. A. Pichon *et al.*, Recurrent DNA virus domestication leading to different parasite virulence strategies. *Sci Adv* **1**, e1501150 (2015).
110. G. R. Burke, T. J. Simmonds, B. J. Sharanowski, S. M. Geib, Rapid Viral Symbiogenesis via Changes in Parasitoid Wasp Genome Architecture. *Mol Biol Evol* **35**, 2463-2474 (2018).
111. R. I. B. Francki, C. Fauquet, D. Knudson, F. Brown, *Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Viruses. Virology Division of the International Union of Microbiological Societies*. (Springer Science & Business Media, 2012), vol. 2.
112. M. R. Strand, G. R. Burke, Polydnaviruses: Evolution and Function. *Current issues in molecular biology* **34**, 163-182 (2019).
113. Y. Wang, J. A. Jehle, Nudiviruses and other large, double-stranded circular DNA viruses of invertebrates: New insights on an old topic. *Journal of Invertebrate Pathology* **101**, 187-193 (2009).
114. S. Asgari *et al.*, ICTV Virus Taxonomy Profile: Ascoviridae. *The Journal of general virology* **98**, 4-5 (2017).
115. R. L. Harrison *et al.*, ICTV Virus Taxonomy Profile: Nudiviridae. *Journal of General Virology* **101**, 3-4 (2020).
116. Y. Zhang, J. Wang, G.-Z. Han, Chalcid wasp paleoviruses bridge the evolutionary gap between bracoviruses and nudiviruses. *Virology* **542**, 34-39 (2020).
117. A. Lorenzi *et al.*, RNA interference identifies domesticated viral genes involved in assembly and trafficking of virus-derived particles in ichneumonid wasps. *PLOS Pathogens* **15**, e1008210 (2019).
118. D. B. Stoltz, S. B. Vinson, in *Advances in Virus Research,* M. A. Lauffer, F. B. Bang, K. Maramorosch, K. M. Smith, Eds. (Academic Press, 1979), vol. 24, pp. 125-171.
119. L. Gasmi *et al.*, Recurrent Domestication by Lepidoptera of Genes from Their Parasites Mediated by Bracoviruses. *Plos Genet* **11**, (2015).
120. M. R. Strand, G. R. Burke, Polydnaviruses as symbionts and gene delivery systems. *PLoS Pathog* **8**, e1002757 (2012).
121. G. R. Burke, H. M. Hines, B. J. Sharanowski, Endogenization from diverse viral ancestors is common and widespread in parasitoid wasps. *bioRxiv*, 2020.2006.2017.148684 (2020).
122. J. Martinez, D. Lepetit, M. Ravallec, F. Fleury, J. Varaldi, Additional heritable virus in the parasitic wasp Leptopilina boulardi: prevalence, transmission and phenotypic effects. *J Gen Virol* **97**, 523-535 (2016).

123. J. Varaldi, D. Lepetit, Deciphering the behaviour manipulation imposed by a virus on its parasitoid host: insights from a dual transcriptomic approach. *Parasitology* **145**, 1979-1989 (2018).
124. D. Lepetit, B. Gillet, S. Hughes, K. Kraaijeveld, J. Varaldi, Genome Sequencing of the Behavior Manipulating Virus LbFV Reveals a Possible New Virus Family. *Genome Biol Evol* **8**, 3718-3739 (2016).
125. F. Dong, J. Wang, R. Deng, X. Wang, Autographa californica multiple nucleopolyhedrovirus gene ac81 is required for nucleocapsid envelopment. *Virus Research* **221**, 47-57 (2016).
126. J. E. Galán, M. Lara-Tejero, T. C. Marlovits, S. Wagner, Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells. *Annu Rev Microbiol* **68**, 415-438 (2014).
127. K. Nothelfer *et al.*, B lymphocytes undergo TLR2-dependent apoptosis upon Shigella infection. *J Exp Med* **211**, 1215-1229 (2014).
128. O. Arizmendi, W. D. Picking, W. L. Picking, Macrophage Apoptosis Triggered by IpaD from Shigella flexneri. *Infect Immun* **84**, 1857-1865 (2016).
129. T. L. Hale, P. J. Sansonetti, P. A. Schad, S. Austin, S. B. Formal, Characterization of virulence plasmids and plasmid-associated outer membrane proteins in Shigella flexneri, Shigella sonnei, and Escherichia coli. *Infect Immun* **40**, 340-350 (1983).
130. T. L. Hale, Genetic basis of virulence in Shigella species. *Microbiological reviews* **55**, 206-224 (1991).
131. K. Kaniga, D. Trollinger, J. E. Galan, Identification of two targets of the type III protein secretion system encoded by the inv and spa loci of Salmonella typhimurium that have homology to the Shigella IpaD and IpaA proteins. *Journal of bacteriology* **177**, 7078-7085 (1995).
132. C. J. Hueck *et al.*, Salmonella typhimurium secreted invasion determinants are homologous to Shigella Ipa proteins. *Molecular microbiology* **18**, 479-490 (1995).
133. M. P. Stevens *et al.*, An Inv/Mxi-Spa-like type III protein secretion system in Burkholderia pseudomallei modulates intracellular behaviour of the pathogen. *Molecular microbiology* **46**, 649-659 (2002).
134. H. G. S. Consortium, Insights into social insects from the genome of the honeybee Apis mellifera. *Nature* **443**, 931-949 (2006).
135. Y. Wurm *et al.*, The genome of the fire ant Solenopsis invicta. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 5679-5684 (2011).
136. C. G. Elsik *et al.*, Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* **15**, 86 (2014).
137. G. R. Burke, T. J. Simmonds, S. A. Thomas, M. R. Strand, Microplitis demolitor Bracovirus Proviral Loci and Clustered Replication Genes Exhibit Distinct DNA Amplification Patterns during Replication. *J Virol* **89**, 9511-9523 (2015).
138. K. Kraaijeveld *et al.*, Decay of Sexual Trait Genes in an Asexual Parasitoid Wasp. *Genome Biology and Evolution* **8**, 3685-3695 (2016).
139. V. E. Gokhman *et al.*, Genomic and karyotypic variation in Drosophila parasitoids (Hymenoptera, Cynipoidea, Figitidae). *Comp Cytogenet* **5**, 211-221 (2011).
140. S. Khan, D. T. Sowpati, A. Srinivasan, M. Soujanya, R. K. Mishra, Long-Read Genome Sequencing and Assembly of. *G3 (Bethesda)* **10**, 1485-1494 (2020).
141. T. Baba *et al.*, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006.0008-2006.0008 (2006).
142. J. B. Duffy, GAL4 system in Drosophila: a fly geneticist's Swiss army knife. *Genesis* **34**, 1-15 (2002).
143. A. H. Brand, A. S. Manoukian, N. Perrimon, Ectopic expression in Drosophila. *Methods Cell Biol* **44**, 635-654 (1994).
144. T.-b. Chou, N. Perrimon, The Autosomal FLP-DFS Technique for Generating Germline Mosaics in Drosophila melanogaster. *Genetics* **144**, 1673-1679 (1996).
145. W. A. Zehring *et al.*, P-element transformation with period locus DNA restores rhythmicity to mutant, arrhythmic Drosophila melanogaster. *Cell* **39**, 369-376 (1984).
146. A. C. Spradling *et al.*, Gene disruptions using P transposable elements: an integral component of the Drosophila genome project. *Proceedings of the National Academy of Sciences* **92**, 10824-10830 (1995).

147. Y. Tomoyasu *et al.*, Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in Tribolium. *Genome Biol* **9**, R10 (2008).
148. J. H. Werren, D. W. Loehlin, J. D. Giebel, Larval RNAi in Nasonia (parasitoid wasp). *Cold Spring Harb Protoc* **2009**, pdb.prot5311 (2009).
149. W. Hunter *et al.*, Large-Scale Field Application of RNAi Technology Reducing Israeli Acute Paralysis Virus Disease in Honey Bees (Apis mellifera, Hymenoptera: Apidae). *PLOS Pathogens* **6**, e1001160 (2010).
150. C.-Y. Hsu, H.-F. Lo, N. S. Mutti, G. V. Amdam, Ferritin RNA interference inhibits the formation of iron granules in the trophocytes of worker honey bees (Apis mellifera). *Scientific Reports* **9**, 10098 (2019).
151. D. Colinet *et al.*, Development of RNAi in a Drosophila endoparasitoid wasp and demonstration of its efficiency in impairing venom protein production. *J Insect Physiol* **63**, 56-61 (2014).
152. D. R. G. Price, J. A. Gatehouse, RNAi-mediated crop protection against insects. *Trends in Biotechnology* **26**, 393-400 (2008).
153. R. Asokan, G. S. Chandra, M. Manamohan, N. K. Kumar, Effect of diet delivered various concentrations of double-stranded RNA in silencing a midgut and a non-midgut gene of Helicoverpa armigera. *Bull Entomol Res* **103**, 555-563 (2013).
154. R. Asokan, G. S. Chandra, M. Manamohan, N. K. K. Kumar, T. Sita, Response of various target genes to diet-delivered dsRNA mediated RNA interference in the cotton bollworm, Helicoverpa armigera. *J Pest Sci* **87**, 163-172 (2014).
155. K. B. Rebijith *et al.*, Diet-Delivered dsRNAs for Juvenile Hormone-Binding Protein and Vacuolar ATPase-H Implied Their Potential in the Management of the Melon Aphid (Hemiptera: Aphididae). *Environ Entomol*, (2015).
156. H. Tabara, A. Grishok, C. C. Mello, RNAi in C. elegans: Soaking in the Genome Sequence. *Science* **282**, 430-431 (1998).
157. R. S. Kamath, J. Ahringer, Genome-wide RNAi screening in Caenorhabditis elegans. *Methods* **30**, 313-321 (2003).
158. R. S. Kamath *et al.*, Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* **421**, 231-237 (2003).
159. I. E. M. EIJS, J. ELLERS, G.-J. VAN DUINEN, Feeding strategies in drosophilid parasitoids: the impact of natural food resources on energy reserves in females. *Ecological Entomology* **23**, 133-138 (1998).
160. M. Li, M. Bui, O. S. Akbari, Embryo Microinjection and Transplantation Technique for Nasonia vitripennis Genome Manipulation. *J Vis Exp*, (2017).
161. M. Li *et al.*, Generation of heritable germline mutations in the jewel wasp Nasonia vitripennis using CRISPR/Cas9. *Scientific Reports* **7**, 901 (2017).
162. D. Chaverra-Rodriguez *et al.*, Germline mutagenesis of Nasonia vitripennis through ovarian delivery of CRISPR-Cas9 ribonucleoprotein. *bioRxiv*, 2020.2005.2010.087494 (2020).
163. A. R. Whiting, The Biology of the Parasitic Wasp Mormoniella vitripennis [=Nasonia brevicornis] (Walker). *The Quarterly Review of Biology* **42**, 333-406 (1967).
164. J. E. Losey, M. Vaughan, The Economic Value of Ecological Services Provided by Insects. *BioScience* **56**, 311-323 (2006).
165. S. Macfadyen, A. P. Davies, M. P. Zalucki, Assessing the impact of arthropod natural enemies on crop pests at the field scale. *Insect Science* **22**, 20-34 (2015).
166. M. H. Schmidt *et al.*, Relative importance of predators and parasitoids for cereal aphid control. *Proc Biol Sci* **270**, 1905-1909 (2003).
167. A. Dereeper, S. Audic, J. M. Claverie, G. Blanc, BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol Biol* **10**, 8 (2010).
168. A. Dereeper *et al.*, Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**, W465-469 (2008).
169. C. Notredame, D. G. Higgins, J. Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology* **302**, 205-217 (2000).
170. J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540-552 (2000).
171. S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704 (2003).

172. M. Anisimova, O. Gascuel, Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* **55**, 539-552 (2006).
173. F. Chevenet, C. Brun, A. L. Banuls, B. Jacq, R. Christen, TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**, 439 (2006).
174. P. J. Sansonetti, D. J. Kopecko, S. B. Formal, Involvement of a plasmid in the invasive ability of Shigella flexneri. *Infect Immun* **35**, 852-860 (1982).
175. O. Attree, I. Attree, A second type III secretion system in Burkholderia pseudomallei: who is the real culprit? *Microbiology* **147**, 3197-3199 (2001).
176. W. L. Picking *et al.*, IpaD of Shigella flexneri is independently required for regulation of Ipa protein secretion and efficient insertion of IpaB and IpaC into host membranes. *Infect Immun* **73**, 1432-1440 (2005).
177. M. P. Stevens *et al.*, Attenuated virulence and protective efficacy of a Burkholderia pseudomallei bsa type III secretion mutant in murine models of melioidosis. *Microbiology* **150**, 2669-2676 (2004).
178. S. Johnson *et al.*, Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD. *The Journal of biological chemistry* **282**, 4035-4044 (2007).
179. A. D. Roehrich, E. Guillossou, A. J. Blocker, I. Martinez-Argudo, Shigella IpaD has a dual role: signal transduction from the type III secretion system needle tip and intracellular secretion regulation. *Molecular microbiology* **87**, 690-706 (2013).
180. T. Rathinavelan, C. Tang, R. N. De Guzman, Characterization of the interaction between the Salmonella type III secretion system tip protein SipD and the needle protein PrgI by paramagnetic relaxation enhancement. *The Journal of biological chemistry* **286**, 4922-4930 (2011).
181. J. E. Galan, Interactions of Salmonella with host cells: encounters of the closest kind. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14006-14008 (1998).
182. T. Miki, N. Okada, Y. Shimada, H. Danbara, Characterization of Salmonella pathogenicity island 1 type III secretion-dependent hemolytic activity in Salmonella enterica serovar Typhimurium. *Microbial pathogenesis* **37**, 65-72 (2004).
183. M. Lara-Tejero, J. E. Galan, Salmonella enterica serovar typhimurium pathogenicity island 1-encoded type III secretion system translocases mediate intimate attachment to nonphagocytic cells. *Infect Immun* **77**, 2635-2642 (2009).
184. S. K. Myeni, L. Wang, D. Zhou, SipB-SipC complex is essential for translocon formation. *PloS one* **8**, e60499 (2013).
185. D. Hersh *et al.*, The Salmonella invasin SipB induces macrophage apoptosis by binding to caspase-1. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 2396-2401 (1999).
186. L. D. Hernandez, M. Pypaert, R. A. Flavell, J. E. Galan, A Salmonella protein causes macrophage cell death by inducing autophagy. *J Cell Biol* **163**, 1123-1131 (2003).
187. S. Suparak *et al.*, Multinucleated giant cell formation and apoptosis in infected host cells is mediated by Burkholderia pseudomallei type III secretion protein BipB. *Journal of bacteriology* **187**, 6556-6560 (2005).
188. E. J. McGhie, R. D. Hayward, V. Koronakis, Cooperation between actin-binding proteins of invasive Salmonella: SipA potentiates SipC nucleation and bundling of actin. *EMBO J* **20**, 2131-2139 (2001).

CHAPTER 1

Immune suppressive extracellular vesicle proteins of *Leptopilina heterotoma* are encoded in the wasp genome

ABSTRACT

Leptopilina heterotoma are obligate parasitoid wasps that develop in the body of their *Drosophila* hosts. During oviposition, female wasps introduce venom into the larval hosts' body cavity. The venom contains discrete, 300 nm-wide, mixed-strategy extracellular vesicles (MSEVs), until recently referred to as virus-like particles. While the crucial immune suppressive functions of *L. heterotoma* MSEVs have remained undisputed, their biotic nature and origin still remain controversial. In recent proteomics analyses of *L. heterotoma* MSEVs, we identified 161 proteins in three classes: conserved eukaryotic proteins, infection and immunity related proteins, and proteins without clear annotation. Here we report 246 additional proteins from the *L. heterotoma* MSEV proteome. An enrichment analysis of the entire proteome supports vesicular nature of these structures. Sequences for more than 90% of these proteins are present in the whole-body transcriptome. Sequencing and *de novo* assembly of the 460 Mb-sized *L. heterotoma* genome revealed 90% of MSEV proteins have coding regions within the genomic scaffolds. Altogether, these results explain the stable association of MSEVs with their wasps, and like other wasp structures, their vertical inheritance. While our results do not rule out a viral origin of MSEVs, they suggest that a similar strategy for co-opting cellular machinery for immune suppression may be shared by other wasps to gain advantage over their hosts. These results are relevant to our understanding of the evolution of figitid and related wasp species.

INTRODUCTION

Parasitic wasps are amongst the most abundant insects; they are vital to biodiversity and contribute to biological control of agricultural pests (*1, 2*). A common strategy for reproductive success of parasitic wasps is suppression of immunity in their larval hosts. Parasitic wasps produce viruses or virus-like particles in tissues associated with the ovary. Wasps of the Ichneumonoidea superfamily produce symbiotic polydnaviruses (PDVs), which package circular dsDNA. PDV (Bracovirus (BV) in braconid wasps; Ichnovirus (IV) in ichneumonid wasps) genomes are integrated within the wasp genome as islands of viral genes. Upon oviposition, PDVs suppress host immunity. BVs and IVs derive from nudivirus and large DNA cytoplasmic viruses, respectively (reviewed in (*3-5*) and references therein).

Immune-suppressive virus-like particles (VLPs) (e.g., VcVLPs in the ichneumonid *Venturia canescens* and FaENVs in the braconid *Fopius arisanus*) lack proviral DNA segments, but are of viral origin and transfer virulence proteins into host cells (*6, 7*). Viral genes encoding VLP proteins are either dispersed in the wasp genome (as in VcVLP) or present in discrete genomic areas (as in FaENV). Thus, various independent viral endogenization events have been important for successful parasitism by these wasps (*3, 5*).

Here, we focus on immune-suppressive particles of figitid wasps of the genus *Leptopilina,* that infect *Drosophila* spp. and are gaining importance as models for natural host-parasite interactions (*8*). *Leptopilina heterotoma* (*Lh*), *L. victoriae* (*Lv*), and *L. boulardi* (*Lb*) produce VLPs in their venom glands. The VLPs of *Leptopilina* spp. and their proteins have been linked to parasite success (*9-14*). Evidence for DNA in *Leptopilina* VLPs is lacking, and because of the absence of a published wasp genome, the chromosomal versus extrachromosomal location of MSEV protein genes is not known. Our goals here are (a) to describe additional proteins in the MSEV proteome and examine their relationship with PDV and other viral proteins, and (b) determine whether MSEV genes are encoded within the wasp genome.

We recently described 161 proteins in the VLPs from two *Lh* strains in three classes: conserved eukaryotic with cellular function (Class 1), infection- and immunity-related (Class 2), and unannotated (novel) proteins without similarity to known proteins (Class 3) (*15*). Class 1 proteins include several vesicular transport and endomembrane system proteins. Class 2 proteins include predicted modulators of immune response, e.g., metalloendopeptidases, RhoGAPs, a knottin-like protein, and a new family of

prokaryotic-like GTPases whose genes lack introns. A striking example of Class 3 proteins is p40, with three-dimensional structural similarity to Type 3 secretion system (T3SS) needle tip proteins, IpaD/SipD/BipD from Gram-negative bacteria, *Salmonella, Shigella* and *Burkholderia*. Earlier results have indicated that the *p40* gene (unlike the *GTPase* genes) is expected to have introns. These results suggested that *Lh* VLPs have novel properties with elements of the prokaryotic and eukaryotic secretion systems and possess a functionally diverse array of immune-suppressive proteins. We therefore renamed VLPs as Mixed Strategy Extracellular Vesicles (MSEVs). Their variable morphologies distinguish them from ordered PDV morphologies. Additionally, genes encoding abundant MSEV proteins p40 and GTPase are present even in antibiotic-treated wasps. These results favored a non-microbial nature of MSEVs (*15*).

Here, we present an analysis of an additional 246 proteins from the *Lh* 14 MSEV proteome to obtain a more comprehensive description. A combined analysis of these and previous results reinforce the idea that the MSEV proteome is enriched in exosomal proteins and that Class 3 proteins are not shared with either *Lb* or an unrelated *Ganaspis* spp. Whole-body transcriptome of adult *Lh* wasps validated the expression of the MSEV genes. *De novo* genomic assembly and analyses revealed 90% of conserved Insecta Benchmarking Universal Single-Copy Orthologs (BUSCOs), as well as a majority (375/407; ~90%) of the MSEV proteins are encoded in the wasp genome. While we cannot rule out a viral origin of MSEVs, in aggregate, our results provide a clearer understanding of the extant nature of these complex structures and strengthen the idea that specialized extracellular vesicles transfer wasp virulence factors and other parasite proteins into *Drosophila* host cells.

MATERIALS AND METHODS

Insects: Isogenized *Lh* strains New York (NY; (*13, 16*)) and *Lh* 14 (*17*), were raised on the *y w* strain of *D. melanogaster* that were reared on standard cornmeal, yeast, and agar fly food at 25°C as described (*18*). Adult wasps were collected from parasitized hosts, 25 days after infection at 25ºC. Male and female wasps were stored on fly food with 70% honey on "buzz" plugs.

Analysis of MSEV super-set ORFs: Previously undescribed open reading frames (ORFs) from the *Lh* 14 MSEV proteome and sequenced as part of (*15*) (PXD005632) are analyzed in the context of the published female abdominal (*19*) and whole body (this study) *Lh* 14 transcriptomes. We have not observed any difference in venom activities of *Lh* 14 and *Lh* NY (*11, 17*), or in wasp success under laboratory conditions. The *Lh* 14 ORFs were aligned against transcripts from BioProject: PRJNA202370, Accession number GAJC0000000 (*19*) as previously described in (*15*). Proteins with an ORF and a transcript were run through the BLAST2GO (v 5.2; downloaded June 2018) annotation pipeline with an E-value threshold of $1x10^{-7}$ (*20, 21*). Results were organized and classified based on Gene Ontology (GO) terms from UniProt and InterPro (*22-25*). Proteins were considered "virulence-related" based on GO terms indicating involvement with infection, host evasion, inflammation, and immune response. ORFs that did not return results via BLAST or InterProScan (Class 3 proteins) were run through Conserved Domain Search (CDD) on NCBI (version 3.16) (*26*). The E-value cut off for CDD search was $1x10^{-2}$. Proteins were considered to have a signal peptide if one was predicted using Phobeus and Signal P (*27-30*). Transmembrane domains were considered to be present if they were predicted using Phobeus and TMHMM (*27, 28, 31, 32*).

The GhostKOALA algorithm (*33*) was used to assign KEGG ortholog (KO) numbers for the MSEV superset protein sequences. If a primary KO number failed to be assigned by GhostKOALA, a secondary number assignment with a score >= 50 was used. Redundant KO numbers were excluded.

MSEV proteins were included in the enrichment analyses only if a human ortholog exists; the gene identifiers for human orthologs were obtained from the MSEV KO and the UniProt mapping utility (*24, 34*). (Human orthologs were chosen because a robust proportion of Vesiclepedia's data is derived

from human vesicle proteomes.) The orthologs of human genes were analyzed for enrichment with the FunRich algorithm (*35, 36*) against the Vesiclepedia database (*37, 38*).

Finally, the MSEV proteome was used as a query using BLASTp for the following databases: "non-redundant" (nr), nr restricted to Taxid: Viridae (10239), nr restricted to Taxid: Polydnaviridae (10482), and nr restricted to Taxid: Unclassified Polydnaviridae (40273) (E-value threshold: $1.0x10^{-3}$, %ID minimum: 20%, performed 04/16/2019). tBLASTn of *L. boulardi* and *G. hookeri* (previously called *Ganaspis spp. 1*) (*19*) transcriptomes was performed on 03/10/2019; the threshold for homologs in *Lb* and *G. hookeri* were 25% ID and an E-value of $1.0x10^{-10}$.

Genomes sequencing and assembly: Library preparations, sequencing reactions, and associated validations were conducted by GENEWIZ, Inc. (South Plainfield, NJ, USA). Genomic DNA was extracted from ~50 mg of tissue (~100 wasps) of *Lh* males and females separately using mixed bead beating and PureLink Genomic DNA extraction kits following manufacturer's protocol. Quantification of extracted DNA was performed using Nanodrop and Qubit2.0 Fluorometer (Live Technologies, Carlsbad, CA, USA). Integrity of genomic DNA was verified by gel electrophoresis (0.6% agarose). DNA libraries were prepared for each wasp gender by acoustic shearing fragmentation using a Covaris S220. Fragments were end repaired and adenylated prior to adapter ligation on 3' ends (NEB NextUltra DNA Library Preparation kit, Illumina, San Diego, CA, USA). Enrichment and indexing of adapter-ligated DNA was done through limited cycle PCR. DNA library validation was performed using TapeStation (Agilent Technologies, Palo Alto, CA, USA). Libraries were quantified using Qubit 2.0 Fluorometer.

Real time PCR (Applied Biosystems, Carlsbad, CA, USA) was used to quantify DNA molar mass for each library before multiplexing in equal molar mass. DNA libraries were sequenced using a 2x150 paired-end (PE) configuration on one lane on an Illumina HiSeq 4000. Image analysis and base calling were performed using the HiSeq Control Software (HCS) on the HiSeq instrument.

The average size of inserts (without adaptors) in the Illumina library was ~300-350 bp. *De novo* assembly of reads and scaffolding of contigs was performed using ABySS 2.2 (*39*) by the New York Genome Center. *De novo* assembly of combined male/female genome was performed using Platanus-

allee (*40*) and scaffolding was improved using AGOUTI (*41*) on the University of Delaware's BIOMIX cluster.

Sequences from *Drosophila*-associated bacteria such as *Wolbachia* spp., *Acetobacter pasteurianus,* and *Lactobacillus plantarum* were identified in both assemblies. *Wolbachia* are endosymbionts of many insects including *Leptopilina* spp. (*42-44*). *Lactobacilli* and *Acetobacter* are symbionts and commensals of sugar-consuming insects (*45, 46*). Among the three bacterial species, *Wolbachia* sequences were the most abundant. BLASTx analysis showed that predicted genes from *Wolbachia* scaffolds were associated with *Wolbachia* proteins in GenBank. These bacterial and mitochondrial sequence-containing scaffolds were identified during the NCBI submission process and were manually removed from the submission.

Evaluation of genome assemblies: Assemblies made with ABySS and Platanus-allee with AGOUTI were run through QUAST v4.0 (*47*) to determine scaffold number, N50, and GC%. All assemblies were examined for conserved genes and orthologs with BUSCO v9 (*48, 49*) using the Insecta set and training parameters set to "Nasonia". NCBI BLAST+ (v 2.7.1) was used to compare select scaffolds produced from male and female genome assemblies (*50, 51*). E-value threshold was set at $1x10^{-7}$. E-values of alignments were considered acceptable if within the range of 0 to $1x10^{-10}$.

K-mer analysis was performed using the K-mer Analysis Toolkit (KAT) (*52*) and heat maps were used to compare multiplicity (coverage plus repeats) of K-mers to GC content of the reads, coloring bins according to the number of distinct K-mers in each. This analysis was used to determine whether there were separate clusters of multiplicity/GC content that might arise from different sources, such as contamination. BLAST (*50, 51, 53*) was used to search for homologs of a random sample of genomic scaffolds to which reads from each cluster mapped. The joint assembly of the *Lh 14* genome was compared to the published *L. clavipes* genome (Bioproject: PRJNA84205 (*54*)) through maps of 27-mer multiplicity versus GC content. Finally, 27-mer multiplicity/GC content of the scaffolds (9.6 Mb) containing MSEV genes was compared to a random subset of scaffolds (9.6 Mb) without MSEV genes. Statistical differences between *Lh 14* and *L. clavipes* genomes and between MSEV-gene containing scaffolds and non-MSEV-gene containing scaffolds were calculated using a multivariate Cramér test (*55-57*).

Gene predictions, gene annotation, and viral gene searches: Gene predictions were performed on parallel and anti-parallel strands using AUGUSTUS (v3.3.1; August 2018) (*58-60*) with the *Nasonia* training set. The AUGUSTUS readout was separated into mRNA, coding DNA sequence (CDS), and translations by gffread (*61*).

Gene predictions were annotated by performing a BLASTx of all gene predictions against the entire nr database (Downloaded on January, 2019) and InterProScan on the University of Delaware BIOMIX Cluster before using BLAST2GO (*20, 21*) to finish annotation based on BLASTx and InterProScan results.

NCBI BLAST+ (v 2.7.1) was used on a local machine to search predictions and scaffolds, cutoff was %ID >70%, E-value < 1E-50, and query coverage > 70%. MSEV genes and $1x10^{-2}$ for Polydnavirus and Nudivirus proteins. Family *Polydnaviridae* and *Nudiviridae* protein sequences for the 11 species available on OrthoDB v9 were downloaded on February 2019 (*50, 51*).

Whole-body transcriptome sequencing and assembly: Total RNA extraction, library preparations, sequencing reactions, and bioinformatics analysis were conducted at GENEWIZ, INC (South Plainfield, NJ, USA). RNA was extracted from frozen tissue with the Qiagen RNeasy Plus Universal mini kit using manufacturer's instructions (Qiagen, Hilden, Germany). The extracted RNA was quantified using a Qubit 2.0 Fluorometer and its integrity was checked with the 4200 TapeStation (Agilent Technologies, Palo Alto, CA, USA).

RNA samples were enriched for mRNA using Oligo d(T) beads. RNA sequencing libraries were prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina following manufacturer's instructions (NEB, Ipswich, MA, USA). The sequencing libraries were validated by using the Agilent TapeStation. Quantification was performed using the Qubit 2.0 Fluorometer and quantitative PCR (KAPA Biosystems, Wilimington, MA, USA).

Sequencing libraries were clustered on a single lane of a flow cell and sequenced on the Illumina HiSeq 4000 instrument using a 2x150 PE configuration. Image analysis and base calling were conducted

by the HCS. Raw sequence data (.bcl files) was converted into fastq files and de-multiplexed using Illumina's bcl2fastq 2.17 software. One mismatch was allowed for index sequence identification.

The Trinity v2.5 (*62*), *de novo* assembler was used to assemble the *Lh 14* transcripts. The *de novo* assembled transcriptome was created with a minimum contig length of 200 bp per sample. Transrate v1.0.3 (*63*) was used to generate statistics for the *de novo* assembled transcriptome. EMBOSS tools getorf were then used to find the ORFs within the *de novo* assembled transcriptome. The *de novo* transcriptome assembly was then annotated using Diamond BLASTx (*64*).

The transcriptome reads were mapped to the genomic scaffolds for downstream analyses using HISAT2 or BWA (*65, 66*).

Preparation of template DNA and PCR: Male and female wasps (n=12, for each sex), were separated and washed in 70% ethanol, and then rinsed twice in deionized water. Genomic DNA (gDNA) was extracted using a Qiagen DNeasy Blood and Tissue kit following provided protocols. gDNA was eluted in Tris-EDTA buffer, pH 8.0, and stored at 4°C. The concentration of gDNA was determined by NanoDrop (Thermo Fisher).

For cDNA preparation, male and female wasps (n=12 for each sex), were separated and washed in 70% ethanol and rinsed twice in deionized water. Total body RNA was extracted using 100 µL of Trizol (Invitrogen) following manufacturer's protocols. RNA was resuspended in 0.1% DEPC treated water and treated with DNase I to remove contaminating DNA (Thermo Fisher Scientific). The RNA concentration was determined by NanoDrop (Thermo Fisher). cDNA was synthesized using Proto-Script First Strand cDNA Synthesis Kit (New England Biolabs).

Analysis of select genes: Primers for *p40 and SmGTPase01* are as follows:

p40 forward: GAATCATTGTTCGTTTGCTTGAAGAAAGAATTGG

p40 reverse: CATTATTAATGGGCCTTTACAATAATTTTAGCC

SmGTPase01 forward: CGTTGCACTACCTTGTTTGTCA

SmGTPase01 reverse: TTGTCTTTGCCCTGAGCGTT

PCRs were performed with Taq polymerase (gift of C. Li lab, CCNY), PCR buffer (300 mM Tris HCl pH 9.5, 75 mM $(NH_4)_2SO_4$, 10 mM $MgCl_2$) and deoxyribonucleotides (0.2 mM; Thermo Fisher Scientific). The PCR products were resolved on a 1% agarose gel in Tris acetic acid EDTA buffer (40 mM Tris HCl pH 7.6, 20 mM acetic acid, 1mM EDTA pH 8.0). Ethidium bromide (Sigma Aldrich)-stained gels were visualized on an ultra violet Trans-Illuminator (UVP) and gel images were taken using the DigiDocIt Imaging System (UVP). Gel images were processed in Adobe Photoshop for clarity only.

gDNA or cDNA-containing PCR products were cloned into pCR TOPO II plasmids (Invitrogen) and transformed into DH10β competent cells (New England Biolabs). For plasmid preparation, colonies were screened via PCR and positive colonies were cultured in Luria Broth with ampicillin (100 µg/mL) at 37°C overnight. Plasmids were extracted using Plasmid Miniprep kit (Qiagen) and sequenced (GENEWIZ, INC. South Plainfield, NJ, USA). Sequences were aligned using NCBI BLAST+ (*50, 51*) and Clustal Omega (*34*). Expected PCR band sizes were determined using SerialCloner (v2.6.1).

Data availability: *L. heterotoma* strains (*13, 17*) are available upon request. All supplemental files and figures can be downloaded from figshare (https://figshare.com/s/3a4598308909307c2ae0). File S1 contains details of supplemental files and tables. File S2 contains listing of accession numbers for all sequences reported in this work. Figure S1 contains the 27-mer vs GC count comparison of MSEV containing scaffolds to non-MSEV containing scaffolds. Table S1 contains annotations and related data for proteins. Table S2 contains BLAST search results of the MSEV proteome against the nr database. Table S3 contains all BUSCOs found in male, female, and joint genome assemblies. MSEV protein sequences are available upon request. Accession numbers for datasets are as follows: *Leptopilina heterotoma* strain *Lh* 14, genome assembly: Male genome: QYUB0000000, Female genome: QYUC0000000, Joint genome: VOOK00000000. *Leptopilina heterotoma* strain *Lh* 14, whole-body transcriptome: GHUQ00000000. *Leptopilina heterotoma* abdominal transcriptome by Goecks et al.: GAJC0000000. *Leptopilina clavipes* genome Bioproject: PRJNA84205. *Leptopilina heterotoma* strain *Lh* 14 proteome: PRIDE: PXD005632

RESULTS AND DISCUSSION

The MSEV proteome superset: A comparative study of the proteomes of the MSEVs from *Lh* 14 and *Lh* NY strains previously generated a list of 161 “common” MSEV proteins (*15*). More than 90% of the 161 proteins are part of the *Lh* 14 MSEV proteome. To describe MSEVs more completely, we characterized a larger set of 407 MSEV proteins from *Lh* 14 (161 common and 246 *Lh* 14) and define this set as the *Lh* MSEV “super-set” (Fig. 1A). Key results from annotation-based classification, analysis for signal peptide and/or transmembrane domain, and presence/absence of proteins in related wasps are summarized below and in Table S1.

The presence/absence of signal peptide (SP) alone, or SP with/without the transmembrane (TM) domain(s) in MSEV proteins reveals their possible location (i.e., potentially secreted into the venom gland lumen or associated with MSEV membrane). We therefore searched the 246 *Lh* 14 proteins for SP and TM domains. Of the 246 proteins, 55 (22.35%) have a predicted SP domain, 37 (15.04%) have a predicted TM domain, while 6 (2.44%) have both a predicted SP and TM domain.

After annotation, we found that a majority (183/246 or 74.39%) of the 246 proteins can be classified as core eukaryotic cell biology proteins (Class 1); 13/246 (5.28%) proteins as virulence- and immunity-related based on associated GO terms (*22, 25*) (Class 2); and 50/246 or 20.32 % as novel sequences without high confidence annotation (Class 3) (Table S1). A presence/absence analysis of these 246 proteins in published transcriptomes (*19*) of *Lb* or a more distantly related wasp, *G. hookeri* (for thresholds see Methods) revealed the following: only 43/246 (17.47%) *Lh* MSEV proteins are expected to be found in *Lb* and/or *G. hookeri* (Table S1). Of these, 33/43 (76.74%) proteins were in Class 1 but only 7/43 (16.27%) and 3/43 (6.98%) were in Class 2 and Class 3 categories, respectively. These results support the idea that, multiple but different, infection strategies and/or host evasion strategies might exist among different wasps infecting the same hosts. (For a complete comparison of *Lh* MSEVs to *Lb* VLPs, see Addendum to Chapter 1.)

While most of the Class 1 proteins were annotated as ribosomal or mitochondrial-related, a few were described as integral membrane proteins, vesicle trafficking protein SEC22b (E-value: 6.22E-145), and the ion channels sideroflexin 1 and 2 (E-value: 0). We also identified an apolipophorin (E-value:

1.02E-7) (Table S1). The presence of these membrane-associated proteins reinforces the vesicular nature of MSEVs.

Examples of Class 2 proteins include the neural/ectodermal development factor IMP-L2 (E-value: $5.29x10^{-50}$) and a protein involved in pain reception, CG9231 (E-value: $4.39x10^{-15}$). A viral-like Diedel protein (E-value: $1.77x10^{-7}$), viral Enhancin (E-value: $6.02x10^{-5}$), l(2)37Cc (E-value: $3.39x10^{-165}$), odorant binding protein 56d-like (E-value: $5.64x10^{-50}$), major royal jelly protein (E-value: $8.59x10^{-135}$), and two venom acid phosphatases Acph-1 (E-value: $4.12x10^{-5}$) were also found in the Class 2 category; their cDNA sequences were published previously (*67*) (Table S1). It is possible that these MSEV proteins modulate the hosts' immune responses and/or influence host development to facilitate successful parasitism.

Within Class 3, 45 proteins (90%) lacked BLASTp and InterProScan results. However, Conserved Domain Database (CDD) (*26*) searches returned 9 hits identifying potentially functional domains (Table 1). This included (a) a CD99L2 like antigen (%ID: 24%, E-value: $3x10^{-3}$), (b) a DEAD-like helicases superfamily member (%ID: 22%, E-value $2x10^{-4}$) and (c) a herpes outer envelope glycoprotein 350 (gp350), (%ID: 28%, E-value: $4x10^{-3}$) (Table 1).

A BLASTp DELTA-BLAST of the potential gp350 domain against the nr database specifying "Vira" (taxid: 10239) under organism resulted in Crimean-Congo hemorrhagic fever orthonairovirus envelope glycoprotein (%ID: 30%, E-value: $2.5x10^{-1}$), Lymphocryptovirus Macaca gp350 (%ID: 29%, E-value: $7.2x10^{-1}$), and Gallid Alphaherpesvirus 1 envelope glycoprotein J (%ID: 26%, E-value: 1.2) as top hits. BLASTp DELTA-BLAST of the potential gp350 domain against the nr database for Hymenoptera yielded an uncharacterized protein as the best hit (%ID: 24%, E-value: $8x10^{-6}$) in the ant *Vollenhovia emeryi*. This ant protein is predicted to contain calcium-binding EGF domains. The second hit in this search is from *N. vitripennis* for a predicted mucin-3A like glycoprotein (*68*) (%ID: 24%, E-value: $2x10^{-4}$). Interestingly, transcripts related to the potential *Lh* gp350-like protein are not found in the *Lb* or *G. hookeri* transcriptomes (Table S1) (*19*). Presence of this gp350-like protein in *Lh* MSEVs, but its absence in *Lb* MSEVs, suggests that it somehow contributes to differences in *Lh*/*Lb* host-parasite interactions and is therefore worthy of future studies. Complement receptor type 2 (CR2) in human B lymphocytes interacts

with gp350 during Epstein-Barr infection (*69*) and finding a verified homologue of CR2 in *Drosophila* hosts would be interesting in future research.

Because more than 200 proteins have now been added to the previously described MSEV proteome (*15*), we re-evaluated our previous enrichment analysis. In an ortholog-based comparison of the superset to human extracellular vesicle (EV) proteomes in Vesiclepedia (the most current and robust source of EV data (*37*)), we found that the largest proportion of superset proteins (49%) are proteins specifically associated with exosomes (Fig. 1B). In human and mouse EV proteomes, mitochondrial and ribosomal proteins are enriched (*37*). Accordingly, protein components of mitochondria (e.g., respiratory chain) and ribosomes (e.g., large and small subunit proteins) are found to be highly enriched in the *Lh* MSEV superset. However, we found that the significance of the enrichment was higher between the superset and exosomal proteins than mitochondrial or ribosomal proteins (Fig. 1C). These results demonstrate the similarities in the protein profiles of MSEVs and EVs.

Do *Lh* MSEVs contain homologs of PDV or viral proteins? Even though figitid *Leptopilina* wasps are distantly related to PDV-containing Ichneumonid and Braconid wasps (*3, 70*), an association of PDV-like viruses in figitid wasps cannot be discounted because of shared evolutionary history. Recent publications have identified capsid-less VLPs in Ichneumonidae wasps (*6, 7, 71*) and it is possible that *Lh* MSEVs have a similar viral origin. We therefore analyzed the *Lh* MSEV proteome superset against the GenBank PDV database, and then against its entire Viridae database.

To identify false positives, MSEV proteins with positive PDV hits (E-values were less than $1.0x10^{-3}$, %ID was 20% or greater, and query coverage was 30% or higher) were also searched against the unrestricted nr database to compare relatedness. If an MSEV protein is similar to a viral or virus-related PDV protein, we expected that, in the unrestricted nr database search, the MSEV query sequence would align again with the same viral subject sequences, but with a lower E-value (Table S2).

For PDV searches (Taxid: 10482 and Taxid: 40273), four proteins returned hits with E-values better than $1.0x10^{-20}$ and query coverage greater than 30%. Three of these hits are conserved proteins (cytochrome P450 and histone 4) while the fourth result identified an uncharacterized *Cotesia congregata* bracovirus (CcBV) protein (%ID: 31.08%, E-value: $1.38x10^{-17}$, query coverage: 77%) (Table S2). The

unbiased BLASTp search against the entire nr database however had better results against eukaryotic proteins (E values: 0 to $2.0x10^{-7}$ and %ID from 100 to 56.25) (Table S2). In fact, the query that yielded the CcBV protein was better matched to a eukaryotic ribonuclease (%ID: 26.06%, E-value $1.14x10^{-16}$, query coverage: 84%) (Table S2). These results suggest that MSEV sequence similarities with PDV proteins may not be significant.

We also searched the *Lh* MSEV superset for presence of *L. boulardi* Filamentous Virus (LbFV) homologs (LbFV is a behavior manipulating virus of *Lb* (*72, 73*)). Of LbFV's 108 genes, 13 are present in genomes of *Lb*, *Lh* and related species, and the 13 transcripts are expressed in the *Lb* venom gland (*74*). Within our thresholds, we obtained only three (of 13) sequences with similarity to LbFV ORFs. However, these three *Lh* MSEV proteins, with hits for LbFV sequences obtained better scoring hits in the unrestricted nr database, suggesting that the *Lh* MSEV proteins are not highly related to the LbFV proteins (Table S2).

When comparing MSEV proteins to the entirety of Viridae, a total of 35 MSEV proteins had hits with %IDs ranging from 30% to 71% and E-values ranging from $1.0x10^{-22}$ to $1.0x10^{-178}$ (Table S2). However, a BLASTp search against the entire nr database found that proteins with results for viral hits had better scores when searched against the entire nr database, indicating that while viral hits are possible, they are not the best match (Table S2). This result, in addition to the fact that 372 other MSEV proteins (including the Diedel and Enhancin (Table S1)) did not return viral hits, would indicate that a majority of MSEV proteins are not closely related to viral proteins.

The whole-body transcriptome contains expressed MSEV transcripts: We performed a mixed-gender whole-body transcriptome sequencing and *de novo* assembly of *Lh* transcripts. This assembly generated 104,066 transcripts. This dataset is more than three times larger than the published data derived from abdomens of female wasps that has 31,400 transcripts (*19*). A BLAST analysis of the female abdominal transcripts against the male/female whole-body transcripts showed that a majority (21,493/31,400, 68.44%) were present in the latter data set.

We searched the whole-body transcriptome for MSEV protein coding sequences using tBLASTn. Of 407 MSEV superset proteins, we identified transcripts for 371 (91.15%) proteins. Despite the ~9%

discrepancy (likely due to differences in expression levels due to different experimental conditions), these results largely verify the transcript data from (*19*) that we have based our proteomic analyses on. Of the 371 MSEV transcripts identified, 233 (62.8%) encode Class 1, 44 (11.86%) encode Class 2, and 94/371 (25.34%) encode Class 3 proteins.

Assembly of the *Lh* genome: We separately sequenced *Lh* 14 male and female genomic DNA and assembled the paired-end reads *de novo*, using ABySS (*39*). These assemblies have a modest scaffold N50 of 4,800 with more than 100,000 scaffolds and an average coverage of 87% (Table 2). Assembly with MaSurCa (*75*) provided similar results (data not shown), indicating that our assembly quality is limited likely due to factors such as large genome size and repetitive sequence regions (*76*). Although the N50 values and large number of scaffolds indicate that the genome is not highly assembled, we found at least 80% of BUSCOs shared in the Insecta set in both assemblies (Table 2, BUSCOs in Table S3).

While still fragmented, a *de novo* joint assembly of male and female sequences using Platanus-allee and AGOUTI improved assembly and scaffolding statistics (N50: 11,906, average coverage 91.1%). The number of found BUSCOs in the joint assembly rose to 90% (Table 2, BUSCOs in Table S3).

Analysis of K-mer multiplicity versus GC content in the genome sequencing reads using the K-mer analysis tool, KAT (*52*) showed three possible clusters, although they are difficult to distinguish (Fig. 2A). Cluster 1 has high multiplicity (450-650), Cluster 2 has lower multiplicity and a wide range of GC content, and Cluster 3 has the lowest multiplicity and the highest GC content. Cluster 3 overlaps with Cluster 2 making them hard to fully separate. BLAST searches of a random sample (1,672/4,482) from Cluster 1 contigs hit insect homologs 73% (1,220/1,672) of the time, *Acetobacter* homologs 13% (216/1,672) of the time, and then a variety of mostly Eukaryotic hits. Cluster 2 represents a majority of the wasp genome (>94%), and blast hits of a random sample (316/68,173) of its contigs almost exclusively had homologs in Hymenoptera (311/316; 98%) and mostly in *L. heterotoma* (227/316; 71%). Cluster 3 is the smallest of the three and contigs from Cluster 3 had homologs exclusively in *Acetobacter* (110/110). There was no evidence for contamination from a viral source or discrete MSEV-specific set of nucleic acid sequence.

Furthermore, K-mer multiplicity versus GC content for the joint assembly of the *Lh 14* genome (Fig. 2B) showed a very similar heat map to that using the published assembly of *L. clavipes* (Fig. 2C; Bioproject: PRJNA84205 (*54*)). The two genomes have a highly similar 27-mer/GC profiles that do not differ statistically (multivariate Cramér test statistic = 114,119, $P = 0.73$, number bootstrap-replicates = 1000). The *L. heterotoma* assembly has 27-mers with approximately twice the multiplicity of those found in *L. clavipes*, which may represent increased repeat content in *L. heterotoma* and is supported by an assembly size of over 200 Mb larger than the *L. clavipes* genome (463 Mb vs 255 Mb) (*54*).

MSEV genes are encoded in the wasp genome: Using our annotation pipeline, 28,481 predicted genes were annotated. Within the annotated genes, we found 8 genes for the body color *yellow*, 3 *major royal jelly protein* (*mrjp*) genes, 25 odorant receptor/odorant binding protein coding genes, and 94 gene predictions for *cytochrome P450*. Some of these nuclear genes are not only involved in development and cellular processes, but are also included in the MSEV proteome (Table S1 and (*15*)). A search of gene predictions for MSEV proteins via tBLASTn identified 325 of 407 (79.85%) MSEV sequences (Table 3). Of these, 153/407 (37.6%) had a percent identity of 95% or greater. Presence or removal of scaffolds with bacterial DNA sequences from either the separate male/female or the joint assembly did not affect this number, supporting the nuclear location of a majority of the MSEV genes.

As gene prediction software can potentially miss genes (*77*), we searched the genomic scaffolds directly for MSEV-coding sequence regions using known protein sequences as queries via tBLASTn before and after removal of bacterial sequences. In both cases, 375/407 (92.13%) MSEV sequences were at least 70% complete as determined by query coverage (Table 3). Of these, 191/407 (46.9%) had a percent identity of 95% or greater.

The scaffolds containing MSEV genes (Fig. S1A) were also compared to a random subset of scaffolds without MSEV genes (Fig S1B) for their 27-mer/GC profiles. These appeared to not differ statistically (multivariate Cramér test statistic = 3755, $P = 0.80$, number bootstrap-replicates = 1000), indicating that the MSEV genes lie on scaffolds that resemble the rest of the genome.

Characterization of select MSEV genes: We spot checked small portions of the genome for gene structure predictions of MSEV virulence protein genes *SmGTPse01* (Class 2) and *p40* (Class 3). For this, we sequenced PCR products of gDNA corresponding to these genes.

The MSEV SmGTPase01 has prokaryotic-like GTPase domains and its gene is expected to lack introns (*15*). The predicted *SmGTPase01* CDS spans 936 bp, which contains the functional GTPase domain (*78*). Scaffolds from male and female genomes confirmed the absence of coding region introns (data not shown). We hypothesized that primers in 5' and 3' untranslated regions (UTRs) should amplify the exact fragment from cDNA/gDNA as template based on manual characterization of the *SmGTPase01* locus (*78*) (Fig. 3A). This prediction was borne out and we amplified an 873 bp fragment only from female cDNA and from both male and female gDNA (Fig. 3C). The sequenced PCR products were identical to corresponding sequence within the assembly and the published transcript sequence from Goecks et al. (data not shown).

The *p40* gene encodes a protein that is structurally similar to T3SS bacterial needle tip proteins IpaD/SipD from *Shigella* and *Salmonella*. However, *p40*'s genomic sequence is expected to have introns (*15*). The full p40 gene was computationally assembled and predicted within both male and female genomes. Primers designed for *p40*'s 5' and 3' UTRs (*19, 79*) (Fig. 3B), allowed amplification of *p40*'s 939 bp cDNA only in preparations from female wasp extracts, but gDNA bands at 1,630 bp were detected from reactions when either male or female genome was used as template, indicating the presence of introns (Fig. 3D). Sequencing the cloned cDNA product from females confirmed the published cDNA sequence (Heavner et al., 2017). We also cloned and sequenced the gDNA products from male and female wasps and found the sequences to be identical (data not shown).

Unlike the well-characterized *Drosophila* hosts, the biology and molecular-genetics of their parasitic wasps have remained relatively obscure with only recent characterizations of *Leptopilina* and *Ganaspis* spp. (*15, 19, 67, 74, 80-83*). Our proteomic, transcriptomic and genomic results here expand the available information on *L. heterotoma*. Bioinformatics analysis of the additional MSEV proteins does not alter the initial interpretation of the original 161 proteomic data. Genomic sequencing and analysis of scaffolds reveals that more than 92% of the MSEV genes reside on the wasp genome. We did not find evidence for MSEV gene association with endosymbiont or commensal bacterial DNA. We suspect that

the remaining ~8% are also nuclear genes and this association will be confirmed in higher quality assemblies. Altogether, these results strongly suggest that, like other subcellular structures, MSEVs are encoded in the wasp nuclear genome.

The cellular nature of *Lh* vesicles is likely to be shared by closely related *Lv* and *Lb* wasps. Our previous work has shown that the overall morphologies of *Lh* and *Lv* MSEVs are similar (*11, 13*). However, this is not the case for *Lb* MSEVs; different *Lb* strains have varying MSEV morphologies (*10, 84, 85*). Interpretation of their identity also varies. For example, Di Giovanni et al. (*74*) contend that MSEVs/VLPs are derived from a virus ancestral to the LbFV. Our analysis of the expanded proteomic superset does not lend strong support to this line of thinking.

We did not find convincing evidence of PDV or other viral structural proteins in the *Lh* MSEV proteome. However, we cannot discount that MSEVs have a viral origin as our analysis is limited by the fragmentation of the genome. It is also possible that a virus related to MSEVs may not have been identified to date. Mechanistically, eukaryotic viruses and vesicles share cellular pathways involving the endomembrane systems of their cells of origin or their target cells (*86*), leading to overlap in protein functionality, but not necessarily origin. Thus, at least some of the Class 1 proteins in the MSEV proteome may be central to MSEV biogenesis in the wasp or for their interactions with the host hemocytes' endomembrane machinery despite potentially being related to viruses. It is noteworthy that energy metabolism genes appear to be involved in rapid speciation and adaptation to new environments (*87, 88*), raising the possibility that MSEV mitochondrial proteins might contribute to this process. How *Lh* MSEVs are functionally similar to other insect or mammalian EVs remains to be explored experimentally. Functional characterization of predicted infection and immunity Class 2 proteins should explain the immune-suppressive strategies of these wasps. RNA interference, infection assays, and other experimental strategies should make this line of inquiry feasible.

Functional assignments are difficult for the unannotated Class 3 proteins. These are likely to be quite interesting, due to their different expression profiles in *Lh* versus *Lb* and *G. hookeri* species. This difference in expression may stem either from *cis* changes in their regulatory sequences, or from absence of these genes in the *Lb* or *G. hookeri* genomes. Recent comparative genomics analysis has shown that over 40% of venom genes in the closely-related species *N. vitripennis* and *N. giraulti* have diverged

significantly and up to 25% of venom genes are specific to a species (*89*). A proteomic analysis of the venom genes of *Leptopilina spp*. and a molecular understanding of their expression will provide insights into how key activities within MSEVs evolved to parasitize the range of fruit fly hosts.

A key question regarding *Lh* virulence proteins critical to wasp success is whether their genes reside in a discrete region of the genome like a “virulence island” found in some microbial genomes (*90, 91*), or whether some genes are dispersed within the genome, while others occur in one or more clusters as in wasps with PDVs (*6, 71*). More complete assemblies, scaffolded to the level of chromosomes, will describe the genome-wide distribution of these genes in *Lh* and related wasps. Key MSEV genes could serve as genetic markers in future studies. Comparative genomics will uncover additional gene family members of MSEV proteins in other *Leptopilina* wasps and enable the development of new functional genomics tools such as CRISPR-disrupted mutant alleles made in *N. vitripennis* (*92-95*). These approaches will open new avenues for understanding the biology of this host-parasite model.

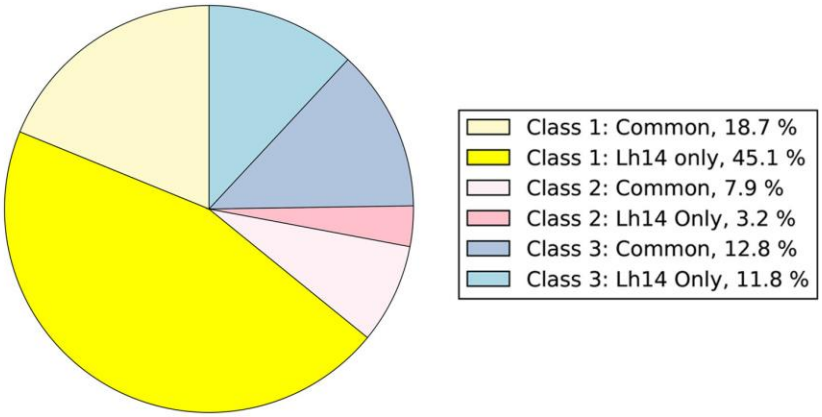
ACKNOWLEDGEMENTS

We thank Drs. W. Qiu and S. Singh, and J. Chou for discussions and critical comments, and our reviewers for insightful feedback. We are grateful to A. Corvelo at the New York Genome Center for help with genome assemblies. Bioinformatics work was conducted in-house and with the BIOMIX Shared Computing Cluster at Delaware Biotechnology Institute, University of Delaware. This work was supported by grants from NASA (NNX15AB42G), NSF (IOS-1121817), NIH (1F31GM111052-01A1, 5G12MD007603-30, and GM103446).

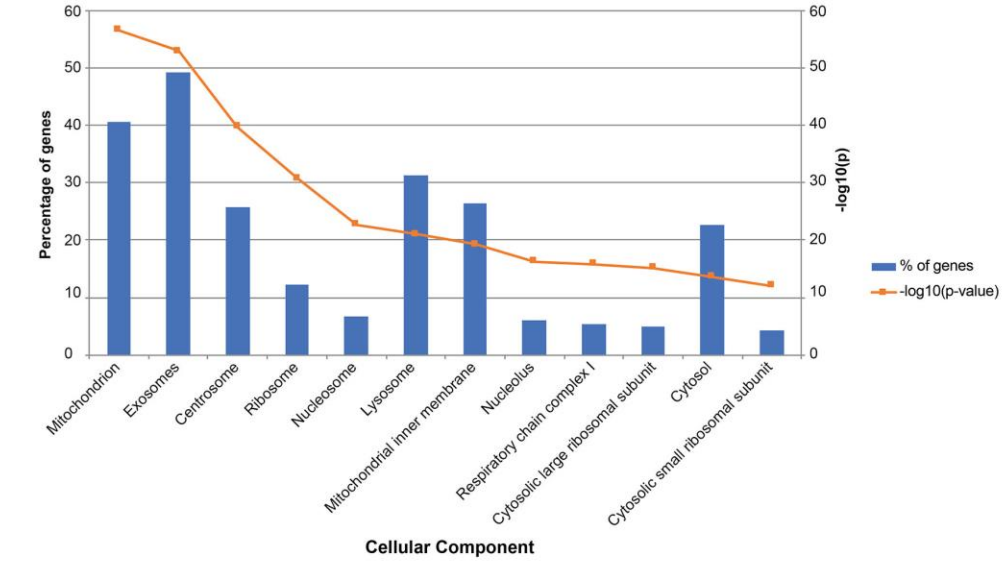
FIGURES

Figure 1. The superset of MSEV proteins: (A) *Lh* 14 MSEV proteins were annotated using BLAST2GO prior to class sorting via annotation and GO Terms. Wedges denoted as "Common," were previously published in (*15*) and represent proteins found in both *Lh* 14 and *Lh* NY MSEV proteomes. New proteins analyzed in this work are in wedges labeled "*Lh* 14 Only." A majority of proteins belong to Class 1. Table S1 lists 246 proteins added to the superset *Lh* 14 proteome. (B) and (C) Enrichment analysis of MSEV superset shows high association with exosomes and mitochondria compared to other cellular organelles according to Vesiclepedia. $-\log_{10}$ (p-value) trend shown in orange for both graphs. The p-values were calculated with the Bonferroni method. (B) Percentage of MSEV genes associated with specific cellular compartments found in Vesiclepedia, relative to all MSEV genes. Of the superset proteins, 41 and 49% are associated with mitochondria and exosomes, respectively ($p = 3 \times 10^{-57}$; 1×10^{-53}). (C) Fold-enrichment of the MSEV dataset in specific cellular compartments. Although many protein classes are present in the proteome, exosomal and mitochondrial proteins show more significant enrichments.

A.


Class 1: Common, 18.7 %
Class 1: Lh14 only, 45.1 %
Class 2: Common, 7.9 %
Class 2: Lh14 Only, 3.2 %
Class 3: Common, 12.8 %
Class 3: Lh14 Only, 11.8 %

B.


Percentage of genes
-log10(p)
% of genes
-log10(p-value)
Mitochondrion
Exosomes
Centrosome
Ribosome
Nucleosome
Lysosome
Mitochondrial inner membrane
Nucleolus
Respiratory chain complex I
Cytosolic large ribosomal subunit
Cytosol
Cytosolic small ribosomal subunit
Cellular Component

C.

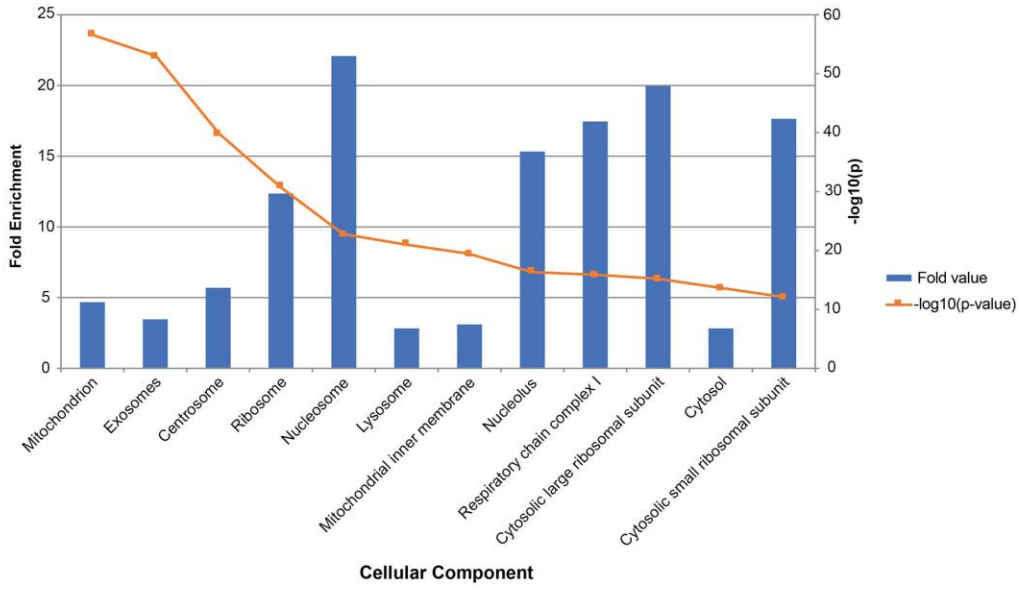

Fold Enrichment
-log10(p)
Fold value
-log10(p-value)
Mitochondrion
Exosomes
Centrosome
Ribosome
Nucleosome
Lysosome
Mitochondrial inner membrane
Nucleolus
Respiratory chain complex I
Cytosolic large ribosomal subunit
Cytosol
Cytosolic small ribosomal subunit
Cellular Component

Figure 2. Analysis of K-mer coverage versus GC count. (A) Analysis of genomic reads. 27-mers generated from the cleaned Illumina reads used to assemble the *L. heterotoma* genome binned by their GC count vs multiplicity (total counts among the reads). Bins are colored by the number of distinct K-mers. Different clusters are identified as shown and described in the text. (B and C): A map of 27-mer multiplicity versus GC content of the joint assembly of the *Lh 14* genome (B) to a map from the published *L. clavipes* genome (Bioproject: PRJNA84205) (C).

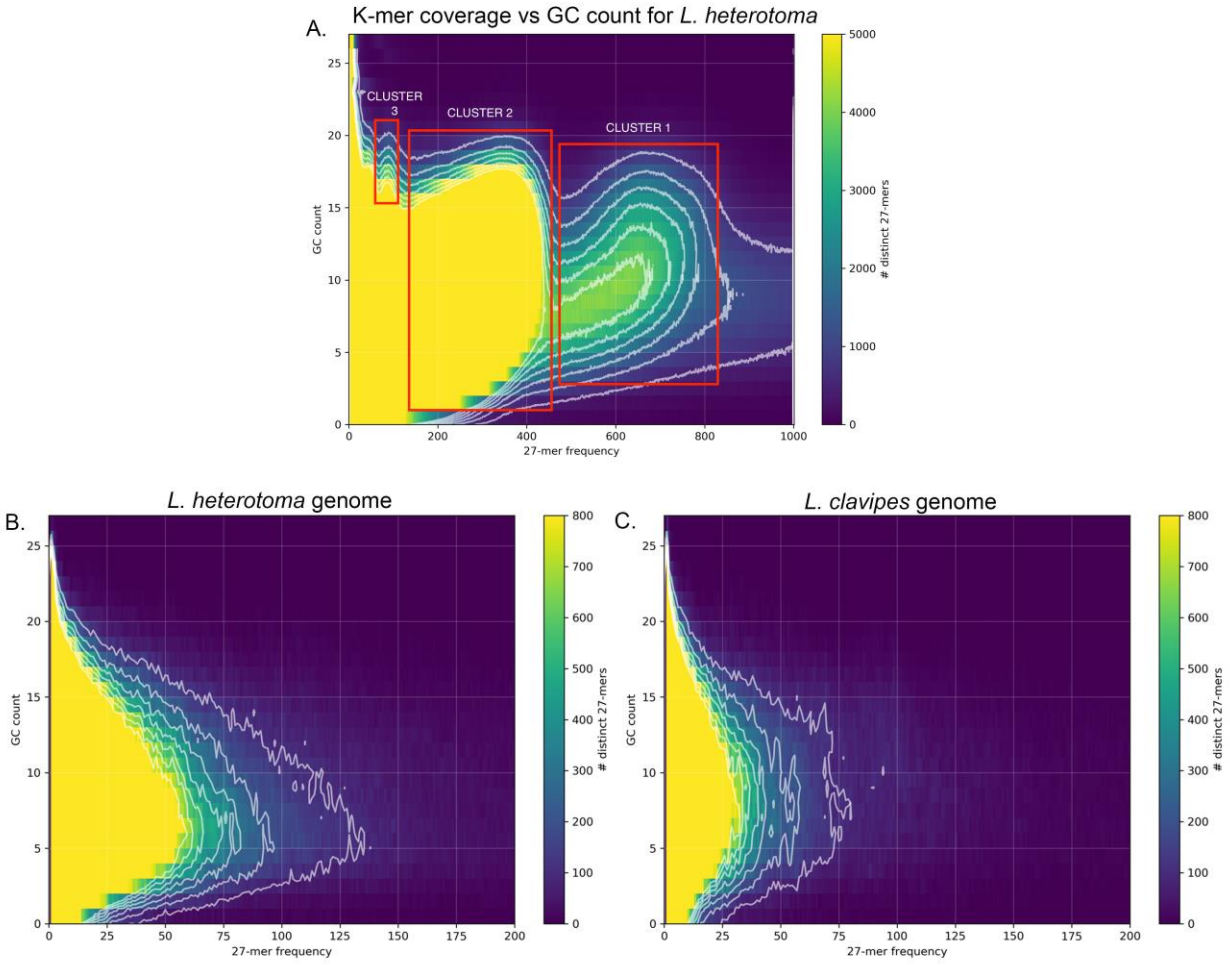
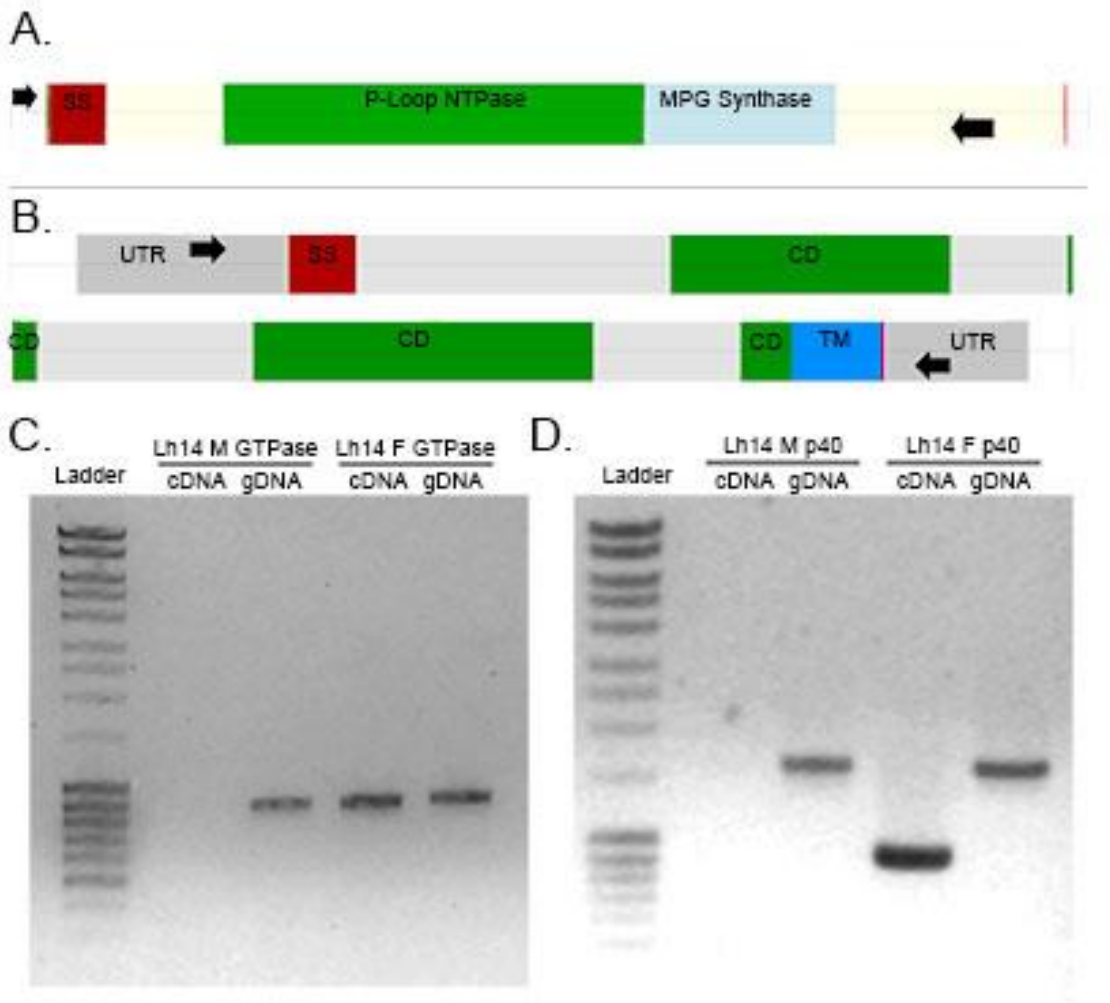


Figure 3. Predicted gene structures verified by PCR amplification experiments (A, B). Diagrams showing primer locations and predicted gene structures of *SmGTPase01* (A) and *p40* (B). Black arrows indicate primer locations, light gray indicates introns, UTR regions are dark gray and labeled, exons encoding potential protein domains are labeled as shown. Cream colored regions in panel A do not have a specified domain. Diagrams were drawn using GenomeDiagram as part of the Biopython (v. 1.6) package (*96, 97*). Each row in the panels A and B diagrams corresponds to approximately 1,000 bp. For primer sequences, see methods. (C and D) Ladder is Thermo Fisher MassRuler ladder. (C) PCR products for *SmGTPase01* from male or female cDNA and gDNA. All products are 873 bp long. Male cDNA PCR was negative. (D) PCR products for *p40* from male or female cDNA and gDNA. The expected band for *p40* cDNA is 939 bp and for gDNA is 1,630 bp. Male cDNA PCR was negative. Sequence analysis of PCR amplification products confirmed gene prediction results.



TABLES

Table 1. CDD-search results of MSEV "un-annotated" proteins in the super-set. MSEV ORFs that completed the BLAST2GO pipeline and did not return any results were run through the NCBI CDD-Search Version 3.16 (Accessed: Aug. 2018). Of 45 queries, only 9 returned hits with threshold set to $1x10^{-2}$. The ninth result came from a search with E-value threshold set to 1. Results listed are all unique, high scoring hits for each ORF that returned hits from the search.

MSEV SUPERSET UNKNOWNS CDD-SEARCH RESULTS						
Query (in-house ID)	PSSM-ID	From	To	E-Value	Accession	Short name
GAJC01013214.1_14	**331760**	**25**	**98**	**0.000176**	cl26939	DEXDc superfamily
GAJC01012558.1_12	**330317**	**39**	**205**	**0.003987**	cl25496	Herpes_BLLF1 superfamily/gp350
GAJC01011863.1_13	**311912**	**86**	**187**	**0.003653**	cl07006	RNA_polI_A34 superfamily
GAJC01011463.1_48	**315064**	**234**	**335**	**0.002964**	cl13702	CD99L2 superfamily
GAJC01010930.1_16	**328726**	**32**	**61**	**0.001252**	cl21457	ICL_KPHMT superfamily
GAJC01010353.1_14	**331876**	**31**	**121**	**0.001483**	cl27055	MutS_III superfamily
GAJC01009713.1_25	**311628**	**138**	**225**	**0.000133**	cl06688	TSGP1 superfamily
GAJC01009493.1_4	**328724**	**79**	**96**	**0.001983**	cl21455	P-loop_NTPase superfamily
GAJC01002124.1_43	**330572**	**4**	**269**	**0.0073146**	cl25751	DUF4045 superfamily

Table 2. Assembly statistics: Statistics of male, female, and combined (male plus female) *Lh* genomes as assessed by QUASTv4.0 and BUSCOv9.0. Percent coverage was found by mapping sequencing reads back to assembly using HISAT2. The identified BUSCOs can be found in Table S1. The QUAST program was run with parameters set for eukaryotic genomes and scaffolds. The BUSCO program was run with species set to 'Nasonia.' Contigs smaller than 500 bp were excluded.

ASSEMBLY STATISTICS				
		Male	Female	Joint
Assembly	N50 (bp)	**4,779**	**4,843**	**11906**
	No. scaffolds	**147,558**	**147,549**	**83,487**
	Largest scaffold (bp)	**306,667**	**176,371**	**375,275**
	Total length (bp)	**474,383,205**	**472,302,230**	**462,564,754**
	GC%	**27.54**	**27.28**	**27.84**
	Coverage (%)	**87.7**	**86.8**	**91.1**
BUSCOs (Insecta)	Complete	**80.20%**	**81.30%**	**90.90%**
	Single	**69.40%**	**71.80%**	**89.20%**
	Duplicated	**10.80%**	**9.50%**	**1.70%**
	Fragmented	**15.90%**	**15.10%**	**6.50%**
	Missing	**3.90%**	**3.60%**	**2.60%**
	n	**1658**	**1658**	**1658**

Table 3. MSEV genes found in scaffolds and predictions: Gene predictions from genome assembly scaffolds and AUGUSTUS gene predictions were searched for MSEV genes using tBLASTn. Results better than %ID >70%, E-value < $1x10^{-50}$, and query coverage > 70% were retained.

MSEV GENES FOUND IN GENOME ANALYSIS				
	MSEV BLASTn scaffold results		AUGUSTUS prediction results	
	Found	Percentage	Found	Percentage
Female	**278**	**68.3**	**169**	**41.52**
Male	**275**	**67.58**	**166**	**40.78**
Shared in M+F	**265**		**159**	
Joint Assembly	**375**	**92.13%**	**325**	**79.85%**

REFERENCES

1. T. C. Narendran, in *Biocontrol Potential and its Exploitation in Sustainable Agriculture: Volume 2: Insect Pests,* R. K. Upadhyay, K. G. Mukerji, B. P. Chamola, Eds. (Springer US, Boston, MA, 2001), pp. 1-12.
2. J. J. Rodriguez *et al.*, Extrapolations from field studies and known faunas converge on dramatically increased estimates of global microgastrine parasitoid wasp species richness (Hymenoptera: Braconidae). *Insect Conserv Diver* **6**, 530-536 (2013).
3. M. R. Strand, G. R. Burke, Polydnaviruses: From discovery to current insights. *Virology* **479-480**, 393-402 (2015).
4. J. M. Drezen *et al.*, Endogenous viruses of parasitic wasps: variations on a common theme. *Curr Opin Virol* **25**, 41-48 (2017).
5. J. Gauthier, J. M. Drezen, E. A. Herniou, The recurrent domestication of viruses: major evolutionary transitions in parasitic wasps. *Parasitology* **145**, 713-723 (2018).
6. A. Pichon *et al.*, Recurrent DNA virus domestication leading to different parasite virulence strategies. *Sci Adv* **1**, e1501150 (2015).
7. G. R. Burke, T. J. Simmonds, B. J. Sharanowski, S. M. Geib, Rapid Viral Symbiogenesis via Changes in Parasitoid Wasp Genome Architecture. *Mol Biol Evol* **35**, 2463-2474 (2018).
8. E. S. Keebaugh, T. A. Schlenke, Insights from natural host-parasite interactions: The Drosophila model. *Developmental and Comparative Immunology* **42**, 111-123 (2014).
9. T. M. Rizki, R. M. Rizki, Y. Carton, Leptopilina heterotoma and L. boulardi: strategies to avoid cellular defense responses of Drosophila melanogaster. *Exp Parasitol* **70**, 466-475 (1990).
10. S. Dupas, M. Brehelin, F. Frey, Y. Carton, Immune suppressive virus-like particles in a Drosophila parasitoid: significance of their intraspecific morphological variations. *Parasitology* **113 (Pt 3)**, 207-212 (1996).
11. J. Morales *et al.*, Biogenesis, structure, and immune-suppressive effects of virus-like particles of a Drosophila parasitoid, Leptopilina victoriae. *J Insect Physiol* **51**, 181-195 (2005).
12. C. Labrosse *et al.*, A RhoGAP protein as a main immune suppressive factor in the Leptopilina boulardi (Hymenoptera, Figitidae) - Drosophila melanogaster interaction. *Insect Biochem Molec* **35**, 93-103 (2005).
13. H. Chiu, J. Morales, S. Govind, Identification and immuno-electron microscopy localization of p40, a protein component of immunosuppressive virus-like particles from Leptopilina heterotoma, a virulent parasitoid wasp of Drosophila. *J Gen Virol* **87**, 461-470 (2006).
14. M. E. Heavner, A. D. Hudgins, R. Rajwani, J. Morales, S. Govind, Harnessing the natural D*rosophila*-parasitoid model for integrating insect immunity with functional venomics. *Curr Opin Insect Sci* **6**, 61-67 (2014).
15. M. E. Heavner *et al.*, Novel Organelles with Elements of Bacterial and Eukaryotic Secretion Systems Weaponize Parasites of Drosophila. *Current biology : CB* **27**, 2869-2877.e2866 (2017).
16. H. Chiu, S. Govind, Natural infection of D. melanogaster by virulent parasitic wasps induces apoptotic depletion of hematopoietic precursors. *Cell death and differentiation* **9**, 1379-1381 (2002).
17. T. A. Schlenke, J. Morales, S. Govind, A. G. Clark, Contrasting infection strategies in generalist and specialist wasp parasitoids of Drosophila melanogaster. *PLoS Pathog* **3**, 1486-1501 (2007).
18. C. Small, I. Paddibhatla, R. Rajwani, S. Govind, An introduction to parasitic wasps of Drosophila and the antiparasite immune response. *J Vis Exp*, e3347 (2012).
19. J. Goecks *et al.*, Integrative approach reveals composition of endoparasitoid wasp venoms. *PloS one* **8**, e64125 (2013).
20. A. Conesa *et al.*, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676 (2005).
21. S. Götz *et al.*, High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* **36**, 3420-3435 (2008).
22. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
23. P. Jones *et al.*, InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).
24. U. Consortium, UniProt: a hub for protein information. *Nucleic Acids Res* **43**, D204-212 (2015).

25. S. Carbon *et al.*, The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research* **47**, D330-D338 (2019).
26. A. Marchler-Bauer *et al.*, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* **45**, D200-D203 (2017).
27. L. Kall, A. Krogh, E. L. Sonnhammer, A combined transmembrane topology and signal peptide prediction method. *Journal of molecular biology* **338**, 1027-1036 (2004).
28. L. Kall, A. Krogh, E. L. L. Sonnhammer, Advantages of combined transmembrane topology and signal peptide prediction - the Phobius web server. *Nucleic Acids Research* **35**, W429-W432 (2007).
29. H. Nielsen, Predicting Secretory Proteins with SignalP. *Methods Mol Biol* **1611**, 59-73 (2017).
30. J. J. Almagro Armenteros *et al.*, SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420-423 (2019).
31. E. L. Sonnhammer, G. von Heijne, A. Krogh, A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**, 175-182 (1998).
32. A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* **305**, 567-580 (2001).
33. M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of molecular biology* **428**, 726-731 (2016).
34. W. Li *et al.*, The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res* **43**, W580-584 (2015).
35. M. Pathan *et al.*, FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics* **15**, 2597-2601 (2015).
36. M. Pathan *et al.*, A novel community driven software for functional enrichment analysis of extracellular vesicles data. *J Extracell Vesicles* **6**, 1321455 (2017).
37. H. Kalra *et al.*, Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation. *PLoS Biol* **10**, e1001450 (2012).
38. M. Pathan *et al.*, Vesiclepedia 2019: a compendium of RNA, proteins, lipids and metabolites in extracellular vesicles. *Nucleic Acids Res* **47**, D516-D519 (2019).
39. S. D. Jackman *et al.*, ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res* **27**, 768-777 (2017).
40. R. Kajitani *et al.*, Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. *Nat Commun* **10**, (2019).
41. S. V. Zhang, L. T. Zhuo, M. W. Hahn, AGOUTI: improving genome assembly and annotation using transcriptome data. *Gigascience* **5**, (2016).
42. B. A. Pannebakker, L. P. Pijnacker, B. J. Zwaan, L. W. Beukeboom, Cytology of Wolbachia-induced parthenogenesis in Leptopilina clavipes (Hymenoptera: Figitidae). *Genome* **47**, 299-303 (2004).
43. J. H. Werren, L. Baldo, M. E. Clark, Wolbachia: master manipulators of invertebrate biology. *Nature reviews. Microbiology* **6**, 741-751 (2008).
44. G. Gueguen, B. Onemola, S. Govind, Association of a new Wolbachia strain with, and its effects on, Leptopilina victoriae, a virulent wasp parasitic to Drosophila spp. *Applied and environmental microbiology* **78**, 5962-5966 (2012).
45. E. Crotti *et al.*, Acetic Acid Bacteria, Newly Emerging Symbionts of Insects. *Applied and environmental microbiology* **76**, 6963-6970 (2010).
46. P. Engel, N. A. Moran, The gut microbiota of insects - diversity in structure and function. *FEMS microbiology reviews* **37**, 699-735 (2013).
47. A. Mikheenko, G. Valin, A. Prjibelski, V. Saveliev, A. Gurevich, Icarus: visualizer for de novo assembly evaluation. *Bioinformatics* **32**, 3321-3323 (2016).
48. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
49. R. M. Waterhouse *et al.*, BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*, (2017).

50. M. Johnson *et al.*, NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**, W5-9 (2008).
51. C. Camacho *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
52. D. Mapleson, G. Garcia Accinelli, G. Kettleborough, J. Wright, B. J. Clavijo, KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* **33**, 574-576 (2016).
53. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410 (1990).
54. K. Kraaijeveld *et al.*, Decay of Sexual Trait Genes in an Asexual Parasitoid Wasp. *Genome Biology and Evolution* **8**, 3685-3695 (2016).
55. R. Ihaka, R. Gentleman, R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **5**, 299-314 (1996).
56. L. Baringhaus, C. Franz, On a new multivariate two-sample test. **88**, 190-206 (2004).
57. C. Franz. (2019).
58. M. Stanke, R. Steinkamp, S. Waack, B. Morgenstern, AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* **32**, W309-312 (2004).
59. M. Stanke, B. Morgenstern, AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* **33**, W465-467 (2005).
60. O. Keller, M. Kollmar, M. Stanke, S. Waack, A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **27**, 757-763 (2011).
61. C. Trapnell *et al.*, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578 (2012).
62. M. G. Grabherr *et al.*, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-U130 (2011).
63. R. Smith-Unna, C. Boursnell, R. Patro, J. M. Hibberd, S. Kelly, TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Research* **26**, 1134-1144 (2016).
64. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59-60 (2015).
65. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
66. D. Kim, B. Landmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-U121 (2015).
67. M. E. Heavner *et al.*, Partial venom gland transcriptome of a Drosophila parasitoid wasp, Leptopilina heterotoma, reveals novel and shared bioactive profiles with stinging Hymenoptera. *Gene* **526**, 195-204 (2013).
68. S. J. Gendler, A. P. Spicer, Epithelial mucin genes. *Annu Rev Physiol* **57**, 607-634 (1995).
69. K. A. Young, A. P. Herbert, P. N. Barlow, V. M. Holers, J. P. Hannan, Molecular basis of the interaction between complement receptor type 2 (CR2/CD21) and Epstein-Barr virus glycoprotein gp350. *J Virol* **82**, 11217-11227 (2008).
70. B. Misof *et al.*, Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763-767 (2014).
71. A. N. Volkoff *et al.*, Analysis of virion structural components reveals vestiges of the ancestral ichnovirus genome. *PLoS Pathog* **6**, e1000923 (2010).
72. J. Varaldi, S. Petit, M. Bouletreau, F. Fleury, The virus infecting the parasitoid Leptopilina boulardi exerts a specific action on superparasitism behaviour. *Parasitology* **132**, 747-756 (2006).
73. S. Patot, R. Allemand, F. Fleury, J. Varaldi, An inherited virus influences the coexistence of parasitoid species through behaviour manipulation. *Ecol Lett* **15**, 603-610 (2012).
74. D. Di Giovanni *et al.*, A behavior-manipulating virus relative as a source of adaptive genes for parasitoid wasps. *bioRxiv*, 342758 (2019).
75. A. V. Zimin *et al.*, The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669-2677 (2013).
76. V. Dominguez Del Angel *et al.*, Ten steps to get started in Genome Assembly and Annotation. *F1000Res* **7**, (2018).
77. Z. Wang, Y. Chen, Y. Li, A brief review of computational gene prediction methods. *Genomics Proteomics Bioinformatics* **2**, 216-221 (2004).
78. M. E. Heavner, City University of New York, Graduate Center (2018).
79. J. Ramroop, Dissertation, Graduate Center, City University of New York, (2016).

80. J. P. Melk, S. Govind, Developmental analysis of Ganaspis xanthopoda, a larval parasitoid of Drosophila melanogaster. *J Exp Biol* **202**, 1885-1896 (1999).
81. D. Colinet *et al.*, Extensive inter- and intraspecific venom variation in closely related parasites targeting the same host: the case of Leptopilina parasitoids of Drosophila. *Insect Biochem Mol Biol* **43**, 601-611 (2013).
82. N. T. Mortimer *et al.*, Parasitoid wasp venom SERCA regulates Drosophila calcium levels and inhibits cellular immunity. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9427-9432 (2013).
83. S. Khan, D. T. Sowpati, R. K. Mishra, Long-read genome sequence and assembly of Leptopilina boulardi: a specialist Drosophila parasitoid. *bioRxiv*, (2018).
84. G. Gueguen, R. Rajwani, I. Paddibhatla, J. Morales, S. Govind, VLPs of Leptopilina boulardi share biogenesis and overall stellate morphology with VLPs of the heterotoma clade. *Virus Res* **160**, 159-165 (2011).
85. B. Wan *et al.*, Venom Atypical Extracellular Vesicles as Interspecies Vehicles of Virulence Factors Involved in Host Specificity: The Case of a Drosophila Parasitoid Wasp. *Frontiers in Immunology* **10**, (2019).
86. E. Nolte-'t Hoen, T. Cremer, R. C. Gallo, L. B. Margolis, Extracellular vesicles and viruses: Are they close relatives? *Proceedings of the National Academy of Sciences* **113**, 9155-9161 (2016).
87. M. Gershoni, A. R. Templeton, D. Mishmar, Mitochondrial bioenergetics as a major motive force of speciation. *Bioessays* **31**, 642-650 (2009).
88. N. Lane, On the origin of bar codes. *Nature* **462**, 272-274 (2009).
89. E. O. Martinson, Mrinalini, Y. D. Kelkar, C. H. Chang, J. H. Werren, The Evolution of Venom by Co-option of Single-Copy Genes. *Current biology : CB* **27**, 2007-2013.e2008 (2017).
90. U. Dobrindt, B. Hochhut, U. Hentschel, J. Hacker, Genomic islands in pathogenic and environmental microorganisms. *Nature reviews. Microbiology* **2**, 414-424 (2004).
91. O. Gal-Mor, B. B. Finlay, Pathogenicity islands: a molecular toolbox for bacterial virulence. *Cell Microbiol* **8**, 1707-1719 (2006).
92. J. H. Werren, D. W. Loehlin, J. D. Giebel, Larval RNAi in Nasonia (parasitoid wasp). *Cold Spring Harb Protoc* **2009**, pdb.prot5311 (2009).
93. A. L. Siebert, D. Wheeler, J. H. Werren, A new approach for investigating venom function applied to venom calreticulin in a parasitoid wasp. *Toxicon*, (2015).
94. M. Li, M. Bui, O. S. Akbari, Embryo Microinjection and Transplantation Technique for Nasonia vitripennis Genome Manipulation. *J Vis Exp*, (2017).
95. M. Li *et al.*, Generation of heritable germline mutations in the jewel wasp Nasonia vitripennis using CRISPR/Cas9. *Sci Rep* **7**, 901 (2017).
96. L. Pritchard, J. A. White, P. R. J. Birch, I. K. Toth, GenomeDiagram: a python package for the visualization of large-scale genomic data. *Bioinformatics* **22**, 616-617 (2006).
97. P. J. A. Cock *et al.*, Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423 (2009).

SUPPLEMENTAL DATA

Figure S1: K-mer coverage vs GC count of MSEV-containing scaffolds vs non MSEV-containing scaffolds. (A) MSEV scaffolds (9.6 Mb). (B) Random subset of non-MSEV scaffolds (9.6 Mb). See Fig 2 for plot details.

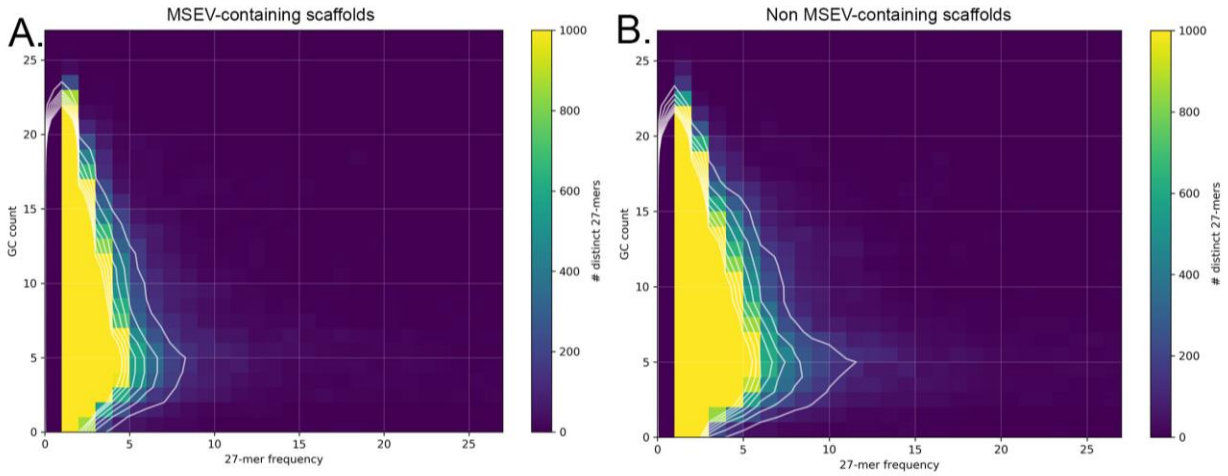


Table S1: Lh 14 Proteins of the MSEV Superset. 246 proteins were annotated using BLAST2GO prior to sorting into classes and additional analysis. Table indicates in-house query accession number, class, annotation, E-value for best BLASTp hit during annotation, presence/absence of signal peptide or transmembrane domain, and presence/absence in L. boulardi (Lb) or G. hookeri (G1) transcriptomes with accession values and E-values. NA in annotation column indicates no annotation.

TABLE S1: *Lh 14* Proteins of the MSEV Superset										
In-house Accession number	Class	Annotation	E-value	Signal Peptide	Transmembrane domain	Lb Accession number	Lb E-value	G1 Accession number	G1 E-value	
GAJC01031099.1_25	Class 1	charged multivesicular body protein 6	4.45E-74	SP: N	TM: N	Lb:GGGI01014746.1	1.35E-89	G1:GAIW01000405.1	5.36E-87	
GAJC01030581.1_43	Class 1	translocon-associated protein subunit beta	3.16E-99	SP: Y	TM: Y	Lb:GAJA01018646.1	1.46E-127	G1:GAIW01010586.1	8.88E-118	
GAJC01030396.1_14	Class 1	mitochondrial pyruvate carrier 1	1.19E-47	SP: N	TM: Y	Lb:GGGI01010001.1	1.65E-58	G1:GAIW01013369.1	3.42E-57	
GAJC01030130.1_52	Class 1	UDP-glucuronosyltransferase 1-9-like	0	SP: Y	TM: Y	Lb:GAJA01016209.1	0	G1:GAIW01022629.1	0	
GAJC01030096.1_24	Class 1	serine/threonine-protein phosphatase Pgam5,	9.03E-129	SP: N	TM: N	Lb:GAJA01020341.1	2.72E-156	G1:GAIW01017059.1	3.07E-146	
GAJC01029414.1_9	Class 1	MICOS complex subunit MIC27 isoform X2	6.18E-56	SP: N	TM: N	Lb:GAJA01006068.1	1.84E-107	G1:GAIW01015086.1	6.47E-95	
GAJC01029363.1_47	Class 1	NADH dehydrogenase [ubiquinone] iron-sulfur	6.23E-79	SP: N	TM: N	Lb:GGGI01013808.1	4.97E-122	G1:GAIW01012284.1	1.09E-101	

0	0	4.35E-153	7.18E-165	2.11E-97	3.25E-61	6.88E-108	0	2.77E-92	0	3.02E-114
G1: GAIW01010953.1	G1: GAIW01023990.1	G1: GAIW01009663.1	G1: GAIW01024386.1	G1: GAIW01021519.1	G1: GAIW01014286.1	G1: GAIW01016848.1	G1: GAIW01022553.1	G1: GAIW01012941.1	G1: GAIW01012493.1	G1: GAIW01001023.1
0	0	3.91E-160	0	1.02E-112	6.31E-54	8.26E-116	0	3.53E-108	0	2.33E-122
Lb: GAJA01021055.1	Lb: GGGI01009676.1	Lb: GAJA01006659.1	Lb: GAJA01018274.1	Lb: GGGI01004311.1	Lb: GGGI01011192.1	Lb: GAJA01002265.1	Lb: GGGI01008383.1	Lb: GAJA01021073.1	Lb: GGGI01013268.1	Lb: GAJA01002259.1
TM: N	TM: Y	TM: N	TM: N	TM: Y	TM: N	TM: Y	TM: Y	TM: N	TM: N	TM: N
SP: N	SP: N	SP: N	SP: N	SP: N	SP: Y	SP: N	SP: N	SP: N	SP: N	SP: N
0	0	1.65E-155	0	5.31E-53	2.24E-37	5.01E-113	0	5.36E-65	0	1.69E-93
2-oxoglutarate dehydrogenase, mitochon	ATP-binding cassette sub-family D member 3	V-type proton ATPase subunit D	dihydrolipoyllysine-residue succinyltransferase compone	NADH dehydrogenase [ubiquinone] 1 beta subcompl	50S ribosome-binding GTPase family protein	transmembrane and coiled-coil domain-containing protein	dolichyl-diphosphooligosaccharide--protein glycosyltr	NADH dehydrogenase [ubiquinone] 1 beta subcompl	V-type proton ATPase subunit d	putative OPA3-like protein CG13603
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01029325.1_93	GAJC01029192.1_25	GAJC01028989.1_23	GAJC01028891.1_55	GAJC01028693.1_7	GAJC01028672.1_28	GAJC01028414.1_19	GAJC01028343.1_20	GAJC01028170.1_18	GAJC01028123.1_9	GAJC01027933.1_33

0	0	0	7.91E-143	8.42E-81	3.54E-132	6.02E-57	0	8.29E-85	4.31E-180	4.86E-171
G1: GAIW01011910.1	G1: GAIW01015768.1	G1: GAIW01009654.1	G1: GAIW01023671.1	G1: GAIW01004976.1	G1: GAIW01004263.1	G1: GAIW01008755.1	G1: GAIW01023592.1	G1: GAIW01019874.1	G1: GAIW01012694.1	G1: GAIW01006128.1
0	0	0	4.15E-145	5.59E-118	1.07E-150	1.17E-65	0	1.59E-86	0	0
Lb: GGGI01013012.1	Lb: GGGI01009154.1	Lb: GGGI01012102.1	Lb: GGGI01005551.1	Lb: GGGI01000557.1	Lb: GGGI01011192.1	Lb: GGGI01001485.1	Lb: GGGI01008067.1	Lb: GAJA01012944.1	Lb: GAJA01012361.1	Lb: GAJA01008204.1
TM: N	TM: N	TM: N	TM: Y	TM: N	TM: N	TM: N	TM: N	TM: N	TM: Y	TM: Y
SP: N	SP: N	SP: N	SP: N	SP: Y	SP: Y	SP: Y	SP: Y	SP: Y	SP: N	SP: Y
0	0	0	6.22E-145	1.26E-82	1.91E-96	2.24E-43	0	2.15E-78	6.49E-178	3.89E-155
isocitrate dehydrogenase [NAD] subunit beta,	NADH dehydrogenase [ubiquinone] flavoprote	electron transfer flavoprotein subunit alpha, mitochon	vesicle-trafficking protein SEC22b-B	lipase 3	50S ribosome-binding GTPase	ribonuclease Oy	protein disulfide-isomerase A6	lysosomal alpha-mannosidase isoform X2	Trimeric intracellular cation channel type B	thioredoxin domain-containing protein
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01027925.1_11	GAJC01027894.1_58	GAJC01027824.1_14	GAJC01027799.1_61	GAJC01027753.1_25	GAJC01027746.1_33	GAJC01027725.1_6	GAJC01027531.1_60	GAJC01024852.1_14	GAJC01024725.1_36	GAJC01022480.1_6

0	0	0	1.12E-129	N/A	N/A	N/A	N/A	N/A	N/A	N/A
G1: GAIW01006852.1	G1: GAIW01010630.1	G1: GAIW01023112.1	G1: GAIW01019098.1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
0	0	0	6.92E-139	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Lb: GGGI01015284.1	Lb: GGGI01008526.1	Lb: GAJA01000548.1	Lb: GAJA01011987.1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: Y	TM: N	TM: N	TM: Y	TM: Y	TM: Y	TM: N	TM: N	TM: N	TM: N	TM: N
SP: N	SP: Y	SP: N	SP: Y	SP: N	SP: N	SP: Y	SP: Y	SP: Y	SP: Y	SP: N
0	0	0	9.24E-[illegible]	0	9.19E-05	8.59E-[illegible]	0	2.74E-75	1.42E-43	1.52E-[illegible]
sideroflexin-1	trehalase-like	paramyosin, long form-like	malectin-A	sideroflexin-2	angiopoietin-1 receptor isoform X3	major royal jelly protein 1	sarcalumenin isoform X2	lipase 3-like	endocuticle structural protein SgAbd-6-like	60S ribosomal protein L13
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01022382.1_37	GAJC01021538.1_51	GAJC01021129.1_25	GAJC01021107.1_29	GAJC01020023.1_41	GAJC01017999.1_17	GAJC01017428.1_36	GAJC01017322.1_30	GAJC01017199.1_29	GAJC01017061.1_11	GAJC01016301.1_14

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: N	TM: Y	TM: N	TM: N	TM: Y	TM: Y	TM: N	TM: Y	TM: N	TM: N	TM: N
SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N
0	5.10E-160	1.35E-126	0	0	2.13E-82	4.14E-120	4.42E-174	1.68E-107	5.59E-34	1.92E-65
60S ribosomal protein L4	NADH-cytochrome b5 reductase 2 isoform X1	cytochrome b-c1 complex subunit Rieske, mitochon	60S acidic ribosomal protein P0	cytochrome P450 4g15-like	membrane-associated progesterone	60S ribosomal protein L13a	cytochrome c1, heme protein, mitochondrial	NADH-quinone oxidoreductase subunit B 2-like	NADH dehydrogenase [ubiquinone] 1 alpha	fatty acid-binding protein, muscle isoform X1
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01013832.1_13	GAJC01013802.1_37	GAJC01013770.1_23	GAJC01013702.1_11	GAJC01013642.1_14	GAJC01013595.1_37	GAJC01013539.1_20	GAJC01013495.1_10	GAJC01013427.1_7	GAJC01013335.1_5	GAJC01013333.1_21

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: N	TM: Y	TM: N	TM: N	TM: N	TM: N	TM: Y	TM: N	TM: N	TM: N	TM: N
SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: Y	SP: N	SP: N	SP: N	SP: N
7.74E-178	1.86E-92	1.02E-94	1.92E-48	0	0	1.92E-120	0	1.12E-68	6.89E-58	2.97E-122
protein NipSnap	ATP synthase subunit b, mitochondrial	60S ribosomal protein L23	NADH dehydrogenase [ubiquinone] 1 alpha	T-complex protein 1 subunit zeta	60S ribosomal protein L5	transloco n-associated protein subunit alpha	Proline dehydrogenase 1, mitochondrial	myosin regulatory light chain 2	ATP synthase subunit d, mitochondrial	RNA-binding protein squid-like
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01013300.1_83	GAJC01013291.1_59	GAJC01013205.1_12	GAJC01013187.1_17	GAJC01013185.1_14	GAJC01013182.1_24	GAJC01013160.1_30	GAJC01013157.1_73	GAJC01013138.1_9	GAJC01013096.1_16	GAJC01012976.1_11

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: N	TM: N	TM: Y	TM: N	TM: Y	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N
SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: Y	SP: N	SP: Y
8.72E-100	0	3.96E-55	0	3.97E-25	0	5.30E-39	1.48E-180	3.30E-39	0	3.70E-29
RNA-binding protein squid isoform X2	heat shock 70 kDa protein cognate 4	cytochrome b5	glutamine synthetase 2 cytoplasmic isoform X1	NADH dehydrogenase [ubiquinone] 1 beta subcompl	annexin B9	ATP synthase subunit g, mitochondrial	14-3-3 protein zeta isoform X1	50S ribosome-binding GTPase	pyruvate kinase-like isoform X2	50S ribosome-binding GTPase
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01012952.1_13	GAJC01012897.1_26	GAJC01012804.1_6	GAJC01012771.1_15	GAJC01012769.1_5	GAJC01012766.1_49	GAJC01012746.1_16	GAJC01012689.1_23	GAJC01012610.1_15	GAJC01012565.1_23	GAJC01012525.1_26

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: N	TM: N	TM: N	TM: N	TM: Y	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N
SP: Y	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N
1.03E-32	1.80E-100	0	0	1.24E-24	2.36E-118	0	7.29E-180	1.04E-50	0	2.17E-96
50S ribosome-binding GTPase	ER membrane protein complex subunit 8/9	actin-5C	actin, muscle	cytochrome c oxidase subunit NDUFA4	60S ribosomal protein L11	very long-chain-fatty-acid--CoA ligase bubblegu	malate dehydrogenase, cytoplasmic	60S ribosomal protein L14	very-long-chain 3-oxoacyl-CoA reductase	60S ribosomal protein L10a
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01012524.1_31	GAJC01012472.1_23	GAJC01012412.1_24	GAJC01012411.1_49	GAJC01012253.1_8	GAJC01012249.1_18	GAJC01012240.1_36	GAJC01012232.1_17	GAJC01012230.1_7	GAJC01012219.1_12	GAJC01012216.1_25

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: N	TM: Y	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N
SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N
0	7.83E-57	2.44E-91	1.00E-85	2.38E-73	2.01E-64	6.21E-80	4.02E-97	8.94E-36	0	0
spectrin alpha chain isoform X1	NADH dehydrogenase [ubiquinone] 1 alpha	40S ribosomal protein S15Aa	myosin light chain alkali isoform X2	probable NADH dehydrogenase [ubiquinone] 1	farnesol dehydrogenase-like	farnesol dehydrogenase-like	40S ribosomal protein S10	cytochrome b-c1 complex subunit 7-like	mitochondrial 2-oxoglutarate/malate carrier protein-	putative ATP-dependent RNA helicase me31b
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01012184.1_54	GAJC01012155.1_9	GAJC01012150.1_20	GAJC01012137.1_24	GAJC01012129.1_15	GAJC01012103.1_17	GAJC01012102.1_28	GAJC01012069.1_15	GAJC01012068.1_6	GAJC01011949.1_55	GAJC01011927.1_54

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: Y	TM: N	TM: Y	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N
SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: Y	SP: N	SP: N	SP: N	SP: Y
2.76E-51	0	6.70E-102	6.27E-66	3.22E-146	5.73E-127	1.54E-11	0	0	9.76E-110	7.70E-27
cytochrome b5-like	Protein disulfide-isomerase	protein RER1	myosin regulatory light chain 2	40S ribosomal protein S2	3-hydroxyacyl-CoA dehydrogenase type-2	leucine-rich repeat, immunoglobulin-like domain	dihydrolipoyl dehydrogenase, mitochondrial	NADH-ubiquinone oxidoreductase 49 kDa	40S ribosomal protein S7	kielin/chordin-like protein
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01011893.1_31	GAJC01011807.1_10	GAJC01011782.1_10	GAJC01011653.1_5	GAJC01011619.1_22	GAJC01011605.1_14	GAJC01011586.1_19	GAJC01011558.1_41	GAJC01011550.1_14	GAJC01011454.1_14	GAJC01011419.1_7

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: N	TM: Y	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N
SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: Y	SP: N	SP: N
0	2.31E-71	4.57E-103	0	1.15E-154	8.42E-52	1.08E-143	4.30E-143	0	0	4.17E-73
MICOS complex subunit Mic60 isoform X1	microsomal glutathione S-transferase 1-like	peroxiredoxin-5, mitochondrial	40S ribosomal protein S4	60S ribosomal protein L10	MICOS complex subunit MIC19	peroxiredoxin-6-like	NADH dehydrogenase [ubiquinone] flavoprote	calreticulin	mitochondrial import inner membrane	60S ribosomal protein L36
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01011326.1_68	GAJC01011277.1_17	GAJC01011256.1_29	GAJC01011252.1_23	GAJC01011236.1_30	GAJC01011232.1_6	GAJC01011192.1_150	GAJC01011169.1_8	GAJC01011133.1_10	GAJC01011105.1_9	GAJC01011050.1_5

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: Y
SP: N	SP: N	SP: N	SP: N	SP: Y	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N
4.32E-126	0	3.18E-57	2.95E-97	4.39E-33	1.91E-79	2.76E-27	3.71E-122	5.10E-78	6.08E-42	6.38E-102
40S ribosomal protein S8	tubulin beta-1 chain	single-stranded DNA-binding protein, mitochon	troponin C	50S ribosome-binding GTPase	plasminogen activator inhibitor 1 RNA-binding	NADH dehydrogenase [ubiquinone] 1 beta subcompl	60S ribosomal protein L9	60S ribosomal protein L35a	ATP synthase-coupling factor 6, mitochon drial	very-long-chain (3R)-3-hydroxyacyl-CoA dehydrata
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01011037.1_17	GAJC01011036.1_41	GAJC01011006.1_24	GAJC01010984.1_6	GAJC01010915.1_9	GAJC01010865.1_16	GAJC01010822.1_3	GAJC01010807.1_12	GAJC01010760.1_13	GAJC01010666.1_10	GAJC01010573.1_13

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: Y	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N
SP: N	SP: N	SP: N	SP: Y	SP: N	SP: N	SP: N	SP: N	SP: Y	SP: N	SP: N
0	0	7.75E-50	1.34E-107	5.67E-78	0	0	1.94E-75	1.82E-38	3.34E-32	0
protein bark beetle isoform X1	moesin/ezrin/radixin homolog 1 isoform X2	cytochrome c oxidase subunit 5B, mitochon	spermine oxidase-like	40S ribosomal protein S20	glyceraldehyde-3-phosphate dehydrogenase 2	elongation factor 1-alpha	CDGSH iron-sulfur domain-containing protein 2	50S ribosome-binding GTPase family protein	cytochrome b-c1 complex subunit 8	myosin heavy chain, muscle isoform X2
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01010572.1_196	GAJC01010488.1_72	GAJC01010481.1_3	GAJC01010465.1_24	GAJC01010390.1_9	GAJC01010319.1_30	GAJC01010280.1_48	GAJC01010278.1_4	GAJC01010274.1_32	GAJC01010260.1_26	GAJC01010230.1_54

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: N	TM: N	TM: N	TM: N	TM: N	TM: Y	TM: N	TM: Y	TM: N	TM: N	TM: Y
SP: N	SP: N	SP: N	SP: Y	SP: Y	SP: N	SP: N	SP: Y	SP: N	SP: Y	SP: N
0	1.63E-168	4.73E-17	6.90E-28	2.39E-27	1.52E-100	2.40E-66	4.32E-101	1.76E-55	6.59E-81	5.34E-59
pyruvate carboxylase, mitochondrial isoform	60S ribosomal protein L8	ATP synthase subunit e, mitochondrial	protein NPC2 homolog	icarapin-like	B-cell receptor-associated protein 31	histone H2A.V	transmembrane emp24 domain-containing protein	40S ribosomal protein S27	transmembrane protein 214-A	60S ribosomal protein L24
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01010173.1_113	GAJC01010142.1_8	GAJC01010139.1_45	GAJC01010058.1_13	GAJC01009987.1_9	GAJC01009972.1_32	GAJC01009958.1_42	GAJC01009924.1_35	GAJC01009895.1_5	GAJC01009838.1_20	GAJC01009770.1_24

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: N	TM: N	TM: N	TM: Y	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N
SP: Y	SP: N	SP: Y	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N
1.12E-07	0	5.64E-50	0	1.84E-64	2.67E-53	1.37E-153	1.37E-69	1.68E-103	8.55E-46	2.62E-53
apolipophorin-III-like protein	Tubulin alpha-1 chain	general odorant-binding protein 56d-like	acyl-CoA Delta(11) desaturase-like	Y-box factor homolog isoform X1	NADH dehydrogenase [ubiquinone] 1 beta subcompl	adenylate kinase	cytochrome c oxidase subunit 5A, mitochon	PDZ and LIM domain protein 3 isoform X2	muscle LIM protein 1 isoform X3	NADH dehydrogenase [ubiquinone] iron-sulfur
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01009754.1_11	GAJC01009730.1_31	GAJC01009718.1_4	GAJC01009624.1_17	GAJC01009588.1_5	GAJC01009568.1_14	GAJC01009563.1_35	GAJC01009552.1_6	GAJC01009529.1_21	GAJC01009527.1_6	GAJC01009482.1_8

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N
SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: Y	SP: Y	SP: N
0	1.02E-37	7.36E-60	0	0	1.11E-121	7.28E-122	6.09E-60	8.08E-09	1.34E-109	7.55E-96
succinate dehydrogenase [ubiquinone] iron-sulfur	protein lethal(2)essential for life	60S ribosomal protein L22-like	asparagine synthetase [glutamine-	alcohol dehydrogenase class-3	muscle-specific protein 20	60S ribosomal protein L17	ubiquitin-like protein FUBI	sushi, von Willebrand factor type A, EGF and pentraxin	endoplasmin	40S ribosomal protein S16
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1
GAJC01009424.1_7	GAJC01009366.1_63	GAJC01009365.1_14	GAJC01009336.1_20	GAJC01009315.1_44	GAJC01009301.1_5	GAJC01009290.1_24	GAJC01009288.1_7	GAJC01009225.1_18	GAJC01009190.1_6	GAJC01009175.1_28

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	6.76E-41
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	G1: GAIW01006143.1
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	8.78E-71
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Lb: GGGI01003289.1
TM: Y	TM: N	TM: N	TM: N	TM: Y	TM: N	TM: N	TM: N	TM: Y	TM: N	TM: N
SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: Y
1.75E-62	1.60E-113	3.11E-130	1.78E-69	0	2.06E-50	1.02E-125	0	4.39E-15	5.29E-50	2.50E-47
mitochondrial import inner membrane	ATP synthase subunit O, mitochondrial	NADH dehydrogenase [ubiquinone] iron-sulfur	40S ribosomal protein S15	sodium/potassium-transporting ATPase subunit	histone H4-like	60S ribosomal protein L12	electron transfer flavoprotein-ubiquinone	UPF0389 protein CG9231	neural/ectodermal development factor IMP-L2 isoform	A disintegrin and metalloproteinase with
Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 1	Class 2
GAJC01009175.1_13	GAJC01009149.1_50	GAJC01009149.1_29	GAJC01008186.1_8	GAJC01007074.1_63	GAJC01003958.1_11	GAJC01003331.1_6	GAJC01002392.1_58	GAJC01002381.1_6	GAJC01001939.1_7	GAJC01029468.1_54

2.04E-66	N/A	5.05E-59	4.80E-11	4.31E-24	1.93E-15	N/A	N/A	N/A	N/A	N/A
G1: GAIW010 26852.1	N/A	G1: GAIW010 14286.1	G1: GAIW010 09606.1	G1: GAIW010 09100.1	G1: GAIW010 08281.1	N/A	N/A	N/A	N/A	N/A
5.74E-157	2.52E-06	5.63E-51	1.16E-26	3.25E-53	4.28E-16	N/A	N/A	N/A	N/A	N/A
Lb: GAJA010 19749.1	Lb: GGGI010 16314.1	Lb: GGGI010 11192.1	Lb: GAJA010 13720.1	Lb: GAJA010 13720.1	Lb: GAJA010 08052.1	N/A	N/A	N/A	N/A	N/A
TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N
SP: Y	SP: N	SP: Y	SP: N	SP: N	SP: Y	SP: N	SP: Y	SP: Y	SP: Y	SP: Y
6.02E-05	2.90E-16	2.82E-44	4.80E-18	2.30E-33	3.53E-19	1.01E-28	7.56E-06	9.01E-16	0	1.77E-07
enhancing factor (viral)	rho GTPase-activating protein 9 isoform X1	ATP-binding protein	neprilysin isoform X1	endothelin-converting enzyme 1 isoform X1	membrane metallo-endopeptidase-like 1 isoform X1	GTPase RsgA	Endothelin-converting enzyme 1	endothelin-converting enzyme 1 isoform X1	lysosomal aspartic protease	protein Diedel-like
Class 2	Class 2	Class 2	Class 2	Class 2	Class 2	Class 2	Class 2	Class 2	Class 2	Class 2
GAJC010 28852.1_18	GAJC010 27751.1_9	GAJC010 21576.1_36	GAJC010 20722.1_12	GAJC010 20720.1_15	GAJC010 20718.1_27	GAJC010 18250.1_20	GAJC010 18015.1_12	GAJC010 13610.1_18	GAJC010 12383.1_11	GAJC010 10415.1_14

N/A	N/A	2.68E-11	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	G1: GAIW01000379.1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	9.42E-28	N/A	N/A	N/A	3.13E-11	4.11E-14	N/A	N/A	N/A
N/A	N/A	Lb: GGGI01000204.1	N/A	N/A	N/A	Lb: GAJA01012767.1	Lb: GGGI01000916.1	N/A	N/A	N/A
TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N
SP: N	SP: Y	SP: Y	SP: N	SP: N	SP: Y	SP: N	SP: N	SP: N	SP: Y	SP: Y
8.51E-150	N/A	3.41E-20	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ras-related protein Rab-11A	---NA---	Macrophage receptor MARCO	---NA---	---NA---	---NA---	---NA---	---NA---	---NA---	---NA---	---NA---
Class 2	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3
GAJC01010126.1_52	GAJC01030678.1_22	GAJC01029385.1_87	GAJC01029271.1_8	GAJC01028966.1_1	GAJC01027902.1_6	GAJC01027679.1_5	GAJC01027678.1_12	GAJC01027452.1_12	GAJC01027372.1_5	GAJC01027336.1_3

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: N	TM: N	TM: N	TM: N	TM: N	TM: Y	TM: N	TM: N	TM: Y	TM: N	TM: Y
SP: Y	SP: N	SP: N	SP: Y	SP: Y	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1.24E-50
---NA---	---NA---	---NA---	---NA---	---NA---	---NA---	---NA---	---NA---	---NA---	---NA---	PREDICTED: uncharacterized protein LOC1071
Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3
GAJC01024744.1_6	GAJC01024743.1_12	GAJC01023574.1_5	GAJC01021659.1_29	GAJC01021632.1_11	GAJC01017940.1_12	GAJC01017287.1_25	GAJC01017051.1_6	GAJC01013355.1_10	GAJC01012558.1_12	GAJC01012408.1_7

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: N	TM: N	TM: N	TM: N	TM: Y	TM: N	TM: N	TM: Y	TM: N	TM: N	TM: N
SP: Y	SP: Y	SP: Y	SP: Y	SP: N	SP: Y	SP: Y	SP: N	SP: N	SP: Y	SP: N
N/A	N/A	N/A	N/A	1.01E-24	N/A	N/A	N/A	1.14E-08	N/A	4.18E-76
---NA---	---NA---	---NA---	---NA---	signal recognition particle receptor subunit alpha-like	---NA---	---NA---	---NA---	hypothetical protein TSAR_010554	---NA---	hypothetical protein SINV_10737, partial
Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3
GAJC01011873.1_7	GAJC01011863.1_13	GAJC01011465.1_5	GAJC01011463.1_48	GAJC01011457.1_12	GAJC01011137.1_32	GAJC01011097.1_122	GAJC01010930.1_16	GAJC01010444.1_106	GAJC01010353.1_14	GAJC01010250.1_8

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: Y	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N	TM: N
SP: N	SP: N	SP: Y	SP: Y	SP: Y	SP: Y	SP: N	SP: Y	SP: N	SP: N	SP: N
N/A	N/A	N/A	2.19E-11	N/A	N/A	N/A	N/A	N/A	N/A	N/A
---NA---	---NA---	---NA---	uncharacterized protein LOC107274787	---NA---	---NA---	---NA---	---NA---	---NA---	---NA---	---NA---
Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3
GAJC01010050.1_21	GAJC01009713.1_25	GAJC01009493.1_4	GAJC01009363.1_49	GAJC01009309.1_26	GAJC01007489.1_11	GAJC01007380.1_16	GAJC01006038.1_15	GAJC01005406.1_20	GAJC01003173.1_5	GAJC01002124.1_43

N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TM: N	TM: N	TM: N	TM: N	TM: N	TM: Y	TM: N	TM: N
SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N	SP: N
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	---NA---	---NA---	---NA---	---NA---	---NA---	---NA---	---NA---
Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3	Class 3
GAJC01001115.1_93	GAJC01000654.1_10	GAJC01000653.1_18	GAJC01000527.1_1	GAJC01000524.1_2	GAJC01000367.1_14	GAJC01000363.1_6	GAJC01000213.1_10

11

12

Table S2: Lh 14 MSEV Proteins Are Not Highly Related to Viral Proteins. Lh MSEV superset proteins were searched for similarities to LbFV proteins via BLASTp. They were also searched against the entire PDV nr database and Viridae nr database via BLASTp. Results from an unrestricted nr BLASTp search score better than any alignment obtained from LbFV, PDV nr, and Viridae nr.

TABLE S2: *Lh 14* MSEV Proteins Are Not Highly Related to Viral Proteins						
LbFV (3 results)						
In-House Query Accesssion		Subject	Description	%ID	E-value	Quer Cov
GAJC01010619.1_3	LbFV	AQQ80003.1	hypothetical protein LbFV_ORF83 [Leptopilina boulardi filamentous virus]	27	0.001	39
	nr	XP_026479782.1	uncharacterized protein LOC113386201 [Ctenocephalides felis]	25.93	2.00E-04	72
GAJC01009336.1_20	LbFV	AQQ79980.1	putative lecithin:cholesterol acyltransferase [Leptopilina boulardi filamentous virus]	25	9.00E-04	39
	nr	XP_015108849.1	PREDICTED: asparagine synthetase [glutamine-hydrolyzing]-like [Diachasma alloeum]	69.12	0	99
GAJC01011893.1_31	LbFV	AQQ80007.1	hypothetical protein LbFV_ORF87 [Leptopilina boulardi filamentous virus]	31.25	3.00E-05	43
	nr	KMQ95331.1	cytochrome b5 [Lasius niger]	63.57	2.00E-53	92
PDV (4 results)						
In-House Query Accesssion	Database	Subject	Description	%ID	E-value	Quer Cov
GAJC01027725.1_6	nr:PDV	YP_184862.1	hypothetical protein CcBV_23.4 [Cotesia congregata bracovirus]	31.088	1.38E-17	77
	nr	XP_004958409.1	ribonuclease 1 [Setaria italica]	26.068	1.14E-16	84
GAJC01013642.1_14	nr:PDV	AFN42311.1	cytochrome P450 [Cotesia sesamiae Mombasa bracovirus]	27.053	3.50E-37	72
	nr	NP_001165993.1	cytochrome P450 4G44 [Nasonia vitripennis]	71.3	0	99
GAJC01010919.1_17	nr:PDV	AFN42313.1	cytochrome P450 [Cotesia sesamiae Mombasa bracovirus]	22.558	1.23E-31	93

	nr	XP_012279668.1	cytochrome P450 4g15 [Orussus abietinus]	84.956	0	99
GAJC01003958.1_11	nr: PDV	AAV98010.1	hypothetical protein ORF3006 [Cotesia plutellae polydnavirus]	84.444	4.70E-52	89
	nr	OWK59279.1	Histone H4 [Lonchura striata domestica]	99.01	2.89E-63	100
Viridae (35 results)						
In-House Query Accesssion	Database	Subject	Description	%ID	E-value	Query Cov
GAJC01013393.1_42	nr: Viridae	ARF08651.1	heat shock protein 83 kDa [Catovirus CTV1]	36.111	1.88E-65	88
	nr	KMQ97651.1	endoplasmin-like protein [Lasius niger]	77.473	0	100
GAJC01009190.1_6	nr: Viridae	AYV76070.1	hsp82-like protein [Terrestrivirus sp.]	58.871	1.14E-39	62
	nr	XP_015592342.1	endoplasmin [Cephus cinctus]	84.848	6.19E-109	100
GAJC01010930.1_14	nr: Viridae	YP_009213665.1	putative fructose-1,6-bisphosphate aldolase [Prochlorococcus phage P-TIM68]	52.557	1.44E-119	96
	nr	XP_011139291.1	fructose-bisphosphate aldolase isoform X1 [Harpegnathos saltator]	88.219	0	100
GAJC01012897.1_26	nr: Viridae	YP_009173627.1	Hsp70 protein [Chrysochromulina ericina virus]	70.936	0	92
	nr	XP_017757713.1	PREDICTED: heat shock 70 kDa protein cognate 4 isoform X1 [Eufriesea mexicana]	91.818	0	100
GAJC01012804.1_6	nr: Viridae	AYV75195.1	putative cytochrome b5 [Terrestrivirus sp.]	43.75	5.20E-22	56
	nr	XP_015509615.1	PREDICTED: cytochrome b5 [Neodiprion lecontei]	69.925	6.49E-64	94
GAJC01010280.1_48	nr: Viridae	AYV78014.1	elongation factor 1 alpha long form [Edafosvirus sp.]	43.29	1.53E-122	96
	nr	NP_001166227.1	elongation factor 1-alpha [Nasonia vitripennis]	97.609	0	100
GAJC01017322.1_30	nr: Viridae	AYV76711.1	hypothetical protein Terrestrivirus12_14 [Terrestrivirus sp.]	31.078	9.76E-50	46

	nr	XP_011330618.1	sarcalumenin isoform X2 [Ooceraea biroi]	63.412	0	99
GAJC01009288.1_7	nr: Viridae	NP_598374.1	ubiquitin-like protein [Murine osteosarcoma virus]	52.273	1.07E-30	98
	nr	OXU26385.1	hypothetical protein TSAR_013419 [Trichomalopsis sarcophagae]	79.389	3.38E-69	100
GAJC01011256.1_29	nr: Viridae	YP_007518064.1	peroxiredoxin [Pelagibacter phage HTVC008M]	46.903	9.39E-29	60
	nr	XP_015601491.1	peroxiredoxin-5, mitochondrial [Cephus cinctus]	77.901	1.78E-102	97
GAJC01011949.1_55	nr: Viridae	QBK91775.1	mitochondrial carrier-like protein [Pithovirus LCPAC304]	33.103	4.15E-44	95
	nr	XP_014220960.1	mitochondrial 2-oxoglutarate/malate carrier protein-like [Trichogramma pretiosum]	90.278	0	98
GAJC01013642.1_14	nr: Viridae	AFN42311.1	cytochrome P450 [Cotesia sesamiae Mombasa bracovirus]	27.053	9.08E-34	72
	nr	OXU28285.1	hypothetical protein TSAR_003383 [Trichomalopsis sarcophagae]	71.3	0	99
GAJC01013073.1_33	nr: Viridae	1604248A	J1 Protein (Moloney murine leukemia virus)	76.923	0	97
	nr	XP_017888099.1	60S ribosomal protein L3 [Ceratina calcarata]	93.689	0	100
GAJC01009958.1_42	nr: Viridae	QBK91702.1	histone 2A-domain-containing protein [Pithovirus LCPAC304]	45.631	1.15E-24	81
	nr	KZC06774.1	Histone H2A.V, partial [Dufourea novaeangliae]	100	5.40E-80	100
GAJC01029718.1_61	nr: Viridae	AYV78414.1	putative ras-related protein Rab-10-like protein [Edafosvirus sp.]	48.37	1.05E-63	81
	nr	XP_015123539.1	PREDICTED: ras-related protein Rab-3 isoform X2 [Diachasma alloeum]	94.57	7.07E-155	100
GAJC01011832.1_33	nr: Viridae	AYV75487.1	GTP-binding protein YPTC1 [Terrestrivirus sp.]	67.647	5.91E-80	80
	nr	XP_001607208.1	PREDICTED: ras-related protein Rab-1A [Nasonia vitripennis]	90.952	1.87E-143	100
GAJC01009336.1_20	nr: Viridae	AYV84840.1	asparagine synthase (glutamine-hydrolysing) [Hyperionvirus sp.]	42.087	9.61E-139	99

	nr	XP_015108849.1	PREDICTED: asparagine synthetase [glutamine-hydrolyzing]-like [Diachasma alloeum]	69.12	0	99
GAJC01009776.1_111	nr: Viridae	AGE57609.1	calcium-transporting ATPase, plasma membrane-type [Acanthocystis turfacea Chlorella virus NTS-1]	29.114	1.37E-81	83
	nr	XP_018058981.1	PREDICTED: sodium/potassium-transporting ATPase subunit alpha isoform X5 [Atta colombica]	96.226	0	100
GAJC01009624.1_17	nr: Viridae	YP_009221129.1	hypothetical protein AV955_gp109 [Diadromus pulchellus ascovirus 4a]	51.579	3.30E-104	79
	nr	XP_011637428.1	acyl-CoA Delta(11) desaturase-like [Pogonomyrmex barbatus]	76.353	0	99
GAJC01028364.1_66	nr: Viridae	YP_004061441.1	hypothetical protein BpV1_011c [Bathycoccus sp. RCC1105 virus BpV1].	41.412	1.33E-119	63
	nr	XP_014470355.1	PREDICTED: AFG3-like protein 2 [Dinoponera quadriceps]	68.642	0	99
GAJC01010919.1_17	nr: Viridae	AFN42313.1	cytochrome P450 [Cotesia sesamiae Mombasa bracovirus]	22.558	2.22E-28	94
	nr	XP_012279668.1	cytochrome P450 4g15 [Orussus abietinus]	84.956	0	100
GAJC01012692.1_68	nr: Viridae	QBK87831.1	Hsp70 protein [Marseillevirus LCMAC202]	63.366	0	92
	nr	XP_015112461.1	PREDICTED: heat shock 70 kDa protein cognate 3 [Diachasma alloeum]	92.683	0	100
GAJC01012771.1_15	nr: Viridae	AYV80407.1	glutamine synthetase [Harvfovirus sp.]	52.632	4.60E-116	90
	nr	XP_008558400.1	PREDICTED: glutamine synthetase 2 cytoplasmic isoform X1 [Microplitis demolitor]	86.957	0	100
GAJC01012526.1_17	nr: Viridae	AYV76018.1	processing protease [Terrestrivirus sp.]	26.699	9.12E-36	79
	nr	XP_011172112.1	mitochondrial-processing peptidase subunit beta [Solenopsis invicta]	76.891	0	100
GAJC01011927.1_54	nr: Viridae	QBK89261.1	DEAD/SNF2-like helicase [Mimivirus LCMiAC02]	39.594	1.05E-87	82
	nr	XP_011501061.1	PREDICTED: putative ATP-dependent RNA helicase me31b [Ceratosolen solmsi marchali]	95.034	0	100
GAJC01012412.1_24	nr: Viridae	CAA25063.1	unnamed protein product [Feline sarcoma virus]	64.706	8.52E-87	58

	nr	QBB01857.1	beta-actin, partial [Cotesia chilonis]	100	0	100
GAJC01009315.1_44	nr: Viridae	YP_009213701.1	putative zinc-containing alcohol dehydrogenase superfamily protein [Prochlorococcus phage P-TIM68]	28.184	1.22E-31	96
	nr	XP_015589544.1	alcohol dehydrogenase class-3 [Cephus cinctus]	84.043	0	100
GAJC01011558.1_41	nr: Viridae	ASD51141.1	putative dihydrolipoamide dehydrogenase [Erysipelothrix phage phi1605]	27.826	5.52E-51	89
	nr	XP_015521697.1	PREDICTED: dihydrolipoyl dehydrogenase, mitochondrial [Neodiprion lecontei]	87.179	0	100
GAJC01010126.1_52	nr: Viridae	AYV82031.1	Ras-like protein Rab-11A [Homavirus sp.]	62.679	7.74E-91	96
	nr	XP_014600829.1	PREDICTED: ras-related protein Rab-11A [Polistes canadensis]	100	2.01E-158	100
GAJC01013185.1_14	nr: Viridae	AQM32732.1	thermosome, partial [uncultured virus]	27.866	2.37E-48	95
	nr	XP_003488200.1	T-complex protein 1 subunit zeta [Bombus impatiens]	85.283	0	100
GAJC01012411.1_49	nr: Viridae	CAA25063.1	unnamed protein product [Feline sarcoma virus]	62.602	2.66E-85	62
	nr	XP_014603494.1	PREDICTED: actin, muscle [Polistes canadensis]	100	0	100
GAJC01012240.1_36	nr: Viridae	QBK88455.1	AMP-binding enzyme [Mimivirus LCMiAC01]	35.738	9.30E-111	84
	nr	XP_008544493.1	PREDICTED: very long-chain-fatty-acid--CoA ligase bubblegum-like isoform X1 [Microplitis demolitor]	73.547	0	100
GAJC01010424.1_32	nr: Viridae	AGE51234.1	calcium-transporting ATPase, plasma membrane-type [Paramecium bursaria Chlorella virus CVG-1]	29.126	2.47E-89	82
	nr	XP_012227425.1	PREDICTED: calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type isoform X2 [Linepithema humile]	94.79	0	99
GAJC01027725.1_6	nr: Viridae	AUL79063.1	T2 family ribonuclease [Tupanvirus deep ocean]	34.021	8.42E-22	86
	nr	XP_012271266.1	ribonuclease Oy [Orussus abietinus]	36.638	2.69E-43	96
GAJC01030130.1_52	nr: Viridae	AXU41563.1	EGT [Mythimna unipuncta nucleopolyhedrovirus]	25.866	2.08E-53	90

	nr	XP_012215174.1	PREDICTED: UDP-glucuronosyltransferase-like [Linepithema humile]	51.145	0	98
GAJC01011881.1_45	nr: Viridae	AQM32696.1	chaperonin GroEL [uncultured virus]	52.7	1.06E-178	94
	nr	XP_011340338.1	heat shock protein 60A [Ooceraea biroi]	90.328	0	96

17

ADDENDUM TO CHAPTER 1

Bioinformatic Comparison of Two *Leptopilina* spp. Proteomes Reveals Similar Profiles

INTRODUCTION:

The diversification of the parasitoid wasp genus, *Leptopilina*, occurred more than 30 million years ago (*1, 2*). Four *Leptopilina* species (*Leptopilina heterotoma* (*Lh*), *Leptopilina boulardi* (*Lb*), *Leptopilina clavipes*, and *Leptopilina victoriae*) are highly studied as proposed pest control agents, the *Leptopilina* genus is of great interest to those studying evolutionary biology and immunology (*3-6*). This interest is due to differences in host immune suppression strategy and the ability to utilize the well-characterized *Drosophila* as a host species (*3-6*). Studies on the origin, nature, and functions of *Leptopilina* VLPs have led to comparative studies in host range (*3*), secreted venom proteins (*6*), transcriptomes (*7, 8*) and proteomes (*9, 10*). In these latter works, no direct comparison of *Lh* MSEVs and *Lb* VLP proteomes indicating whether the two particles share a profile supporting the hypothesis of extracellular vesicles or viruses was performed. Utilizing our own *Lh 14* MSEV proteomes (*8, 10*) and the newly published *Lb* strain Gotheron (*Lb G*) VLP proteome (*9*), we directly compared the two particles not only in terms of similar proteins, but also for virus-like proteins, enrichment profile, and for putative secreted and transmembrane proteins. We hypothesized that *Lb G* VLPs should share an overall similar profile to *Lh 14* MSEVs, possess properties of extracellular vesicles, and do not contain viral proteins. An enrichment profile of the *Lb G* VLPs proteome indicates an exosomal character. The proportions of proteins amongst previously published *Lh* protein classes was also similar even though only a third of all *Lb G* proteins (115/383) were highly similar to *Lh 14* proteins. A majority of the 115 proteins are conserved cellular proteins. Three of these conserved proteins possess a signal peptide, a transmembrane domain, and annotations for proteins associated with the endomembrane system, supporting an endosome membrane system origin for the *Lh* and *Lb* particle membranes. As new wasp genome assemblies have been released on GenBank (*11*), we re-examined the portion of expected particle proteins that can be found in both *Lh* and *Lb* genomes. We report that more than 95% of proteins have coding regions within both genomes, confirming our previous results in Chapter 1.

METHODS:

Insects: Isogenized *L. heterotoma* strain 14 (*Lh 14*) and *L. boulardi* strain 17 (*Lb 17*) (*3*), were raised on the *y w* strain of *D. melanogaster* that were reared on standard cornmeal, yeast, and agar fly food at 25°C as previously described (*12*). Adult wasps were collected from parasitized hosts, 25 days after infection at 25^{o}C. Male and female wasps were stored on fly food with 70% honey on "buzz" plugs.

Analysis of *L. boulardi* VLP proteins: Protein sequences were obtained from Dr. Julien Varaldi (University of Lyon) for *Lb* strain Gotheron (*Lb G*) VLPs (*9*) as these sequences were not available publicly. Proteins that had an equivalent transcript in the *Lb 17* whole body transcriptome (*Lb* Ground, GITC00000000, *Lb* Space, GISX00000000) and the female abdominal transcriptome (GAJA00000000) were identified and this corresponding accession number is noted in Table 1. Equivalence is defined as %ID, E-value, and query coverage being within cutoff threshold of BLAST search (%ID >70%, E-value < 1E-50, and query coverage > 70%). E-value indicates the estimated probability of finding a better match after repeated searches within the database. Query coverage (Quer Cov) indicates percentage of query covered by subject sequence.

Annotation: Proteins were annotated using the OmicsBox annotation pipeline (downloaded April 2020), E-value threshold of 1E-7 (*13, 14*). Results were organized and classified based on Gene Ontology (GO) terms from UniProt and InterPro using scripts to sort annotation results, followed by manual curation (*15-18*). Proteins that failed annotation by OmicsBox but obtained GO terms from UniProt/InterProScan searches and were still classified. Proteins were considered "virulence-related" based on GO terms indicating involvement with infection, inflammation, and immune response through the manual curation step. The above protocol applies to all proteins in both the *Lb* and *Lh* proteomes (and in all tables where a Class is specified in this Addendum). The results may contain minor discrepancies when compared to classification in Heavner et al.

Enrichment: Enrichment analysis (Fig 1C) of *Lb G* VLP and *Lh 14* MSEV proteins was as performed (*8, 10*) using FunRich and the Vesiclepedia database as of November 2020. Simultaneous analysis was done in order to ensure that both proteomes were analyzed with the same available version of the database (*19-21*). Briefly, *Lb G* VLP and *Lh 14* MSEV proteins could be included in the enrichment

analyses only if a human ortholog was present. The gene identifiers for human orthologs were obtained from the MSEV/VLP KO and the UniProt mapping utility (*17, 22*).

Secreted/membrane proteins: Presence of signal peptide and/or transmembrane domain in *Lb G* and the full set of *Lh 14* was examined utilizing SignalP, TMHMM, and Phobius (*23-28*) as in (*10*). Assignments for each parameter was made only if a specific sequence met the criteria of both programs (i.e., SignalP and Phobius for signal peptide, and TMHMM and Phobius for transmembrane domain). Full listing of *Lb G* proteins analyzed for signal peptide and/or transmembrane domain in Table 1. Full listing of *Lh 14* proteins analyzed for signal peptide and/or transmembrane domain in Table 2.

BLAST search using *Lb G* proteins: To determine correspondence of proteins within each proteome, the NCBI BLAST+ (v 2.7.1) (*29-31*) tool was used to query 383 *Lb* VLP proteins (*9*) against (a) NCBI non-redundant (nr) database restricted to Polydnaviridae (Taxid: 10482), Viridae (Taxid: 10239), and unrestricted (nr) as described previously in (*10*) (see Tables 7 and 8); and (b) also against the 407 *Lh 14* MSEV proteins (see Table 5, Figure 1A) (*8, 10*). Proteins were considered similar if E-value was within 1E-5 and query coverage >50-70% in both searches. Percent identity was not used as a threshold in comparing *Lh 14* to *Lb G* but was provided in the table.

BLAST search of new wasp genomes: New genome assemblies for *Lh 14* and *Lb* strain *G486* (*Lb G486*) were released on GenBank by Huang et al. under the accession numbers JABAIE010000000 and JABAIF010000000 respectively and downloaded in December 2020 and published in (*11*). Protein sequences for *Lh 14* MSEVs and *Lb G* VLPs were used as queries for TBLASTN search of genomes using NCBI BLAST+ (v2.7.1) (*29-31*). Cut-off for positive results was query coverage > 50%, %ID > 40% and E-value of 1E-7.

RESULTS AND DISCUSSION:

Similarities between *Lb* VLPs and *Lh* MSEVs

To test the extracellular vesicle hypothesis, we first annotated and sorted all 383 *Lb G* proteins into the same classes we have previously used for *Lh*: Class 1 – Conserved Eukaryotic, Class 2 – Infection and Immunity, and Class 3 – Unannotated Novel (*8, 10*). The full list of proteins with annotations, class, presence/absence of signal peptide and transmembrane domain, and equivalent accession numbers is in Table 1. A repeat of this analysis with a full listing of *Lh 14* proteins can be found in Table 2. A number of similarities became apparent from these analyses which are discussed below.

Comparison of classes

When comparing the 383 proteins found in *Lb G* VLPs to the set of 407 *Lh 14* MSEV proteins (previously described as the "superset" published in (*10*)), we found that only 115 proteins were highly similar between the two species at a stringent threshold of E-value 1E-5 and >70% query coverage (Fig 1A, Table 3). Of these 115 proteins, the vast majority (96/115, 83.5%) were classified as conserved eukaryotic (Class 1) proteins, which is to be expected as they are important for cellular functions in eukaryotes. The remaining 19 proteins were split into infection and immunity related (Class 2 - 10/115, 8.69%) and novel proteins (Class 3 - 9/115, 7.82%) (Fig 1A, Table 3, all proteins that were unique to *Lh 14* and *Lb* are listed in Tables 4 and 5 respectively).

Although only an approximate third of the *Lb G* VLP proteins were found to be highly similar to the *Lh 14* MSEV proteins, we examined if the overall proportions of these proteins in the three classes would also be the same. Compared to *Lh 14* (Class 1: 63.8%, Class 2: 11.1%, Class 3: 24.6% (*10*)), *Lb* VLP proteins are similarly distributed (Class 1: 249/383 - 65%, Class 2: 34/383 - 9%, Class 3: 100/383 - 26%) (Fig 1B). Classes 1 and 2 contain many expected proteins, such as cytoskeletal proteins (actin and tubulin), myosin, and ribosomal proteins (all Class 1), and venom allergen protein, and several diedel-like proteins (Class 2) (Table 1). Class 3 does appear to contain several protein annotations which will require additional analysis and manual curation which will shift proportions. We have also found two Class 3 proteins that are annotated as a QWxxN domain, a predicted protein found in *Enterococcus* bacteria according to RefSeq (accession number WP_016625331.1), with an e-value between 1E-12 to 1E-22.

Future modeling of these proteins may indicate how similar the *Lb G* protein is to the bacterial. If highly similar, this result may be an indication of prokaryotic-like proteins being used as part of the wasp strategy. This would also match the presence of prokaryotic-like proteins in the *Lh* proteome, further describing particles of *Leptopilina* wasps as having a “mixed-strategy” (*8*).

Enrichment analysis and predicted secreted/membrane proteins

An enrichment analysis to see if *Lb G* VLPs share a similar profile to *Lh 14* revealed that just as *Lh 14* MSEVs have a profile enriched in exosome proteins (*8, 10*), *Lb G* VLPs are similarly significantly enriched (Fig 1C). In fact, *Lb G* VLPs have a greater enrichment than *Lh 14* MSEVs (56.8% compared to 48.8%). This interpretation is consistent with the report of Di Giovanni et al., that also used the same ontogeny terms for annotating the *Lh* and *Lb* proteins and found similar results (*9*).

The signal peptide and transmembrane domain searches revealed that more than 80% of the *Lb G* proteins may not contain these sequences (89.3% and 84.3% respectively), and a small portion of proteins may be secreted or inserted on to the VLP membrane. Interestingly, only 11/383 (2.9%) proteins were predicted to have both these protein trafficking signals (Table 1). Thus, only a small number of proteins represent candidates for mediating interaction with the hosts’ (hemocyte) cell membrane.

Class 1 proteins

We also noted that *Lh 14* had 12 proteins and *Lb G* particles had 11 proteins with signal peptides and transmembrane domains, indicating these proteins may be components of the particles responsible for interacting with host cells. Comparing these proteins via BLAST indicated three were highly similar (E-values < 1E-140, %ID and query cov. > 80%) and their annotations were for malectin, translocon-associated protein subunit beta, and transmembrane emp24 domain-containing protein bai (Table 6), and all three were within Class 1. Malectin is a protein found in the endoplasmic reticulum and participates in quality control of glycoproteins (*32-34*). Translocon-associated protein subunit beta is associated with transport of proteins into the endoplasmic reticulum (*35, 36*). Transmembrane emp24 domain-containing protein bai is a cargo receptor and also aids in vesicle formation (*37*). All three of these proteins are highly associated with cellular endoplasmic reticulum and endomembrane system.

Another 2/11 scored poorly, with e-values near 1E-4, %ID < 50%, and query coverage ≤ 10%, These two proteins were also found to be mostly Class 1 proteins, but annotations did not match between the proteins (Table 6). The remaining 6 proteins either did not have significant similarities or were already matched to a protein which had a much higher scoring subject. The presence of such proteins, highly associated with endomembrane system and endoplasmic reticulum support the hypothesis of an extracellular vesicle nature as these organelles are involved in the production of extracellular vesicles in the form of exosomes released via multivesicular bodies (*38-40*).

Class 2 proteins

Examining the 10 Class 2 proteins that were found to be highly similar between *Lb G* and *Lh 14* (Fig 1A, Table 4) revealed three proteins to be annotated as ras-related rab GTPases (3/10, 30%, Table 4). Rab GTPases are involved in not only moving effector proteins onto membrane, but also vesicle formation and fusion (*41*). Also found in the common 115 proteins were metalloproteinases and venom allergen 3-like proteins, which were previously described as components of the *Lh* proteome and transcriptome (Table 4) (*8, 42*). Venom allergen 3 is a known acid phosphatase, which triggers allergic responses after hymenopteran stings and is found in many hymenopteran venoms (*7, 42-48*). As an acid phosphatase, it may also work in conjunction with metalloproteinases to digest host tissue to release nutrients for the growing wasp larva. A single enhancing factor and diedel-like protein were also found in common (Table 5). Both are proteins involved in modulating host immune response, especially diedel which is also found to be co-opted by insect viruses, parasitoid venoms, and as an immune modulatory cytokine in *Drosophila* (*8, 42, 49*).

We previously identified a family of prokaryotic-like GTPases in *Lh* MSEVs that is absent in *Lb 17* transcripts (*8*) . The equivalent GTPase are absent within the *Lb G* proteome. Similarly, p40 was not identified in the *Lb G* proteome. In contrast, LbGAP (a known *Lb* virulence protein (*50-52*), g27718.t1), has counterparts in *Lh 14*. This result supports our previous results that *Lh* and *Lb* may have similar RhoGAP proteins in their proteomes (*8*). At least 6 RhoGAPs were reported in *Lh* MSEVs (*8*), however, only one high scoring candidate (GAJC01024661.1_16) met our strict thresholds.

Comparison of PDV/viral profiles

It is intriguing that the proportions of proteins in both *Lh 14* and *Lb G* proteomes with viral hits (PDV and Viridae) is quite similar. While the *Lh 14* proteome contains 6/407 (1.47%) PDV hits and 84/407 (20.63%) Viridae hits, the *Lb G* proteome has 7/383 PDV (1.8%) and 101/383 Viridae (26.3%) hits. These latter results for *Lh 14* are discussed in Chapter 1. Results from the *Lb G* proteome are presented next.

When searching the PDV protein database, 7/383 (1.8%) *Lb G* proteins returned hits. However, all 7 proteins received better hits against eukaryotic proteins in the nr database (Table 7). When this search was repeated using Viridae protein database, 101/383 (26.3%) proteins had hits. All but one protein (Lb_LbFV_ORF85), had better matches against eukaryotic proteins (Table 8). Interestingly, the Lb_LbFV_ORF85 did not match a LbFV protein sequence, but instead matched a *Drosophila* filamentous virus protein, annotated as Ac81-like (Table 8). The same Ac81-like annotation was also given to Lb_LbFV_ORF85 by Di Giovanni et al. (*9*). Ac81 is involved in nucleocapsid envelopment of *Autographa californica nuclear polyhedrosis virus,* a highly pathogenic baculovirus that targets insects (*53*). The evidence for the Lb_LbFV_ORF85 counterpart's included in the *Lh 14* proteome is not convincing (*9, 10*)

A similar result was obtained in both restricted and unrestricted nr databases with virus-like diedels in the *Lh* proteome (*8, 10*). While diedel homologs are present in viruses and eukaryotic hosts (*49*), this Ac81-like protein in the *Lb G* proteome remains firmly viral and essential for virus envelopment alongside several other core genes (*53*). Together, these results support our method of evaluating viral versus non-viral sequences and also the work of Di Giovanni et al. from Dr. Varaldi's group.

Genomic locations of both Lh MSEV and Lb VLP proteins

New genome assemblies for *Lh 14* and *Lb* strain G486 (*Lb G486*) were released by Huang et al. (*11*). We previously found 375/407 (92%) of MSEV proteins to be encoded the *Lh 14* genome (Chapter 1, (*10*)). When these sequences were searched against a new *Lh 14* assembly, 386/407 (94.8%) genes could be mapped to genomic scaffolds. The bulk (18/21) of the remaining 21 proteins fall within Class 3. The remaining three proteins are Class 1 proteins, two of which have annotations for troponin and Y-box factor. The third Class 1 protein is unannotated. We mapped all but four of the 383 (98.9%) *Lb G* VLP genes against the *Lb G486* genome. Two of the unmapped sequences were in Class 1 and annotated as a 60S acidic ribosomal protein and a mesencephalic astrocyte-derived neurotrophic factor homolog. The

other two unmapped genes were in Class 3. The discrepancies in total proteins to mapped genes in both *Lh* and *Lb* could be due to strain differences, issues with gene prediction (unusual or non-canonical gene structures); locations in heterchromatic genomic regions affecting assembly, close identity with other gene families members or inaccurate alternative splicing predictions.

Concluding remarks

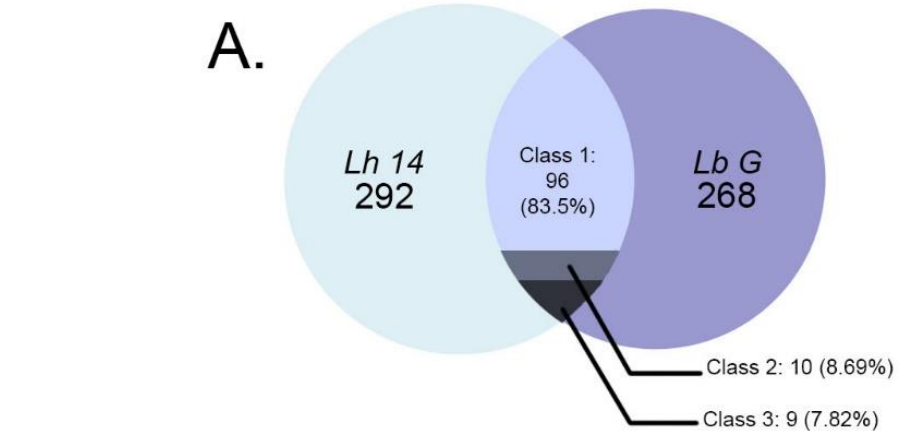
Here we have presented a set of proteins common between both VLPs and MSEVs. A closer examination of this group of proteins should be useful in understanding how these particles are produced by venom gland cells, what proteins influence particle structure, and the mechanism of particle entry/fusion with host immune cells. Our results suggest that like *Lh* MSEVs the *Lb* VLPs should be considered extracellular vesicles. Furthermore, both proteomes have similar profiles and lack Viridae sequences. We believe this analysis provides a basis for further comparative studies between *Leptopilina* species and supports the hypothesis that venom particles found within the *Leptopilina* genus possess extracellular vesicle-like nature. Future studies with the rapidly increasing amount of data available on GenBank and in publications will allow for these comparative studies to include other *Leptopilina* species of interest like *L. clavipes* and *L. victoriae*. These will, in turn, aid in addressing virus integration hypotheses and also drawing a distinction between origin of sequences required to synthesize particles versus the current nature of the particle itself.

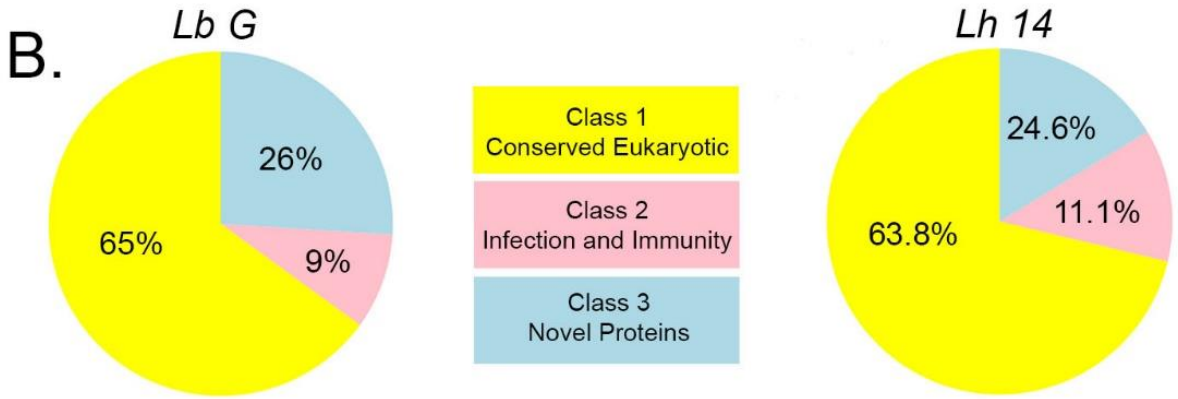
ACKNOWLEDGEMENTS:

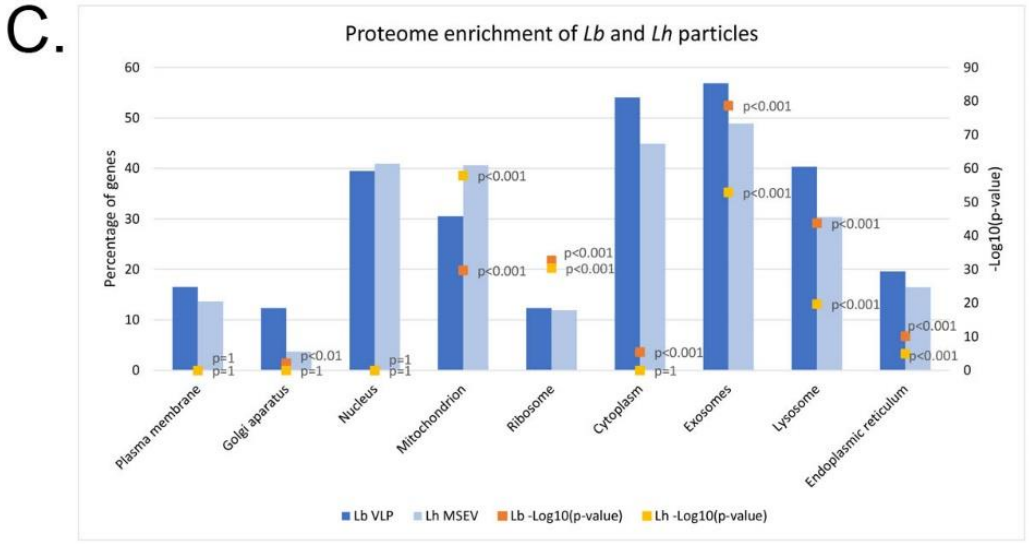
Raw data for *Leptopilina boulardi* VLP protein sequences was provided by Dr. Julien Varaldi (University of Lyon). Bioinformatics work was conducted in-house and with the BIOMIX Shared Computing Cluster at Delaware Biotechnology Institute, University of Delaware. We are grateful to Dr. K. Hopper for continued sharing of these resources. Additional thanks to J. Chou for critical commentary.

FIGURES

Figure 1: Comparison of *Lb G* VLPs to *Lh 14* MSEVs and profiling *Lb G* VLP proteome. A) *Lb G* VLP proteins were compared to *Lh 14* MSEV proteins. Homologous proteins, indicated by class, are within the overlap of the Venn diagram. B) Distribution of *Lb G* VLP proteins and *Lh 14* MSEV proteins by class. C) Enrichment analysis of *Lb G* VLP proteins and *Lh 14* MSEV proteins as percentage of genes analyzed by FunRich. P-value provided as scatterplot points.


A.
Lh 14
292
Class 1:
96
(83.5%)
Lb G
268
Class 2: 10 (8.69%)
Class 3: 9 (7.82%)


B.
Lb G
26%
65%
9%
Class 1
Conserved Eukaryotic
Class 2
Infection and Immunity
Class 3
Novel Proteins
Lh 14
24.6%
63.8%
11.1%


C.
Proteome enrichment of Lb and Lh particles
Percentage of genes
-Log10(p-value)
Plasma membrane
Golgi aparatus
Nucleus
Mitochondrion
Ribosome
Cytoplasm
Exosomes
Lysosome
Endoplasmic reticulum
Lb VLP
Lh MSEV
Lb -Log10(p-value)
Lh -Log10(p-value)

TABLES

Table 1: Full listing of *L. boulardi* VLP proteins analyzed in this work. Lb VLP Protein column indicates protein sequences of the *Lb* Gotheron strain provided by Varaldi et al (*13*). Class indicates assignment of a protein as "conserved eukaryotic" (Class 1), "infection and immunity-related" (Class 2), or "novel" (Class 3). Annotations were determined by OmicsBox (see Methods). The "Equivalent Accession" column provides accession numbers for transcripts from three Lb transcriptomes. (The translated sequence shows a high match to the analyzed VLP protein.) Signal peptide (SP: Y/N) and transmembrane domain (TM: Y/N) presence/absence are indicated in respective columns. N/A indicates no annotation from OmicsBox. NF refers to absence of equivalent accession number within available transcriptomes. Proteins that failed annotation by OmicsBox but obtained GO terms from UniProt/InterProScan searches and were still classified.

***L. boulardi* VLP protein annotations**					
Lb VLP protein	Class	Annotation	Equivalent Accession	SP Y/N	TM Y/N
g10226.t1	1	60S ribosomal protein L35a	GITC01193275,GAJA01021476.1	N	N
g5735.t1	1	protein disulfide-isomerase A6	GITC01132888,GISX01095206,GAJA01020761.1	N	N
g11424.t1	1	adenosine deaminase 2	GITC01181039,GAJA01004793.1	Y	N
g6755.t1	1	protein 5NUC-like	GITC01201240,GISX01083143,GAJA01015634.1	Y	N
g20540.t1	1	N/A	GAJA01008711.1	N	N
g4531.t1	1	trifunctional enzyme subunit alpha, mitochondrial	GITC01050226,GISX01127213,GAJA01008336.1	N	N
g25630.t1	1	glucose dehydrogenase [FAD, quinone]-like	GITC01127954,GISX01065617,GAJA01017927.1	N	N
g1469.t1	1	endoplasmic reticulum resident protein 29	GITC01196819,GISX01099290,GAJA01018369.1	Y	N
g6244.t1	1	trehalase-like isoform X1	GITC01144334,GAJA01008723.1	N	N
g3977.t1	1	carnitine O-palmitoyltransferase 1, liver isoform X1	GITC01173671,GAJA01008019.1	N	Y
g11933.t1	1	ADP-ribosylation factor 2	GAJA01008152.1	N	N
g3337.t1	1	cytochrome b-c1 complex subunit 2, mitochondrial	GAJA01019584.1	N	N
g6174.t1	1	60S ribosomal protein L36	GITC01120985,GAJA01019440.1	N	N
g11454.t1	1	60S ribosomal protein L19	GAJA01019023.1	N	N
g3959.t1	1	extended synaptotagmin-2 isoform X1	GITC01067467,GAJA01008133.1	N	N
g14301.t1	1	60S ribosomal protein L21	GITC01133828,GAJA01020597.1	N	N
g551.t1	1	60S ribosomal protein L9	GAJA01020417.1	N	N
g5708.t1	1	spectrin alpha chain isoform X1	GITC01132619,GISX01094263,GAJA01019739.1	N	N
g8283.t1	1	Myosin heavy chain, muscle	GITC01193182,GISX01003011,GAJA01011172.1	N	N
g7680.t1	1	translocon-associated protein subunit alpha	GAJA01020481.1	N	Y

g12490.t1	1	heat shock 70 kDa protein cognate 3 isoform X2	GITC01027938,GISX01081150,GAJA01019253.1	Y	N
g12.t1	1	dnaJ homolog subfamily C member 3	GITC01132708,GISX01050623,GAJA01020807.1	N	N
g7587.t1	1	Transmembrane 9 superfamily member 3	GITC01084883,GAJA01007370.1	N	Y
g13577.t1	1	40S ribosomal protein S20	GITC01142898,GAJA01018534.1	N	N
g11084.t1	1	probable uridine nucleosidase 1 isoform X1	GITC01157756,GAJA01011580.1	Y	N
g8239.t1	1	lysosome-associated membrane glycoprotein 1	GITC01120700,GAJA01018886.1	N	Y
g1748.t1	1	phosphate carrier protein, mitochondrial	GITC01072000,GISX01097333,GAJA01018389.1	N	N
g9326.t1	1	clathrin heavy chain	GITC01144177,GAJA01018156.1	N	N
g7141.t1	1	multiple coagulation factor deficiency protein 2 homolog isoform X1	GITC01163355,GAJA01007722.1	Y	N
g821.t1	1	protein transport protein Sec61 subunit alpha	GITC01196646,GAJA01019710.1	N	Y
g3528.t1	1	titin homolog isoform X1	GITC01104896,GAJA01012615.1	N	N
g10794.t1	1	3-hydroxyacyl-CoA dehydrogenase type-2	GITC01159264,GISX01126801,GAJA01019916.1	N	N
g13588.t1	1	40S ribosomal protein S5	GITC01003012,GAJA01019151.1	N	N
g4483.t1	1	adipocyte plasma membrane-associated protein	GITC01095244,GAJA01007042.1	N	Y
g2575.t1	1	ATP synthase subunit O, mitochondrial	GITC01160107,GAJA01020429.1	N	N
g12444.t1	1	isocitrate dehydrogenase [NAD] subunit gamma, mitochondrial	GITC01106947,GISX01010734,GAJA01020762.1	N	N
g8679.t1	1	60S ribosomal protein L5	GITC01140390,GAJA01020775.1	N	N
g13417.t1	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	GITC01117032,GAJA01021288.1	Y	Y
g13648.t1	1	protein toll	GITC01201704,GISX01131040,GAJA01009829.1	N	N
g12771.t1	1	stromal cell-derived factor 2	GITC01199234,GAJA01020360.1	Y	N
g1465.t1	1	short-chain specific acyl-CoA dehydrogenase, mitochondrial	GITC01170655,GISX01120676,GAJA01001371.1	N	N
g10210.t1	1	ATP synthase subunit alpha, mitochondrial	GITC01171643,GISX01091541,GAJA01018157.1	N	N
g24769.t1	1	protein 5NUC-like	GITC01099956,GAJA01005448.1	N	N
g669.t1	1	tubulin alpha-1 chain	GITC01004580,GISX01036390,GAJA01021050.1	N	N
g11746.t1	1	putative fatty acyl-CoA reductase CG8306	GITC01065215,GISX01090367,GAJA01007151.1	N	Y
g1069.t1	1	probable citrate synthase 2, mitochondrial	GITC01026697,GAJA01020546.1	N	N
g408.t1	1	dnaJ homolog subfamily A member 1	GITC01053845,GAJA01012937.1	N	N
g19132.t1	1	arginine kinase	GITC01022209,GAJA01020556.1	N	N
g11592.t1	1	protein disulfide-isomerase	GITC01160504,GISX01095206,GAJA01018442.1	Y	N
g2224.t1	1	transketolase-like protein 2 isoform X2	GITC01119652,GISX01102650,GAJA01019937.1	N	N
g3443.t1	1	annexin B9	GITC01010951,GISX01061896,GAJA01019905.1	N	N
g14011.t1	1	aldehyde dehydrogenase, mitochondrial	GITC01048965,GAJA01019471.1	N	N
g15500.t1	1	sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like	GITC01026657,GAJA01006995.1	Y	N
g5201.t1	1	catalase isoform X1	GITC01051492,GISX01143135,GAJA01020919.1	N	N

g3942.t1	1	40S ribosomal protein S4	GITC01106705,GAJA01019641.1	N	N
g14952.t1	1	Protein Dok-7	GITC01200504,GAJA01005216.1	N	N
g12611.t1	1	renin receptor isoform X2	GITC01150552,GAJA01005954.1	Y	Y
g5500.t1	1	acetyl-coenzyme A transporter 1	GITC01199066,GAJA01000834.1	N	Y
g8282.t1	1	myosin heavy chain, muscle isoform X1	GITC01177677,GAJA01020142.1	N	N
g26546.t1	1	alaserpin isoform X5	GITC01198069,GAJA01019393.1	N	N
g5600.t1	1	peptidyl-alpha-hydroxyglycine alpha-amidating lyase 1	GITC01103963,GAJA01016066.1	N	Y
g2029.t1	1	RNA-binding protein squid	GITC01017542,GISX01110266,GAJA01020377.1	N	N
g14212.t1	1	3-ketoacyl-CoA thiolase, mitochondrial	GITC01033377,GISX01158239,GAJA01006839.1	N	N
g2999.t1	1	40S ribosomal protein S15Aa	GITC01179326,GAJA01019089.1	N	N
g3358.t1	1	beta-mannosidase	GITC01094070,GISX01102084,GAJA01002610.1	N	N
g7274.t1	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	GAJA01008035.1	Y	Y
g249.t1	1	60S ribosomal protein L10a	GAJA01018439.1	N	N
g469.t1	1	protein lethal(2)essential for life-like	GITC01052564,GISX01038005,GAJA01019474.1	N	N
g1869.t1	1	multidrug resistance protein 1A	GITC01110448,GISX01006229,GAJA01011165.1	N	Y
g699.t1	1	myosin regulatory light chain 2	GITC01078317,GAJA01021209.1	N	N
g3507.t1	1	peptidylglycine alpha-hydroxylating monooxygenase	GITC01023314,GISX01010886,GAJA01015755.1	N	N
g4524.t1	1	60S ribosomal protein L17	GITC01204473,GAJA01018538.1	N	N
g11675.t1	1	sulfhydryl oxidase 2-like	GITC01158188,GISX01007172,GAJA01014272.1	Y	Y
g15893.t1	1	protein transport protein Sec24C	GITC01133363,GISX01111185,GAJA01007520.1	N	N
g3992.t1	1	60S ribosomal protein L13a	GITC01092991,GAJA01018432.1	N	N
g13184.t1	1	alpha-tocopherol transfer protein-like	GITC01160980,GISX01010018,GAJA01015109.1	N	N
g18935.t1	1	dnaJ homolog shv	GITC01044410,GAJA01015175.1	Y	N
g11607.t1	1	60S ribosomal protein L31	GITC01030749,GAJA01020670.1	N	N
g7824.t1	1	endoplasmin	GAJA01018420.1	N	N
g10788.t1	1	RNA-binding protein squid-like isoform X8	GITC01048110,GAJA01007899.1	N	N
g8876.t1	1	transitional endoplasmic reticulum ATPase TER94	GITC01111868,GAJA01020404.1	N	N
g634.t1	1	Cytochrome c oxidase subunit 4 isoform 1, mitochondrial	GITC01091019,GISX01081319,GAJA01020379.1	N	Y
g5307.t1	1	proteasomal ubiquitin receptor ADRM1 homolog isoform X2	GITC01022665,GAJA01019012.1	N	N
g3893.t1	1	troponin T, skeletal muscle isoform X3	GITC01179722,GAJA01006009.1	N	N
g10473.t1	1	60S ribosomal protein L27a	GITC01093705,GAJA01019244.1	N	N
g8001.t1	1	vesicle transport protein GOT1B	GITC01039099,GAJA01006852.1	N	Y
g20934.t1	1	lipase 3-like	GITC01045500,GISX01029643,GAJA01018823.1	N	N
g5997.t1	1	juvenile hormone epoxide hydrolase 1-like	GITC01204935,GISX01057930,GAJA01006104.1	Y	N
g12024.t1	1	probable beta-hexosaminidase fdl isoform X1	GITC01046075,GISX01007846,GAJA01017821.1	N	N
g1298.t1	1	60S ribosomal protein L6	GITC01167220,GAJA01019243.1	N	N
g4353.t1	1	dolichyl-diphosphooligosaccharide--	GAJA01006801.1	Y	Y

		protein glycosyltransferase 48 kDa subunit			
g11414.t1	1	Hermansky-Pudlak syndrome 4 protein	GITC01086439,GISX01052876,GAJA01002104.1	N	N
g6263.t1	1	annexin B9 isoform X1	GITC01010951,GAJA01019905.1	N	N
g15014.t1	1	sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like	GITC01026654,GAJA01018728.1	N	N
g26437.t1	1	histone H4-like	GITC01147019,GISX01052793,GAJA01018362.1	N	N
g5514.t1	1	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	GITC01138609,GISX01034114,GAJA01006925.1	N	N
g7129.t1	1	aldehyde dehydrogenase, dimeric NADP-preferring isoform X2	GITC01166397,GAJA01002685.1	N	Y
g3059.t1	1	protein wntless	GITC01000968,GISX01132444,GAJA01006542.1	N	Y
g1756.t1	1	sequestosome-1 isoform X1	GISX01052120,GAJA01018105.1	N	N
g11702.t1	1	cytochrome P450 4g15-like	GITC01100946,GISX01110434,GAJA01019633.1	N	Y
g21089.t1	1	alaserpin isoform X5	GITC01198069,GAJA01020187.1	N	N
g3257.t1	1	endoplasmic reticulum-Golgi intermediate compartment protein 3	GITC01108972,GAJA01007709.1	N	Y
g332.t1	1	procollagen-lysine,2-oxoglutarate 5-dioxygenase 1 isoform X1	GITC01002380,GISX01038304,GAJA01019208.1	N	N
g16717.t1	1	Calreticulin	GITC01002549,GAJA01020312.1	N	N
g92.t1	1	40S ribosomal protein S23	GITC01173041,GAJA01019291.1	N	N
g4759.t1	1	putative beta-carotene-binding protein	GITC01034596,GISX01157145,GAJA01011141.1	N	Y
g8776.t1	1	ATP synthase subunit beta, mitochondrial	GITC01204992,GISX01091541,GAJA01018617.1	N	N
g13146.t1	1	FChain F, Histone H4	GITC01104094,GISX01052793,GAJA01018362.1	N	N
g3616.t1	1	probable cytochrome P450 6a13	N/F	N	Y
g3726.t1	1	translocon-associated protein subunit delta	GITC01003513,GISX01126147,GAJA01019639.1	Y	N
g3853.t1	1	probable isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	GITC01145192,GISX01010734,GAJA01008534.1	N	N
g1259.t1	1	protein disulfide-isomerase A3	GITC01089254,GISX01095206,GAJA01019020.1	Y	N
g4237.t1	1	alkaline phosphatase	GITC01145350,GISX01050580,GAJA01014852.1	N	N
g12983.t1	1	vesicle-trafficking protein SEC22b-B	GITC01025850,GAJA01000110.1	N	Y
g3017.t1	1	moesin/ezrin/radixin homolog 1 isoform X2	GITC01073898,GAJA01014209.1	N	N
g8285.t1	1	Myosin heavy chain, muscle	GITC01115643,GAJA01019881.1	N	N
g13472.t1	1	uncharacterized protein LOC105834636	GITC01202287,GAJA01011613.1	N	N
g7864.t1	1	Succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial	GITC01053277,GAJA01020185.1	N	N
g8412.t1	1	40S ribosomal protein S3	GITC01163517,GISX01053088,GAJA01018456.1	N	N
g12533.t1	1	V-type proton ATPase 116 kDa subunit a isoform X2	GITC01063707,GAJA01016358.1	N	Y
g12977.t1	1	elongation factor 1-beta'	N/F	N	N
g732.t1	1	signal peptidase complex subunit 1	GAJA01006328.1	N	Y
g12228.t1	1	voltage-dependent anion-selective channel-like	GAJA01019071.1	N	N
g9610.t1	1	trifunctional enzyme subunit beta, mitochondrial	GITC01033377,GISX01072038,GAJA01021118.1	N	N

g4242.t1	1	ABC transporter G family member 20 isoform X1	GITC01142795,GISX01021523,GAJA01014091.1	N	N
g15809.t1	1	putative methyltransferase TARBP1	GITC01139736,GAJA01007968.1	N	N
g9327.t1	1	60S acidic ribosomal protein P2	GITC01102488,GAJA01020490.1	Y	N
g329.t1	1	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	GAJA01019207.1	N	N
g9718.t1	1	60S ribosomal protein L23a	GITC01095446,GAJA01019658.1	N	N
g12336.t1	1	protein 60A	GITC01022418,GISX01157710,GAJA01004651.1	N	N
g7969.t1	1	60S ribosomal protein L30	GISX01018125,GAJA01018303.1	N	N
g6087.t1	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B isoform X1	GITC01193794,GISX01025473,GAJA01018737.1	N	Y
g729.t1	1	ubiquitin-like protein FUBI	GITC01197277,GAJA01018201.1	N	N
g11435.t1	1	DNA repair protein complementing XP-C cells homolog	N/F	N	N
g658.t1	1	peptidyl-prolyl cis-trans isomerase 5	GITC01176579,GISX01012484,GAJA01018693.1	Y	N
g8002.t1	1	ADP,ATP carrier protein 2	GITC01069314,GAJA01019573.1	N	Y
g13106.t1	1	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	GITC01193666,GAJA01006656.1	N	N
g5585.t1	1	40S ribosomal protein S14	GITC01023987,GISX01113011,GAJA01020790.1	N	N
g13251.t1	1	translocation protein SEC63 homolog	GITC01147975,GAJA01021337.1	N	Y
g10034.t1	1	glyceraldehyde-3-phosphate dehydrogenase 2-like	GITC01193871,GISX01024927,GAJA01018844.1	N	N
g7654.t1	1	mesencephalic astrocyte-derived neurotrophic factor homolog	GITC01166479,GAJA01006568.1	Y	N
g2193.t1	1	V-type proton ATPase subunit B	GITC01062615,GISX01091541,GAJA01021129.1	N	N
g5016.t1	1	ribonuclease Oy	GITC01059225,GISX01069040,GAJA01018081.1	N	N
g7903.t1	1	60S ribosomal protein L14	GITC01108265,GAJA01018420.1	N	N
g8216.t1	1	surfeit locus protein 4 homolog	GITC01019546,GAJA01021123.1	N	Y
g8500.t1	1	coatomer subunit alpha	GITC01070530,GAJA01020686.1	N	N
g3124.t1	1	actin-5C	GITC01186843,GISX01092424,GAJA01020288.1	N	N
g16716.t1	1	Calreticulin	GITC01067668,GAJA01020312.1	Y	N
g12892.t1	1	adenylate cyclase type 10-like	N/F	N	N
g14686.t1	1	vesicular integral-membrane protein VIP36	GITC01175381,GISX01089820,GAJA01016135.1	N	Y
g8915.t1	1	40S ribosomal protein S12	GITC01103797,GISX01123001,GAJA01018094.1	N	N
g12939.t1	1	vesicle-associated membrane protein-associated protein B	GAJA01018960.1	N	Y
g3791.t1	1	fasciclin-2 isoform X2	GITC01164621,GISX01017607,GAJA01019952.1	N	N
g21501.t1	1	lysosomal alpha-mannosidase isoform X1	GITC01143013,GISX01151660,GAJA01012944.1	N	N
g11743.t1	1	40S ribosomal protein S3a	GITC01090593,GAJA01021030.1	N	N
g6685.t1	1	60S ribosomal protein L7a	GITC01204455,GAJA01019490.1	N	N
g22703.t1	1	spermine oxidase	GITC01031358,GISX01103130,GAJA01019353.1	N	N
g15153.t1	1	sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like	GITC01031891,GISX01058185,GAJA01008402.1	N	N

g5494.t1	1	translocon-associated protein subunit beta	GITC01057342,GAJA01018646.1	Y	Y
g11874.t1	1	protein YIPF1	GAJA01018743.1	N	Y
g12308.t1	1	coatomer subunit gamma	GITC01143614,GAJA01008807.1	N	N
g4245.t1	1	protein transport protein Sec61 subunit beta	GITC01183753,GISX01013289,GAJA01021066.1	N	Y
g9652.t1	1	coatomer subunit beta	GITC01184187,GAJA01007247.1	N	N
g7099.t1	1	Golgi reassembly-stacking protein 2	GITC01155695,GAJA01000385.1	N	N
g5466.t1	1	signal peptidase complex catalytic subunit SEC11A	GITC01176630,GISX01128699,GAJA01007462.1	N	Y
g11920.t1	1	mannosyl-oligosaccharide glucosidase GCS1	GITC01094274,GISX01078827,GAJA01010181.1	N	Y
g18649.t1	1	40S ribosomal protein S19-like	GITC01170878,GAJA01018587.1	N	N
g12970.t1	1	coatomer subunit zeta-1	GITC01055847,GISX01005749,GAJA01018348.1	N	N
g8281.t1	1	myosin heavy chain, muscle isoform X4	GITC01019235,GISX01138967,GAJA01019882.1	N	N
g10826.t1	1	heat shock 70 kDa protein cognate 5	GITC01143233,GISX01081150,GAJA01019253.1	N	N
g22586.t1	1	lysosomal alpha-mannosidase isoform X1	GITC01016947,GISX01151651,GAJA01012944.1	N	N
g186.t1	1	superoxide dismutase [Cu-Zn]	GITC01167390,GISX01040095,GAJA01018817.1	N	N
g5010.t1	1	40S ribosomal protein S15	GITC01142395,GISX01126628,GAJA01018404.1	N	N
g15224.t1	1	Polyadenylate-binding protein 1	GITC01140487,GISX01158596,GAJA01018206.1	N	N
g16252.t1	1	pol polyprotein	GITC01062974	N	N
g12249.t1	1	heat shock protein 83	GITC01034157,GAJA01020548.1	N	N
g8711.t1	1	translocon-associated protein subunit gamma	GAJA01020036.1	N	Y
g8740.t1	1	fructose-bisphosphate aldolase-like isoform X1	GITC01051343,GAJA01020057.1	N	N
g13994.t1	1	paramyosin, long form	GITC01138280,GISX01138967,GAJA01000548.1	N	N
g10248.t1	1	FK506-binding protein 2 isoform X2	GITC01095223,GAJA01020242.1	N	N
g4533.t1	1	Trifunctional enzyme subunit alpha, mitochondrial	GITC01050226,GAJA01008336.1	N	N
g14457.t1	1	CD63 antigen	GITC01050580,GISX01143362,GAJA01007761.1	Y	N
g8573.t1	1	malate dehydrogenase, mitochondrial	GAJA01019365.1	N	N
g11819.t1	1	cofilin/actin-depolymerizing factor homolog	GITC01039102,GISX01132084,GAJA01019586.1	N	N
g1500.t1	1	40S ribosomal protein SA	GAJA01020035.1	N	N
g8024.t1	1	aminoacylase-1	GITC01150154,GISX01072045,GAJA01015784.1	N	N
g16496.t1	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit DAD1	GAJA01020840.1	N	N
g11523.t1	1	very long-chain specific acyl-CoA dehydrogenase, mitochondrial	GITC01065884,GISX01120676,GAJA01008439.1	N	N
g6246.t1	1	aconitate hydratase, mitochondrial-like	GAJA01020420.1	N	N
g5064.t1	1	enolase isoform X2	GAJA01020476.1	N	N
g11203.t1	1	probable medium-chain specific acyl-CoA dehydrogenase, mitochondrial	GITC01148053,GISX01098103,GAJA01005888.1	N	N
g7569.t1	1	elongation factor 1-alpha	GISX01095435,GAJA01020170.1	N	N
g9735.t1	1	coatomer subunit epsilon	GITC01135087,GAJA01021164.1	N	N

g8480.t1	1	peptidyl-prolyl cis-trans isomerase	GITC01200692,GISX01017882,GAJA01018970.1	N	N
g7229.t1	1	muscle-specific protein 20	GITC01128855,GAJA01019150.1	N	N
g416.t1	1	60s ribosomal protein l15	GAJA01020535.1	N	N
g11425.t1	1	adenosine deaminase 2-like	GITC01075799,GAJA01004793.1	N	N
g3549.t1	1	40S ribosomal protein S8	GITC01061995,GAJA01018620.1	N	N
g11199.t1	1	transmembrane emp24 domain-containing protein eca	GITC01098662,GISX01075598,GAJA01020423.1	Y	Y
g13896.t1	1	glucose dehydrogenase [FAD, quinone]-like	GITC01058197,GISX01058493,GAJA01006729.1	N	N
g2387.t1	1	metalloproteinase inhibitor 3 isoform X2	GITC01048396,GISX01018531,GAJA01019753.1	Y	N
g13378.t1	1	lipase 3-like	GITC01047891,GISX01029643,GAJA01017509.1	N	N
g15643.t1	1	protein toll	GITC01073170,GISX01131040,GAJA01000999.1	N	N
g2352.t1	1	60S ribosomal protein L4	GITC01194207,GAJA01019555.1	N	N
g9235.t1	1	40S ribosomal protein S11	GITC01194517,GAJA01021186.1	N	N
g23789.t1	1	histone H1/H5 family protein	GITC01028211,GISX01101988,GAJA01020108.1	N	N
g21563.t1	1	40S ribosomal protein S2	GITC01198398,GAJA01020305.1	N	N
g13729.t1	1	histone H2B	GITC01049931,GISX01050887,GAJA01014709.1	N	N
g6787.t1	1	aldose reductase	GITC01133192,GAJA01019227.1	N	N
g6088.t1	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B isoform X1	GITC01072113,GISX01025473,GAJA01018737.1	N	Y
g354.t1	1	40S ribosomal protein S16	GITC01027471,GAJA01019678.1	N	N
g1671.t1	1	V-type proton ATPase catalytic subunit A	GITC01203844,GAJA01020437.1	N	N
g13993.t1	1	Paramyosin, short form	GITC01138272,GISX01138967,GAJA01019881.1	N	N
g5610.t1	1	40S ribosomal protein S28	GITC01103134,GISX01040925,GAJA01020302.1	N	Y
g566.t1	1	polyadenylate-binding protein 4-like	GITC01191850,GISX01158596,GAJA01018206.1	N	N
g176.t1	1	40S ribosomal protein S9	GITC01121164,GAJA01021276.1	N	N
g6386.t1	1	60S acidic ribosomal protein P1	GAJA01020446.1	N	N
g3852.t1	1	probable isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	GITC01106952,GISX01010734,GAJA01008534.1	N	N
g937.t1	1	N/A	GITC01166663,GAJA01002343.1	N	Y
g13332.t1	1	calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type isoform X1	GITC01147818,GAJA01021086.1	N	Y
g5387.t1	1	60S ribosomal protein L8	GAJA01020755.1	N	N
g785.t1	1	40S ribosomal protein S6	GITC01062107,GAJA01020461.1	N	N
g1685.t1	1	poly(U)-specific endoribonuclease homolog	GITC01136711,	Y	N
g2033.t1	1	26S proteasome non-ATPase regulatory subunit 12	GITC01037055,GAJA01019681.1	N	N
g3398.t1	1	neuroglian isoform X1	GITC01149026,GISX01078699,GAJA01000094.1	N	N
g18360.t1	1	40S ribosomal protein S18	GITC01179505,GAJA01018178.1	N	N
g3260.t1	1	peroxiredoxin	GITC01143132,GISX01034181,GAJA01020403.1	N	N
g11248.t1	1	60S ribosomal protein L3	GITC01062754,GAJA01020552.1	N	N
g8413.t1	1	Krueppel homolog 2	GITC01204983,GISX01040721,GAJA01020150.1	N	Y

g9134.t1	1	protein 5NUC-like	GITC01050565,GISX01054517,GAJA01015634.1	N	N
g5940.t1	1	superoxide dismutase [Cu-Zn], chloroplastic-like isoform X1	GITC01089403,GISX01040095,GAJA01020703.1	N	N
g15152.t1	1	farnesol dehydrogenase	GITC01173225,GISX01033949,GAJA01018345.1	N	N
g10249.t1	1	FK506-binding protein 2	GITC01095223,GAJA01020242.1	N	N
g14706.t1	1	protein ERGIC-53	GITC01133290,GAJA01006253.1	Y	Y
g107.t1	1	serine protease inhibitor 3/4-like isoform X4	GITC01078508,GAJA01019395.1	Y	N
g27585.t1	1	glucose dehydrogenase [FAD, quinone]	GITC01001100,GISX01062449,GAJA01006341.1	N	N
g3442.t1	1	annexin B9	GISX01061893,GAJA01008236.1	N	N
g25658.t1	2	endothelin-converting enzyme 2-like isoform X1	GITC01033674,GISX01068272,GAJA01001617.1	N	N
g4648.t1	2	venom acid phosphatase Acph-1-like	GITC01165220,GISX01021119,GAJA01011887.1	N	N
g10681.t1	2	ras-related protein Rab-14	GITC01027133,GISX01068352,GAJA01005981.1	N	N
g20818.t1	2	aminopeptidase N	GITC01140264,GISX01066720,GAJA01011355.1	N	N
g11662.t1	2	ras-related protein Rab-1A	GITC01067731,GISX01068352,GAJA01008437.1	N	N
g23613.t1	2	N/A	GITC01120012,GISX01038173,GAJA01006288.1	Y	N
g882.t1	2	venom allergen 3-like	GITC01006168,GAJA01008706.1	N	N
g7656.t1	2	Rac GTPase-activating protein 1	GITC01120006,GISX01038173,GAJA01020436.1	N	N
g9088.t1	2	Tubulin beta-1 chain	GITC01095487,GAJA01006773.1	N	N
g19482.t1	2	protein Diedel-like	GITC01029986,GISX01083966,GAJA01021338.1	N	N
g11841.t1	2	neutral alpha-glucosidase AB	GITC01095440,GAJA01018312.1	Y	N
g4816.t1	2	GTP-binding protein REM 1	GITC01082018,	N	N
g12817.t1	2	ras-related protein Rab-2	GITC01024510,GISX01068352,GAJA01005981.1	N	N
g3715.t1	2	leukotriene A-4 hydrolase	GITC01063867,GAJA01018399.1	N	N
g8353.t1	2	venom metalloproteinase 3-like	GITC01187796,GISX01033600,GAJA01019047.1	N	N
g7877.t1	2	rac GTPase-activating protein 1-like	GITC01087178,GISX01124339,GAJA01019580.1	N	N
g1968.t1	2	Venom dipeptidyl peptidase 4	GITC01163866,GAJA01002345.1	N	N
g23282.t1	2	venom metalloproteinase 3-like	GITC01190322,GISX01033600,GAJA01006373.1	Y	N
g16194.t1	2	rac GTPase-activating protein 1	GITC01087178,GISX01124353,GAJA01019580.1	N	N
g13022.t1	2	nucleobindin-2 isoform X1	GITC01032445,GAJA01008352.1	Y	N
g3802.t1	2	ADP-ribosylation factor 1-like 2	GITC01129126,GISX01054304,GAJA01001633.1	N	N
g8456.t1	2	tissue inhibitor of metalloproteinase	GITC01048396,GISX01018531,GAJA01015719.1	N	N
g9812.t1	2	atlastin isoform X3	GITC01047626,GAJA01018523.1	N	Y
g6700.t1	2	ras-related protein Rab-7a	GITC01047112,GISX01068352,GAJA01000701.1	N	N
g2172.t1	2	lysosomal aspartic protease	GITC01135517,GISX01011342,GAJA01018052.1	N	N
g18394.t1	2	rac GTPase-activating protein 1	GITC01087148,GISX01124339,GAJA01018901.1	N	N
g9493.t1	2	rac GTPase-activating protein 1	GITC01134882,GISX01038173,GAJA01020282.1	N	N
g14621.t1	2	venom serine carboxypeptidase	GAJA01019126.1	N	N
g14524.t1	2	ADP-ribosylation factor-like protein 1	GISX01052920,GAJA01008428.1	N	N
g7363.t1	2	nuclease-sensitive element-binding protein 1 isoform X2	N/F	N	N
g27718.t1	2	RhoGAP precursor/LbGAP	GITC01087168,GISX01124318,GAJA01018901.1	N	N

g2220.t1	2	GTP-binding protein SAR1b	GITC01198816,GISX01054304,GAJA01021124.1	N	N
g12041.t1	2	ras-related protein Rab-5C	GITC01173649,GISX01151112,GAJA01020758.1	N	N
g3640.t1	3	protein canopy homolog 2	GITC01025928,GAJA01014507.1	N	Y
g1195.t1	3	N/A	N/F	N	N
g27107.t1	3	Putative SopA-like secreted protein	GITC01071609,GISX01040514,GAJA01000967.1	N	N
g27935.t1	3	N/A	GITC01059467,GAJA01004319.1	N	N
g7876.t1	3	N/A	N/F	N	N
g26195.t1	3	Macrophage receptor MARCO	GITC01141109,GISX01059024,	N	N
g21977.t1	3	enhancing factor	GITC01047464,GISX01136729,GAJA01014011.1	N	N
g12151.t1	3	malectin-A	GAJA01011987.1	Y	Y
g15481.t1	3	N/A	GITC01071609,GISX01118346,GAJA01000967.1	N	N
g5976.t1	3	microsomal glutathione S-transferase 1	GITC01083393,GISX01021151,GAJA01018221.1	N	Y
g2479.t1	3	endoplasmic reticulum resident protein 44 isoform X2	GITC01177868,GISX01019322,GAJA01007202.1	Y	N
g7087.t1	3	60S acidic ribosomal protein P0	GITC01030978,GAJA01019000.1	N	N
g27018.t1	3	N/A	GITC01156076,GAJA01003572.1	N	N
g23411.t1	3	QWxxN domain	GITC01156068,GISX01079842,GAJA01012562.1	N	N
g26802.t1	3	S-antigen protein-like	GITC01129438,GAJA01002663.1	N	N
g20948.t1	3	four and a half LIM domains protein 2 isoform X8	GAJA01012187.1	N	N
g9570.t1	3	40S ribosomal protein S10	GITC01035718,GAJA01019911.1	N	N
g26603.t1	3	S-antigen protein-like	GITC01129442,GAJA01002663.1	N	N
g15349.t1	3	Signal recognition particle receptor subunit beta	GITC01148522,GAJA01006494.1	N	N
g28919.t1	3	N/A	GITC01141109,GISX01059024,	N	N
g6093.t1	3	tumor suppressor candidate 3	GAJA01021301.1	N	Y
g20411.t1	3	N/A	GITC01149204,GAJA01018809.1	N	N
g26199.t1	3	N/A	GITC01149197,GAJA01015839.1	N	N
g1247.t1	3	N/A	GITC01030182,GISX01130980,GAJA01007629.1	Y	N
g21730.t1	3	N/A	GITC01156069,GISX01118346,GAJA01000967.1	N	N
g10324.t1	3	N/A	GITC01190350,GISX01033592,GAJA01006377.1	N	N
g20177.t1	3	N/A	GITC01071867,GISX01061533,GAJA01012569.1	N	N
g20322.t1	3	N/A	GITC01071867,GISX01061533,GAJA01012568.1	N	N
g1397.t1	3	scavenger receptor cysteine-rich protein	GITC01023641,GISX01102393,GAJA01020201.1	N	N
g24368.t1	3	N/A	GITC01156069,GISX01118363,GAJA01017988.1	N	N
g26235.t1	3	lysosomal alpha-mannosidase isoform X2	GITC01031693,GISX01151651,GAJA01012944.1	N	N
g28902.t1	3	N/A	GITC01029136,GISX01118367,GAJA01016769.1	N	N
g11786.t1	3	nose resistant to fluoxetine protein 6-like	GITC01078514,GAJA01004776.1	N	Y
g18814.t1	3	PREDICTED: uncharacterized protein LOC105569791, partial	GITC01097315,GAJA01004714.1	N	N
g27130.t1	3	Putative SopA-like secreted protein	GITC01156068,GISX01118363,GAJA01017988.1	N	N
g8451.t1	3	N/A	GITC01026684,GAJA01018118.1	N	N
g22070.t1	3	N/A	GITC01029125,GISX01018255,GAJA01011578.1	N	N

g25868.t1	3	N/A	GITC01053848,GISX01079839,GAJA01012166.1	N	N
g15987.t1	3	N/A	GITC01053848,GISX01079839,GAJA01012937.1	N	N
g247.t1	3	N/A	GITC01022452,GISX01039130,GAJA01000836.1	N	N
Lb_LbFV_ORF85	3	Ac81-like protein	GAJA01001673.1	N	Y
g4000.t1	3	14-3-3 protein zeta isoform X2	GITC01190213,GAJA01021146.1	N	N
g19441.t1	3	N/A	GITC01156063,GISX01118346,GAJA01000968.1	N	N
g19774.t1	3	N/A	GISX01050576,GAJA01004072.1	N	N
g23606.t1	3	N/A	GITC01071872,GISX01061522,GAJA01012569.1	N	N
g25753.t1	3	N/A	GITC01156063,GISX01040514,GAJA01017988.1	N	N
g21419.t1	3	N/A	GITC01105368,GISX01017935,GAJA01000968.1	N	N
g385.t1	3	protein THEM6	GITC01039010,GAJA01008840.1	N	Y
g25831.t1	3	N/A	GITC01141109,	N	N
g10080.t1	3	probable uridine nucleosidase 1 isoform X1	GITC01157759,GAJA01005947.1	N	N
g19192.t1	3	Leucine-rich repeat (LRR) protein	GITC01053848,GISX01079839,GAJA01012937.1	N	N
g22788.t1	3	N/A	GITC01204980,GISX01059465,GAJA01006002.1	N	N
g23729.t1	3	N/A	GITC01090731,GAJA01016213.1	N	N
g3916.t1	3	QWxxN domain	GITC01105368,GAJA01012562.1	N	N
g29741.t1	3	N/A	GITC01153128,GISX01024528,GAJA01020180.1	N	N
g11793.t1	3	receptor expression-enhancing protein 5	GAJA01021283.1	N	Y
g24380.t1	3	pentapeptide repeat-containing protein	GITC01105370,GISX01040514,GAJA01012562.1	N	N
g18880.t1	3	N/A	GITC01156063,GISX01136942,GAJA01015791.1	N	N
g20214.t1	3	N/A	GITC01044639	N	N
g9031.t1	3	staphylococcal nuclease domain-containing protein 1	GITC01105858,GAJA01019200.1	N	N
g28195.t1	3	N/A	GITC01029125,GISX01018255,GAJA01011578.1	N	N
g12150.t1	3	heat shock 70 kda protein cognate 4	GITC01027925,GISX01081150,GAJA01018097.1	N	N
g26590.t1	3	N/A	GITC01190350,GISX01033592,GAJA01006377.1	N	N
g2355.t1	3	tumor protein D52 isoform X2	GITC01199611,GAJA01020060.1	N	N
g24649.t1	3	N/A	GITC01053848,GISX01079839,GAJA01012168.1	N	N
g4910.t1	3	inactive hydroxysteroid dehydrogenase-like protein 1	GITC01131709,GISX01000819,GAJA01008678.1	N	Y
g465.t1	3	TM2 domain-containing protein almondex	GITC01052559,GAJA01002341.1	N	N
g27649.t1	3	hypothetical protein AVEN_6331_1	GITC01141109,GISX01059024,	N	N
g9910.t1	3	general odorant-binding protein 56d-like	GITC01197372,GISX01140095,GAJA01015499.1	Y	N
g10682.t1	3	transmembrane emp24 domain-containing protein-like	GITC01098662,GISX01075598,GAJA01020423.1	Y	Y
g11630.t1	3	protein FAM151B isoform X2	GITC01062209,GAJA01015363.1	N	Y
g2472.t1	3	transferrin-like	GITC01031132,GAJA01014176.1	Y	N
g7964.t1	3	hydroxysteroid dehydrogenase-like protein 2	GITC01044418,GISX01050519,GAJA01006815.1	N	N
g27108.t1	3	N/A	GITC01156068,GISX01118346,GAJA01020069.1	N	N
g19442.t1	3	N/A	GITC01035349,GISX01136942,GAJA01012767.1	N	N
g3917.t1	3	hypothetical protein	GITC01105368,GAJA01000968.1	N	N

g20130.t1	3	N/A	GITC01105360,GISX01017935,GAJA01000967.1	N	N
g1887.t1	3	N/A	GITC01010741,GAJA01018131.1	N	Y
g27228.t1	3	S-antigen protein-like	GITC01129437,GAJA01002660.1	N	N
g11642.t1	3	PRA1 family protein 3	GITC01051020,GAJA01020291.1	N	Y
g25079.t1	3	AGAP000929-PA-like protein	GITC01026659,GAJA01006995.1	N	N
g25011.t1	3	N/A	GITC01168882,GAJA01020320.1	N	N
g15068.t1	3	N/A	GITC01156054,GISX01079840,GAJA01009572.1	N	N
g25395.t1	3	N/A	GITC01023910,GISX01033600,GAJA01006374.1	Y	N
g20781.t1	3	N/A	GITC01071613,GISX01079839,GAJA01012562.1	N	N
g3999.t1	3	14-3-3 protein zeta isoform X1	GAJA01021146.1	N	N
g25297.t1	3	N/A	GITC01156063,GISX01040514,GAJA01003572.1	N	N
g24465.t1	3	N/A	GITC01029125,GISX01018255,GAJA01011578.1	N	N
g23005.t1	3	Leucine-rich repeat (LRR) protein	GITC01197868,GISX01079839,GAJA01012937.1	N	N
g22206.t1	3	N/A	GITC01129447,GAJA01002663.1	N	N
g28873.t1	3	N/A	GITC01105360,GISX01017935,GAJA01000968.1	N	N
g22025.t1	3	N/A	GITC01104099,GISX01105974,GAJA01020753.1	N	N
g899.t1	3	N/A	N/F	N	N
g2320.t1	3	putative nuclease HARBI1	GITC01029125,GISX01118346,GAJA01016769.1	N	N
g26022.t1	3	N/A	GITC01016685,GISX01059471,GAJA01004736.1	N	N
g25331.t1	3	N/A	GAJA01015285.1	N	N
g19696.t1	3	N/A	GITC01053838,GISX01079839,GAJA01012937.1	N	N
g19027.t1	3	N/A	GITC01029125,GISX01018255,GAJA01011578.1	N	N
g13391.t1	3	peroxisomal multifunctional enzyme type 2	GITC01143122,GAJA01020391.1	N	N
g21443.t1	3	N/A	GITC01022450,GAJA01000837.1	N	N
g9643.t1	3	uncharacterized protein LOC103318133	GAJA01014376.1	N	N
g9705.t1	3	transmembrane emp24 domain-containing protein bai isoform X2	GITC01140546,GISX01101177,GAJA01020958.1	Y	Y
g17673.t1	3	Leucine-rich repeat (LRR) protein	GITC01053839,GISX01079839,GAJA01012937.1	N	N
g23839.t1	3	N/A	GITC01156054,GISX01079839,GAJA01012937.1	N	N
g9726.t1	3	double-stranded RNA-specific editase 1 isoform X3	GITC01197713,GAJA01021244.1	N	N
g15166.t1	3	N/A	GITC01071872,GISX01061522,GAJA01012569.1	N	N
g17565.t1	3	Leucine-rich repeat (LRR) protein	GITC01105368,GISX01017935,GAJA01000968.1	N	N
g25522.t1	3	N/A	GITC01156069,GISX01118346,GAJA01000968.1	N	N
g17667.t1	3	hypothetical protein CsmBAC4b19.5 [Cotesia sesamiae]	GITC01061878,GAJA01010387.1	N	N
g27699.t1	3	N/A	GITC01156063,GISX01118346,GAJA01000967.1	N	N
g19652.t1	3	N/A	GITC01028117,GISX01084597,GAJA01011758.1	N	N
g13578.t1	3	N/A	GITC01144757,GAJA01018638.1	N	N
g6078.t1	3	N/A	GITC01127826,GISX01133594,GAJA01019220.1	N	N

Table 2: Full listing of Lh 14 MSEV proteins with class, annotation, and presence/absence of signal peptide and transmembrane domain. Signal peptide (SP: Y/N) and transmembrane domain (TM: Y/N) presence/absence are indicated in respective columns. N/A indicates no annotation from OmicsBox. Proteins that failed annotation by OmicsBox but obtained GO terms from UniProt/InterProScan searches and were still classified.

***L. heterotoma* MSEV protein annotations**				
Lh protein	Class	Annotation	SP: Y/N	TM: Y/N
GAJC01031099.1_25	1	charged multivesicular body protein 6	N	N
GAJC01030581.1_43	1	translocon-associated protein subunit beta	Y	Y
GAJC01030396.1_14	1	mitochondrial pyruvate carrier 1	N	Y
GAJC01030130.1_52	1	UDP-glucuronosyltransferase 1-9-like	Y	Y
GAJC01030096.1_24	1	serine/threonine-protein phosphatase Pgam5, mitochondrial	N	N
GAJC01029414.1_9	1	MICOS complex subunit MIC27 isoform X2	N	N
GAJC01029363.1_47	1	NADH dehydrogenase [ubiquinone] iron-sulfur protein 4, mitochondrial	N	N
GAJC01029325.1_93	1	2-oxoglutarate dehydrogenase, mitochondrial isoform X5	N	N
GAJC01029192.1_25	1	ATP-binding cassette sub-family D member 3	N	Y
GAJC01028989.1_23	1	V-type proton ATPase subunit D	N	N
GAJC01028891.1_55	1	dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	N	N
GAJC01028693.1_7	1	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 5, mitochondrial	N	Y
GAJC01028672.1_28	1	50S ribosome-binding GTPase family protein	Y	N
GAJC01028414.1_19	1	transmembrane and coiled-coil domain-containing protein 1	N	Y
GAJC01028343.1_20	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B isoform X1	N	Y
GAJC01028170.1_18	1	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 10	N	N
GAJC01028123.1_9	1	V-type proton ATPase subunit d	N	N
GAJC01027933.1_33	1	putative OPA3-like protein CG13603	N	N
GAJC01027925.1_11	1	isocitrate dehydrogenase [NAD] subunit beta, mitochondrial	N	N
GAJC01027894.1_58	1	NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial	N	N
GAJC01027824.1_14	1	electron transfer flavoprotein subunit alpha, mitochondrial	N	N
GAJC01027799.1_61	1	vesicle-trafficking protein SEC22b-B	N	Y
GAJC01027753.1_25	1	lipase 3	Y	N
GAJC01027746.1_33	1	50S ribosome-binding GTPase	Y	N
GAJC01027725.1_6	1	ribonuclease Oy	Y	N
GAJC01027531.1_60	1	protein disulfide-isomerase A6	Y	N
GAJC01024852.1_14	1	lysosomal alpha-mannosidase isoform X2	Y	N
GAJC01024725.1_36	1	Trimeric intracellular cation channel type B	N	Y
GAJC01022480.1_6	1	thioredoxin domain-containing protein	Y	Y
GAJC01022382.1_37	1	sideroflexin-1	N	Y

GAJC01021538.1_51	1	trehalase-like	Y	N
GAJC01021129.1_25	1	paramyosin, long form-like	N	N
GAJC01021107.1_29	1	malectin-A	Y	Y
GAJC01020023.1_41	1	sideroflexin-2	N	Y
GAJC01017999.1_17	1	angiopoietin-1 receptor isoform X3	N	Y
GAJC01017428.1_36	1	major royal jelly protein 1	Y	N
GAJC01017322.1_30	1	sarcalumenin isoform X2	Y	N
GAJC01017199.1_29	1	lipase 3-like	Y	N
GAJC01017061.1_11	1	endocuticle structural protein SgAbd-6-like	Y	N
GAJC01016301.1_14	1	60S ribosomal protein L13	N	N
GAJC01013832.1_13	1	60S ribosomal protein L4	N	N
GAJC01013802.1_37	1	NADH-cytochrome b5 reductase 2 isoform X1	N	Y
GAJC01013770.1_23	1	cytochrome b-c1 complex subunit Rieske, mitochondrial	N	N
GAJC01013702.1_11	1	60S acidic ribosomal protein P0	N	N
GAJC01013642.1_14	1	cytochrome P450 4g15-like	N	Y
GAJC01013595.1_37	1	membrane-associated progesterone receptor component 1	N	Y
GAJC01013539.1_20	1	60S ribosomal protein L13a	N	N
GAJC01013495.1_10	1	cytochrome c1, heme protein, mitochondrial	N	Y
GAJC01013427.1_7	1	NADH-quinone oxidoreductase subunit B 2-like	N	N
GAJC01013335.1_5	1	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 7-like	N	N
GAJC01013333.1_21	1	fatty acid-binding protein, muscle isoform X1	N	N
GAJC01013300.1_83	1	protein NipSnap	N	N
GAJC01013291.1_59	1	ATP synthase subunit b, mitochondrial	N	Y
GAJC01013205.1_12	1	60S ribosomal protein L23	N	N
GAJC01013187.1_17	1	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 8	N	N
GAJC01013185.1_14	1	T-complex protein 1 subunit zeta	N	N
GAJC01013182.1_24	1	60S ribosomal protein L5	N	N
GAJC01013160.1_30	1	translocon-associated protein subunit alpha	Y	Y
GAJC01013157.1_73	1	Proline dehydrogenase 1, mitochondrial	N	N
GAJC01013138.1_9	1	myosin regulatory light chain 2	N	N
GAJC01013096.1_16	1	ATP synthase subunit d, mitochondrial	N	N
GAJC01012976.1_11	1	RNA-binding protein squid-like	N	N
GAJC01012952.1_13	1	RNA-binding protein squid isoform X2	N	N
GAJC01012897.1_26	1	heat shock 70 kDa protein cognate 4	N	N
GAJC01012804.1_6	1	cytochrome b5	N	Y
GAJC01012771.1_15	1	glutamine synthetase 2 cytoplasmic isoform X1	N	N
GAJC01012769.1_5	1	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 8, mitochondrial	N	Y
GAJC01012766.1_49	1	annexin B9	N	N
GAJC01012746.1_16	1	ATP synthase subunit g, mitochondrial	N	N
GAJC01012689.1_23	1	14-3-3 protein zeta isoform X1	N	N
GAJC01012610.1_15	1	50S ribosome-binding GTPase	Y	N

GAJC01012565.1_23	1	pyruvate kinase-like isoform X2	N	N
GAJC01012525.1_26	1	50S ribosome-binding GTPase	Y	N
GAJC01012524.1_31	1	50S ribosome-binding GTPase	Y	N
GAJC01012472.1_23	1	ER membrane protein complex subunit 8/9 homolog	N	N
GAJC01012412.1_24	1	actin-5C	N	N
GAJC01012411.1_49	1	actin, muscle	N	N
GAJC01012253.1_8	1	cytochrome c oxidase subunit NDUFA4	N	Y
GAJC01012249.1_18	1	60S ribosomal protein L11	N	N
GAJC01012240.1_36	1	very long-chain-fatty-acid--CoA ligase bubblegum isoform X2	N	N
GAJC01012232.1_17	1	malate dehydrogenase, cytoplasmic	N	N
GAJC01012230.1_7	1	60S ribosomal protein L14	N	N
GAJC01012219.1_12	1	very-long-chain 3-oxoacyl-CoA reductase	N	N
GAJC01012216.1_25	1	60S ribosomal protein L10a	N	N
GAJC01012184.1_54	1	spectrin alpha chain isoform X1	N	N
GAJC01012155.1_9	1	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13	N	Y
GAJC01012150.1_20	1	40S ribosomal protein S15Aa	N	N
GAJC01012137.1_24	1	myosin light chain alkali isoform X2	N	N
GAJC01012129.1_15	1	probable NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12	N	N
GAJC01012103.1_17	1	farnesol dehydrogenase-like	N	N
GAJC01012102.1_28	1	farnesol dehydrogenase-like	N	N
GAJC01012069.1_15	1	40S ribosomal protein S10	N	N
GAJC01012068.1_6	1	cytochrome b-c1 complex subunit 7-like	N	N
GAJC01011949.1_55	1	mitochondrial 2-oxoglutarate/malate carrier protein-like	N	N
GAJC01011927.1_54	1	putative ATP-dependent RNA helicase me31b	N	N
GAJC01011893.1_31	1	cytochrome b5-like	N	Y
GAJC01011807.1_10	1	Protein disulfide-isomerase	N	N
GAJC01011782.1_10	1	protein RER1	N	Y
GAJC01011653.1_5	1	myosin regulatory light chain 2	N	N
GAJC01011619.1_22	1	40S ribosomal protein S2	N	N
GAJC01011605.1_14	1	3-hydroxyacyl-CoA dehydrogenase type-2	N	N
GAJC01011586.1_19	1	leucine-rich repeat, immunoglobulin-like domain and transmembrane domain-containing protein 2	Y	N
GAJC01011558.1_41	1	dihydrolipoyl dehydrogenase, mitochondrial	N	N
GAJC01011550.1_14	1	NADH-ubiquinone oxidoreductase 49 kDa subunit	N	N
GAJC01011454.1_14	1	40S ribosomal protein S7	N	N
GAJC01011419.1_7	1	kielin/chordin-like protein	Y	N
GAJC01011326.1_68	1	MICOS complex subunit Mic60 isoform X1	N	N
GAJC01011277.1_17	1	microsomal glutathione S-transferase 1-like	N	Y
GAJC01011256.1_29	1	peroxiredoxin-5, mitochondrial	N	N
GAJC01011252.1_23	1	40S ribosomal protein S4	N	N
GAJC01011236.1_30	1	60S ribosomal protein L10	N	N

GAJC01011232.1_6	1	MICOS complex subunit MIC19	N	N
GAJC01011192.1_150	1	peroxiredoxin-6-like	N	N
GAJC01011169.1_8	1	NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial	N	N
GAJC01011133.1_10	1	calreticulin	Y	N
GAJC01011105.1_9	1	mitochondrial import inner membrane translocase subunit TIM50-C-like	N	N
GAJC01011050.1_5	1	60S ribosomal protein L36	N	N
GAJC01011037.1_17	1	40S ribosomal protein S8	N	N
GAJC01011036.1_41	1	tubulin beta-1 chain	N	N
GAJC01011006.1_24	1	single-stranded DNA-binding protein, mitochondrial	N	N
GAJC01010984.1_6	1	troponin C	N	N
GAJC01010915.1_9	1	50S ribosome-binding GTPase	Y	N
GAJC01010865.1_16	1	plasminogen activator inhibitor 1 RNA-binding protein isoform X1	N	N
GAJC01010822.1_3	1	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 7	N	N
GAJC01010807.1_12	1	60S ribosomal protein L9	N	N
GAJC01010760.1_13	1	60S ribosomal protein L35a	N	N
GAJC01010666.1_10	1	ATP synthase-coupling factor 6, mitochondrial	N	N
GAJC01010573.1_13	1	very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 2	N	Y
GAJC01010572.1_196	1	protein bark beetle isoform X1	N	Y
GAJC01010488.1_72	1	moesin/ezrin/radixin homolog 1 isoform X2	N	N
GAJC01010481.1_3	1	cytochrome c oxidase subunit 5B, mitochondrial	N	N
GAJC01010465.1_24	1	spermine oxidase-like	Y	N
GAJC01010390.1_9	1	40S ribosomal protein S20	N	N
GAJC01010319.1_30	1	glyceraldehyde-3-phosphate dehydrogenase 2	N	N
GAJC01010280.1_48	1	elongation factor 1-alpha	N	N
GAJC01010278.1_4	1	CDGSH iron-sulfur domain-containing protein 2 homolog	N	N
GAJC01010274.1_32	1	50S ribosome-binding GTPase family protein	Y	N
GAJC01010260.1_26	1	cytochrome b-c1 complex subunit 8	N	N
GAJC01010230.1_54	1	myosin heavy chain, muscle isoform X2	N	N
GAJC01010173.1_113	1	pyruvate carboxylase, mitochondrial isoform X1	N	N
GAJC01010142.1_8	1	60S ribosomal protein L8	N	N
GAJC01010139.1_45	1	ATP synthase subunit e, mitochondrial	N	N
GAJC01010058.1_13	1	protein NPC2 homolog	Y	N
GAJC01009987.1_9	1	icarapin-like	Y	N
GAJC01009972.1_32	1	B-cell receptor-associated protein 31	N	Y
GAJC01009958.1_42	1	histone H2A.V	N	N
GAJC01009924.1_35	1	transmembrane emp24 domain-containing protein bai isoform X2	Y	Y
GAJC01009895.1_5	1	40S ribosomal protein S27	N	N
GAJC01009838.1_20	1	transmembrane protein 214-A	Y	N
GAJC01009770.1_24	1	60S ribosomal protein L24	N	Y
GAJC01009754.1_11	1	apolipophorin-III-like protein	Y	N
GAJC01009730.1_31	1	Tubulin alpha-1 chain	N	N

GAJC01009718.1_4	1	general odorant-binding protein 56d-like	Y	N
GAJC01009624.1_17	1	acyl-CoA Delta(11) desaturase-like	N	Y
GAJC01009588.1_5	1	Y-box factor homolog isoform X1	N	N
GAJC01009568.1_14	1	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 6	N	N
GAJC01009563.1_35	1	adenylate kinase	N	N
GAJC01009552.1_6	1	cytochrome c oxidase subunit 5A, mitochondrial	N	N
GAJC01009529.1_21	1	PDZ and LIM domain protein 3 isoform X2	N	N
GAJC01009527.1_6	1	muscle LIM protein 1 isoform X3	N	N
GAJC01009482.1_8	1	NADH dehydrogenase [ubiquinone] iron-sulfur protein 6, mitochondrial	N	N
GAJC01009424.1_7	1	succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	N	N
GAJC01009366.1_63	1	protein lethal(2)essential for life	N	N
GAJC01009365.1_14	1	60S ribosomal protein L22-like	N	N
GAJC01009336.1_20	1	asparagine synthetase [glutamine-hydrolyzing]	N	N
GAJC01009315.1_44	1	alcohol dehydrogenase class-3	N	N
GAJC01009301.1_5	1	muscle-specific protein 20	N	N
GAJC01009290.1_24	1	60S ribosomal protein L17	N	N
GAJC01009288.1_7	1	ubiquitin-like protein FUBI	N	N
GAJC01009225.1_18	1	sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like	Y	N
GAJC01009190.1_6	1	endoplasmin	Y	N
GAJC01009175.1_28	1	40S ribosomal protein S16	N	N
GAJC01009175.1_13	1	mitochondrial import inner membrane translocase subunit TIM14	N	Y
GAJC01009149.1_50	1	ATP synthase subunit O, mitochondrial	N	N
GAJC01009149.1_29	1	NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial	N	N
GAJC01008186.1_8	1	40S ribosomal protein S15	N	N
GAJC01007074.1_63	1	sodium/potassium-transporting ATPase subunit beta-2	N	Y
GAJC01003958.1_11	1	histone H4-like	N	N
GAJC01003331.1_6	1	60S ribosomal protein L12	N	N
GAJC01002392.1_58	1	electron transfer flavoprotein-ubiquinone oxidoreductase, mitochondrial isoform X1	N	N
GAJC01002381.1_6	1	UPF0389 protein CG9231	N	Y
GAJC01001939.1_7	1	neural/ectodermal development factor IMP-L2 isoform X3	N	N
GAJC01030266.1_67	1	myosin-IA	N	N
GAJC01011635.1_5	1	N/A	N	N
GAJC01020281.1_20	1	troponin T, skeletal muscle isoform X1	N	N
GAJC01022194.1_22	1	tropomyosin-1	N	N
GAJC01009928.1_24	1	tropomyosin isoform X8	N	N
GAJC01010049.1_6	1	PREDICTED: titin-like	Y	N
GAJC01006067.1_15	1	tenascin-R isoform X1	N	N
GAJC01002804.1_4	1	fibronectin type III domain protein	N	Y
GAJC01002803.1_8	1	usherin isoform X1	N	N

GAJC01000705.1_23	1	Down syndrome cell adhesion molecule	N	N
GAJC01012684.1_5	1	vesicle-associated membrane protein-associated protein B	N	Y
GAJC01011832.1_33	1	ras-related protein Rab-1A	N	N
GAJC01029718.1_61	1	ras-related protein Rab-3 isoform X2	N	N
GAJC01027512.1_8	1	ER membrane protein complex subunit 10	N	Y
GAJC01011881.1_45	1	60 kDa heat shock protein, mitochondrial	N	N
GAJC01012692.1_68	1	heat shock 70 kDa protein cognate 3	Y	N
GAJC01010010.1_65	1	protein disulfide-isomerase A3	N	N
GAJC01013393.1_42	1	endoplasmin	N	N
GAJC01010424.1_32	1	calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type isoform X3	N	Y
GAJC01002499.1_10	1	cysteine-rich venom protein 6-like	Y	N
GAJC01010919.1_17	1	cytochrome P450 4g15	N	Y
GAJC01009776.1_111	1	sodium/potassium-transporting ATPase subunit alpha isoform X5	N	Y
GAJC01030706.1_19	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	N	Y
GAJC01010064.1_16	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	N	Y
GAJC01009455.1_18	1	uncharacterized family 31 glucosidase KIAA1161	N	N
GAJC01002819.1_26	1	N/A	N	Y
GAJC01028013.1_45	1	facilitated trehalose transporter Tret1-like isoform X1	N	Y
GAJC01013372.1_11	1	D-arabinitol dehydrogenase 1-like	N	N
GAJC01009176.1_51	1	maltase 1-like	N	N
GAJC01012151.1_60	1	maltase 1-like	Y	Y
GAJC01012521.1_22	1	peroxisomal multifunctional enzyme type 2	N	N
GAJC01011146.1_49	1	maltase 1-like	Y	N
GAJC01009589.1_56	1	apolipophorins	N	Y
GAJC01011361.1_59	1	non-specific lipid-transfer protein	N	N
GAJC01012275.1_45	1	protein l(2)37Cc	N	N
GAJC01011407.1_9	1	prohibitin-2 isoform X1	N	N
GAJC01029025.1_65	1	ATPase family AAA domain-containing protein 3	N	N
GAJC01031145.1_15	1	very long-chain specific acyl-CoA dehydrogenase, mitochondrial	N	N
GAJC01013251.1_45	1	voltage-dependent anion-selective channel-like	N	N
GAJC01011176.1_55	1	ATP synthase subunit beta, mitochondrial	N	N
GAJC01012748.1_61	1	ATP synthase subunit gamma, mitochondrial	N	N
GAJC01009259.1_20	1	ATP synthase subunit alpha, mitochondrial	N	N
GAJC01013821.1_13	1	ADP,ATP carrier protein 2	N	Y
GAJC01028052.1_49	1	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial	N	N
GAJC01022442.1_21	1	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	N	N
GAJC01012046.1_10	1	cytochrome c oxidase subunit 6A1, mitochondrial	N	Y
GAJC01028834.1_45	1	cytochrome b-c1 complex subunit 2, mitochondrial	N	N
GAJC01009913.1_9	1	cytochrome c oxidase subunit 4 isoform 1, mitochondrial-like	N	Y
GAJC01011448.1_11	1	cytochrome c oxidase subunit 6B1	N	N

GAJC01013519.1_8	1	cytochrome c oxidase subunit 6C	N	Y
GAJC01009320.1_17	1	PREDICTED: uncharacterized protein LOC100115623 isoform X1	N	Y
GAJC01011968.1_37	1	succinyl-CoA ligase [ADP/GDP-forming] subunit alpha, mitochondrial	N	N
GAJC01030141.1_27	1	succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	N	N
GAJC01012246.1_84	1	malate dehydrogenase, mitochondrial	N	N
GAJC01030132.1_27	1	glycerol-3-phosphate dehydrogenase, mitochondrial isoform X1	N	N
GAJC01028161.1_116	1	pyruvate dehydrogenase E1 component subunit beta, mitochondrial	N	N
GAJC01012529.1_87	1	probable pyruvate dehydrogenase E1 component subunit alpha, mitochondrial isoform X2	N	N
GAJC01010930.1_14	1	fructose-bisphosphate aldolase-like isoform X1	N	N
GAJC01009649.1_69	1	probable aconitate hydratase, mitochondrial	N	N
GAJC01009326.1_29	1	probable isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	N	N
GAJC01012526.1_17	1	mitochondrial-processing peptidase subunit beta	N	N
GAJC01012254.1_64	1	trifunctional enzyme subunit alpha, mitochondrial	N	N
GAJC01010690.1_152	1	phosphate carrier protein, mitochondrial	N	N
GAJC01013593.1_6	1	mitochondrial hydrogen-transporting ATP synthase coupling factor F	N	Y
GAJC01012002.1_19	1	LETM1 and EF-hand domain-containing protein anon-60Da, mitochondrial	N	Y
GAJC01013635.1_19	1	trifunctional enzyme subunit beta, mitochondrial	N	N
GAJC01013073.1_33	1	60S ribosomal protein L3	N	N
GAJC01013055.1_50	1	60S ribosomal protein L7	N	N
GAJC01012427.1_5	1	60S ribosomal protein L31	N	N
GAJC01011271.1_5	1	40S ribosomal protein S13	N	N
GAJC01011103.1_25	1	40S ribosomal protein S3a	N	N
GAJC01012697.1_5	1	40S ribosomal protein S14	N	N
GAJC01009335.1_27	1	60S ribosomal protein L6	N	N
GAJC01009474.1_50	1	60S ribosomal protein L30	N	N
GAJC01010617.1_11	1	60S ribosomal protein L7a	N	N
GAJC01029468.1_54	2	A disintegrin and metalloproteinase with thrombospondin motifs 8 isoform X1	Y	N
GAJC01028852.1_18	2	enhancing factor (viral)	Y	N
GAJC01027751.1_9	2	rho GTPase-activating protein 9 isoform X1	N	N
GAJC01021576.1_36	2	ATP-binding protein	Y	N
GAJC01020722.1_12	2	neprilysin isoform X1	N	N
GAJC01020720.1_15	2	endothelin-converting enzyme 1 isoform X1	N	N
GAJC01020718.1_27	2	membrane metallo-endopeptidase-like 1 isoform X1	Y	N
GAJC01018250.1_20	2	GTPase RsgA	N	N
GAJC01018015.1_12	2	Endothelin-converting enzyme 1	Y	N
GAJC01013610.1_18	2	endothelin-converting enzyme 1 isoform X1	Y	N
GAJC01012383.1_11	2	lysosomal aspartic protease	Y	N
GAJC01010415.1_14	2	protein Diedel-like	Y	N

GAJC01010126.1_52	2	ras-related protein Rab-11A	N	N
GAJC01030906.1_27	2	myb domain-containing protein	N	N
GAJC01029655.1_38	2	50S ribosome-binding GTPase	Y	Y
GAJC01011181.1_31	2	50S ribosome-binding GTPase family protein	Y	N
GAJC01027332.1_14	2	minor histocompatibility protein HA-1 isoform X2	N	N
GAJC01002026.1_8	2	rac GTPase-activating protein 1-like isoform X3	Y	N
GAJC01027331.1_8	2	rac GTPase-activating protein 1	N	N
GAJC01011013.1_3	2	rac GTPase-activating protein 1-like	N	N
GAJC01024661.1_16	2	Rac GTPase-activating protein 1	N	Y
GAJC01010743.1_26	2	rac GTPase-activating protein 1-like	N	N
GAJC01010744.1_22	2	rac GTPase-activating protein 1-like	N	N
GAJC01030768.1_8	2	N/A	N	N
GAJC01016782.1_12	2	membrane metallo-endopeptidase-like 1	Y	N
GAJC01003564.1_5	2	endothelin-converting enzyme 1-like	N	N
GAJC01028412.1_11	2	membrane metallo-endopeptidase-like 1	N	N
GAJC01000870.1_33	2	endothelin-converting enzyme 1-like	N	N
GAJC01028413.1_13	2	endothelin-converting enzyme 1-like	N	N
GAJC01001743.1_44	2	Membrane metallo-endopeptidase-like 1	Y	N
GAJC01000282.1_4	2	membrane metallo-endopeptidase-like 1	N	N
GAJC01024282.1_10	2	membrane metallo-endopeptidase-like 1	N	N
GAJC01020721.1_11	2	membrane metallo-endopeptidase-like 1	N	Y
GAJC01021959.1_38	2	membrane metallo-endopeptidase-like 1	N	N
GAJC01017489.1_17	2	membrane metallo-endopeptidase-like 1	N	N
GAJC01013609.1_18	2	endothelin-converting enzyme 1 isoform X2	Y	N
GAJC01020719.1_22	2	endothelin-converting enzyme 2-like isoform X1	Y	N
GAJC01020902.1_47	2	membrane metallo-endopeptidase-like 1	N	N
GAJC01028364.1_66	2	AFG3-like protein 2	N	Y
GAJC01022240.1_9	2	protein Diedel-like	N	Y
GAJC01010809.1_70	2	chitinase-like protein Idgf4 isoform X2	Y	N
GAJC01009318.1_8	2	venom allergen 3-like	N	N
GAJC01011813.1_4	2	AMP1_MESCRRecName: Full=Antimicrobial peptide 1; Flags: Precursor	Y	N
GAJC01009485.1_6	2	PREDICTED: uncharacterized protein LOC100119416	N	N
GAJC01009308.1_11	2	50S ribosome-binding GTPase family protein	Y	Y
GAJC01030678.1_22	3	N/A	Y	N
GAJC01029385.1_87	3	Macrophage receptor MARCO	Y	N
GAJC01029271.1_8	3	N/A	N	N
GAJC01028966.1_1	3	N/A	N	N
GAJC01027902.1_6	3	N/A	Y	N
GAJC01027679.1_5	3	N/A	N	N
GAJC01027678.1_12	3	N/A	N	N
GAJC01027452.1_12	3	N/A	N	N

GAJC01027372.1_5	3	N/A	Y	N
GAJC01027336.1_3	3	N/A	Y	N
GAJC01024744.1_6	3	N/A	Y	N
GAJC01024743.1_12	3	N/A	N	N
GAJC01023574.1_5	3	N/A	N	N
GAJC01021659.1_29	3	N/A	Y	N
GAJC01021632.1_11	3	N/A	Y	N
GAJC01017940.1_12	3	N/A	N	Y
GAJC01017287.1_25	3	N/A	N	N
GAJC01017051.1_6	3	N/A	N	N
GAJC01013355.1_10	3	N/A	N	Y
GAJC01012558.1_12	3	N/A	N	N
GAJC01012408.1_7	3	PREDICTED: uncharacterized protein LOC107188865	N	Y
GAJC01011873.1_7	3	N/A	Y	N
GAJC01011863.1_13	3	N/A	Y	N
GAJC01011465.1_5	3	N/A	Y	N
GAJC01011463.1_48	3	N/A	Y	N
GAJC01011457.1_12	3	signal recognition particle receptor subunit alpha-like protein	N	Y
GAJC01011137.1_32	3	N/A	Y	N
GAJC01011097.1_122	3	N/A	Y	N
GAJC01010930.1_16	3	N/A	N	Y
GAJC01010444.1_106	3	hypothetical protein TSAR_010554	N	N
GAJC01010353.1_14	3	N/A	Y	N
GAJC01010250.1_8	3	hypothetical protein SINV_10737, partial	N	N
GAJC01010050.1_21	3	N/A	N	Y
GAJC01009713.1_25	3	N/A	N	N
GAJC01009493.1_4	3	N/A	Y	N
GAJC01009363.1_49	3	uncharacterized protein LOC107274787	Y	N
GAJC01009309.1_26	3	N/A	Y	N
GAJC01007489.1_11	3	N/A	Y	N
GAJC01007380.1_16	3	N/A	N	N
GAJC01006038.1_15	3	N/A	Y	N
GAJC01005406.1_20	3	N/A	N	N
GAJC01003173.1_5	3	N/A	N	N
GAJC01002124.1_43	3	N/A	N	N
GAJC01001115.1_93	3	N/A	N	N
GAJC01000654.1_10	3	N/A	N	N
GAJC01000653.1_18	3	N/A	N	N
GAJC01000527.1_1	3	N/A	N	N
GAJC01000524.1_2	3	N/A	N	N
GAJC01000367.1_14	3	N/A	N	Y

GAJC01000363.1_6	3	N/A	N	N
GAJC01000213.1_10	3	N/A	N	N
GAJC01010373.1_18	3	N/A	N	Y
GAJC01009480.1_7	3	N/A	Y	N
GAJC01021790.1_43	3	hypothetical protein TSAR_006518	N	N
GAJC01030870.1_3	3	N/A	N	Y
GAJC01010745.1_7	3	N/A	N	N
GAJC01006954.1_8	3	N/A	N	Y
GAJC01022409.1_24	3	N/A	N	Y
GAJC01020018.1_10	3	N/A	N	N
GAJC01011966.1_13	3	N/A	Y	N
GAJC01010581.1_4	3	N/A	Y	N
GAJC01000652.1_26	3	N/A	N	Y
GAJC01000651.1_24	3	N/A	N	Y
GAJC01020017.1_13	3	N/A	N	N
GAJC01000093.1_11	3	N/A	Y	Y
GAJC01000213.1_7	3	N/A	N	Y
GAJC01000429.1_10	3	N/A	N	N
GAJC01000520.1_5	3	N/A	N	N
GAJC01000522.1_3	3	N/A	N	Y
GAJC01000878.1_4	3	N/A	N	Y
GAJC01001957.1_22	3	N/A	N	Y
GAJC01002522.1_21	3	N/A	N	N
GAJC01009196.1_11	3	N/A	Y	N
GAJC01009447.1_20	3	N/A	N	N
GAJC01010012.1_3	3	N/A	Y	N
GAJC01010137.1_5	3	N/A	N	N
GAJC01010619.1_3	3	phage tail protein	Y	N
GAJC01010776.1_6	3	N/A	N	Y
GAJC01010848.1_5	3	N/A	N	N
GAJC01011338.1_25	3	N/A	N	N
GAJC01011340.1_11	3	N/A	N	N
GAJC01011342.1_8	3	N/A	N	N
GAJC01011343.1_5	3	N/A	N	N
GAJC01012252.1_5	3	N/A	Y	N
GAJC01011462.1_48	3	N/A	N	Y
GAJC01011994.1_8	3	N/A	Y	N
GAJC01012325.1_25	3	N/A	N	N
GAJC01012567.1_7	3	N/A	N	Y
GAJC01013054.1_50	3	N/A	N	N
GAJC01013214.1_14	3	N/A	Y	N

GAJC01013394.1_5	3	“p40”	Y	Y
GAJC01013472.1_4	3	N/A	N	N
GAJC01013671.1_10	3	N/A	Y	N
GAJC01015272.1_9	3	N/A	N	N
GAJC01015330.1_18	3	N/A	N	Y
GAJC01018220.1_18	3	N/A	N	N
GAJC01018847.1_12	3	N/A	N	N
GAJC01024669.1_3	3	N/A	Y	Y
GAJC01025142.1_10	3	N/A	N	N
GAJC01027680.1_7	3	N/A	N	N
GAJC01028202.1_11	3	N/A	Y	N
GAJC01030220.1_17	3	N/A	Y	N
GAJC01030658.1_4	3	N/A	N	Y
GAJC01012741.1_7	3	N/A	N	N
GAJC01027562.1_9	3	N/A	N	N

Table 3: Comparison of *Lh 14* MSEV and *Lb* VLP proteins. All 407 *Lh 14* MSEV proteins and 383 *Lb G* VLP proteins were compared using NCBI BLAST+. The 115 common proteins are listed here with their class and annotation. The last three columns show the percent identity, e-value, and query coverage for the alignment are also provided. Percent identity (%ID) indicates the percentage of amino acid residues that are exactly the same in the query (*Lb G*) protein to the subject (*Lh 14*) protein. Protein IDs in columns 1 and 4, are in-house identifiers only, whereas identifiers in columns 1 are related to accession numbers in NCBI. N/A indicates that no annotation was found by OmicsBox. Some proteins failed annotation by OmicsBox but obtained GO terms from the UniProt/InterProScan searches and were still classified.

Comparison of *Lh 14* and *Lb G* MSEV/VLP proteins

Lh 14 protein	Class	*Lh 14* Annotation	*Lb G* protein	Class	*Lb G* Annotation	% ID	E-value	Query Cov
GAJC01010230.1_54	1	myosin heavy chain, muscle isoform X2	g8282.t1	1	myosin heavy chain, muscle isoform X1	92.76	0	98
GAJC01009649.1_69	1	probable aconitate hydratase, mitochondrial	g6246.t1	1	aconitate hydratase, mitochondrial-like	96.57	0	100
GAJC01013073.1_33	1	60S ribosomal protein L3	g11248.t1	1	60S ribosomal protein L3	99.76	0	100
GAJC01009474.1_50	1	60S ribosomal protein L30	g7969.t1	1	60S ribosomal protein L30	100.00	2.66E-84	100
GAJC01010280.1_48	1	elongation factor 1-alpha	g7569.t1	1	elongation factor 1-alpha	98.27	0	100
GAJC01009149.1_50	1	ATP synthase subunit O, mitochondrial	g2575.t1	1	ATP synthase subunit O, mitochondrial	92.02	2.67E-125	100
GAJC01013182.1_24	1	60S ribosomal protein L5	g8679.t1	1	60S ribosomal protein L5	95.96	0	100
GAJC01012216.1_25	1	60S ribosomal protein L10a	g249.t1	1	60S ribosomal protein L10a	100.00	1.13E-120	70
GAJC01013635.1_19	1	trifunctional enzyme subunit beta, mitochondrial	g9610.t1	1	trifunctional enzyme subunit beta, mitochondrial	96.36	0	100
GAJC01012521.1_22	1	peroxisomal multifunctional enzyme type 2	g13391.t1	1	peroxisomal multifunctional enzyme type 2	89.52	0	99
GAJC01030581.1_43	1	translocon-associated protein subunit beta	g5494.t1	1	translocon-associated protein subunit beta	94.30	3.64E-139	100
GAJC01027799.1_61	1	vesicle-trafficking protein SEC22b-B	g12983.t1	1	vesicle-trafficking protein SEC22b-B	99.53	3.46E-161	100
GAJC01009175.1_28	1	40S ribosomal protein S16	g354.t1	1	40S ribosomal protein S16	96.69	2.04E-107	100
GAJC01011176.1_55	1	ATP synthase subunit beta, mitochondrial	g8776.t1	1	ATP synthase subunit beta, mitochondrial	97.86	0	100
GAJC01012692.1_68	1	heat shock 70 kDa protein cognate 3	g12490.t1	1	heat shock 70 kDa protein cognate 3 isoform X2	96.04	0	100
GAJC01013821.1_13	1	ADP,ATP carrier protein 2	g8002.t1	1	ADP,ATP carrier protein 2	97.33	0	100
GAJC01012766.1_49	1	annexin B9	g6263.t1	1	annexin B9 isoform X1	34.78	5.03E-09	97
GAJC01012684.1_5	1	vesicle-associated membrane protein-associated protein B	g12939.t1	1	vesicle-associated membrane protein-associated protein B	89.70	6.70E-164	100
GAJC01013138.1_9	1	myosin regulatory light chain 2	g699.t1	1	myosin regulatory light chain 2	97.98	5.10E-139	94
GAJC01010424.1_32	1	calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type isoform X3	g13332.t1	1	calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type isoform X1	98.06	0	98
GAJC01021538.1_51	1	trehalase-like	g6244.t1	1	trehalase-like isoform X1	69.04	0	97
GAJC01012184.1_54	1	spectrin alpha chain isoform X1	g5708.t1	1	spectrin alpha chain isoform X1	25.12	2.21E-48	100
GAJC01027531.1_60	1	protein disulfide-isomerase A6	g5735.t1	1	protein disulfide-isomerase A6	92.69	0	100
GAJC01010465.1_24	1	spermine oxidase-like	g22703.t1	1	spermine oxidase	65.18	1.66E-156	92
GAJC01012254.1_64	1	trifunctional enzyme subunit alpha, mitochondrial	g4533.t1	1	Trifunctional enzyme subunit alpha, mitochondrial	96.90	0	100
GAJC01012246.1_84	1	malate dehydrogenase, mitochondrial	g8573.t1	1	malate dehydrogenase, mitochondrial	97.35	0	100
GAJC01009335.1_27	1	60S ribosomal protein L6	g1298.t1	1	60S ribosomal protein L6	92.99	4.35E-156	100

GAJC01011133.1_10	1	calreticulin	g16716.t1	1	Calreticulin	90.15	2.00E-87	99
GAJC01010760.1_13	1	60S ribosomal protein L35a	g10226.t1	1	60S ribosomal protein L35a	93.42	5.14E-49	100
GAJC01009301.1_5	1	muscle-specific protein 20	g7229.t1	1	muscle-specific protein 20	96.74	8.62E-137	100
GAJC01009924.1_35	1	transmembrane emp24 domain-containing protein bai isoform X2	g9705.t1	1	transmembrane emp24 domain-containing protein bai isoform X2	89.47	6.74E-144	100
GAJC01028834.1_45	1	cytochrome b-c1 complex subunit 2, mitochondrial	g3337.t1	1	cytochrome b-c1 complex subunit 2, mitochondrial	87.53	0	100
GAJC01012150.1_20	1	40S ribosomal protein S15Aa	g2999.t1	1	40S ribosomal protein S15Aa	100.00	6.53E-62	97
GAJC01010064.1_16	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	g13417.t1	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	88.34	0	100
GAJC01012697.1_5	1	40S ribosomal protein S14	g5585.t1	1	40S ribosomal protein S14	100.00	1.34E-110	100
GAJC01012219.1_12	1	very-long-chain 3-oxoacyl-CoA reductase	g4910.t1	1	inactive hydroxysteroid dehydrogenase-like protein 1	35.29	9.41E-25	89
GAJC01011192.1_150	1	peroxiredoxin-6-like	g3260.t1	1	peroxiredoxin	30.64	3.98E-19	67
GAJC01013702.1_11	1	60S acidic ribosomal protein P0	g7087.t1	1	60S acidic ribosomal protein P0	98.32	0	99
GAJC01010010.1_65	1	protein disulfide-isomerase A3	g1259.t1	1	protein disulfide-isomerase A3	87.48	0	99
GAJC01011605.1_14	1	3-hydroxyacyl-CoA dehydrogenase type-2	g10794.t1	1	3-hydroxyacyl-CoA dehydrogenase type-2	88.63	7.70E-174	100
GAJC01013160.1_30	1	translocon-associated protein subunit alpha	g7680.t1	1	translocon-associated protein subunit alpha	91.87	3.48E-154	79
GAJC01027331.1_8	1	rac GTPase-activating protein 1	g7877.t1	1	rac GTPase-activating protein 1-like	34.75	4.11E-09	79
GAJC01012383.1_11	1	lysosomal aspartic protease	g2172.t1	1	lysosomal aspartic protease	91.41	0	99
GAJC01010319.1_30	1	glyceraldehyde-3-phosphate dehydrogenase 2	g10034.t1	1	glyceraldehyde-3-phosphate dehydrogenase 2-like	92.24	0	100
GAJC01010930.1_14	1	fructose-bisphosphate aldolase-like isoform X1	g8740.t1	1	fructose-bisphosphate aldolase-like isoform X1	90.76	0	90
GAJC01012230.1_7	1	60S ribosomal protein L14	g7903.t1	1	60S ribosomal protein L14	80.56	1.89E-63	76
GAJC01011050.1_5	1	60S ribosomal protein L36	g6174.t1	1	60S ribosomal protein L36	100.00	8.75E-81	100
GAJC01010488.1_72	1	moesin/ezrin/radixin homolog 1 isoform X2	g3017.t1	1	moesin/ezrin/radixin homolog 1 isoform X2	96.48	0	99
GAJC01012069.1_15	1	40S ribosomal protein S10	g9570.t1	1	40S ribosomal protein S10	98.73	2.03E-114	100
GAJC01012427.1_5	1	60S ribosomal protein L31	g11607.t1	1	60S ribosomal protein L31	98.70	3.35E-53	77
GAJC01009259.1_20	1	ATP synthase subunit alpha, mitochondrial	g10210.t1	1	ATP synthase subunit alpha, mitochondrial	97.99	0	100
GAJC01013832.1_13	1	60S ribosomal protein L4	g2352.t1	1	60S ribosomal protein L4	96.35	0	100
GAJC01017199.1_29	1	lipase 3-like	g13378.t1	1	lipase 3-like	46.26	3.51E-108	99
GAJC01010690.1_152	1	phosphate carrier protein, mitochondrial	g1748.t1	1	phosphate carrier protein, mitochondrial	93.50	0	100
GAJC01003958.1_11	1	histone H4-like	g13146.t1	1	FChain F, Histone H4	100.00	7.20E-69	95
GAJC01009290.1_24	1	60S ribosomal protein L17	g4524.t1	1	60S ribosomal protein L17	99.46	4.24E-140	100
GAJC01012102.1_28	1	farnesol dehydrogenase-like	g15152.t1	1	farnesol dehydrogenase	69.11	2.54E-133	100
GAJC01010142.1_8	1	60S ribosomal protein L8	g5387.t1	1	60S ribosomal protein L8	97.61	1.91E-117	93
GAJC01011252.1_23	1	40S ribosomal protein S4	g3942.t1	1	40S ribosomal protein S4	99.62	0	100
GAJC01013251.1_45	1	voltage-dependent anion-selective channel-like	g12228.t1	1	voltage-dependent anion-selective channel-like	93.31	0	90
GAJC01010807.1_12	1	60S ribosomal protein L9	g551.t1	1	60S ribosomal protein L9	96.32	1.35E-138	100
GAJC01022442.1_21	1	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	g13106.t1	1	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial	93.96	0	100
GAJC01009288.1_7	1	ubiquitin-like protein FUBI	g729.t1	1	ubiquitin-like protein FUBI	91.67	2.71E-91	100
GAJC01013642.1_14	1	cytochrome P450 4g15-like	g11702.t1	1	cytochrome P450 4g15-like	82.77	0	100
GAJC01012897.1_26	1	heat shock 70 kDa protein cognate 4	g12150.t1	1	heat shock 70 kda protein cognate 4	99.08	0	99
GAJC01012411.1_49	1	actin, muscle	g3124.t1	1	actin-5C	97.07	0	100
GAJC01013539.1_20	1	60S ribosomal protein L13a	g3992.t1	1	60S ribosomal protein L13a	100.00	3.13E-152	100

GAJC01011103.1_25	1	40S ribosomal protein S3a	g11743.t1	1	40S ribosomal protein S3a	99.63	0	100
GAJC01012952.1_13	1	RNA-binding protein squid isoform X2	g10788.t1	1	RNA-binding protein squid-like isoform X8	32.63	2.19E-11	98
GAJC01012976.1_11	1	RNA-binding protein squid-like	g2029.t1	1	RNA-binding protein squid	98.90	0	100
GAJC01009913.1_9	1	cytochrome c oxidase subunit 4 isoform 1, mitochondrial-like	g634.t1	1	Cytochrome c oxidase subunit 4 isoform 1, mitochondrial	83.98	8.11E-120	100
GAJC01009588.1_5	1	Y-box factor homolog isoform X1	g7363.t1	2	nuclease-sensitive element-binding protein 1 isoform X2	96.04	1.93E-134	69
GAJC01021129.1_25	1	paramyosin, long form-like	g13994.t1	1	paramyosin, long form	99.16	0	99
GAJC01009326.1_29	1	probable isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	g3853.t1	1	probable isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	96.27	7.48E-117	100
GAJC01009363.1_49	1	uncharacterized protein LOC107274787	g1247.t1	3	---NA---	54.17	1.75E-54	98
GAJC01011036.1_41	1	tubulin beta-1 chain	g9088.t1	1	Tubulin beta-1 chain	99.77	0	98
GAJC01011277.1_17	1	microsomal glutathione S-transferase 1-like	g5976.t1	1	microsomal glutathione S-transferase 1	87.59	9.85E-100	100
GAJC01030706.1_19	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	g7274.t1	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	91.89	0	100
GAJC01010617.1_11	1	60S ribosomal protein L7a	g6685.t1	1	60S ribosomal protein L7a	99.25	0	100
GAJC01011013.1_3	1	rac GTPase-activating protein 1-like	g16194.t1	1	rac GTPase-activating protein 1	24.50	6.72E-11	86
GAJC01009718.1_4	1	general odorant-binding protein 56d-like	g9910.t1	1	general odorant-binding protein 56d-like	77.44	1.43E-75	100
GAJC01031145.1_15	1	very long-chain specific acyl-CoA dehydrogenase, mitochondrial	g11523.t1	1	very long-chain specific acyl-CoA dehydrogenase, mitochondrial	95.09	0	100
GAJC01011037.1_17	1	40S ribosomal protein S8	g3549.t1	1	40S ribosomal protein S8	99.52	2.57E-155	100
GAJC01010390.1_9	1	40S ribosomal protein S20	g13577.t1	1	40S ribosomal protein S20	100.00	2.43E-87	100
GAJC01020281.1_20	1	troponin T, skeletal muscle isoform X1	g3893.t1	1	troponin T, skeletal muscle isoform X3	90.64	0	89
GAJC01024661.1_16	1	Rac GTPase-activating protein 1	g27718.t1	2	RhoGAP precursor/LbGAP	38.46	1.03E-08	88
GAJC01008186.1_8	1	40S ribosomal protein S15	g5010.t1	1	40S ribosomal protein S15	85.83	3.08E-73	86
GAJC01012689.1_23	1	14-3-3 protein zeta isoform X1	g4000.t1	1	14-3-3 protein zeta isoform X2	94.40	9.44E-89	100
GAJC01010744.1_22	1	rac GTPase-activating protein 1-like	g18394.t1	1	rac GTPase-activating protein 1	31.80	4.03E-25	93
GAJC01013393.1_42	1	endoplasmin	g7824.t1	1	endoplasmin	90.45	0	75
GAJC01011619.1_22	1	40S ribosomal protein S2	g21563.t1	1	40S ribosomal protein S2	99.63	0	99
GAJC01021107.1_29	1	malectin-A	g12151.t1	1	malectin-A	85.14	4.49E-148	94
GAJC01028343.1_20	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B isoform X1	g6087.t1	1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B isoform X1	95.49	0	100
GAJC01009589.1_56	1	apolipophorins	g937.t1	1	N/A	82.39	0	86
GAJC01027725.1_6	1	ribonuclease Oy	g5016.t1	1	ribonuclease Oy	34.97	3.54E-26	93
GAJC01009730.1_31	1	Tubulin alpha-1 chain	g669.t1	1	tubulin alpha-1 chain	100.00	0	100
GAJC01029718.1_61	2	ras-related protein Rab-3 isoform X2	g6700.t1	2	ras-related protein Rab-7a	33.65	2.81E-37	97
GAJC01020902.1_47	2	membrane metallo-endopeptidase-like 1	g25658.t1	2	endothelin-converting enzyme 2-like isoform X1	31.40	5.10E-14	97
GAJC01029468.1_54	2	A disintegrin and metalloproteinase with thrombospondin motifs 8 isoform X1	g8353.t1	2	venom metalloproteinase 3-like	40.41	3.42E-51	100
GAJC01009225.1_18	2	sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like	g15014.t1	2	sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like	33.33	1.77E-22	97
GAJC01028364.1_66	2	AFG3-like protein 2	g8876.t1	1	transitional endoplasmic reticulum ATPase TER94	41.25	5.58E-45	58
GAJC01028852.1_18	2	enhancing factor (viral)	g21977.t1	2	enhancing factor	33.18	4.35E-93	97
GAJC01010126.1_52	2	ras-related protein Rab-11A	g12817.t1	2	ras-related protein Rab-2	49.29	9.58E-65	99
GAJC01011832.1_33	2	ras-related protein Rab-1A	g12041.t1	2	ras-related protein Rab-5C	45.27	4.75E-56	91
GAJC01009318.1_8	2	venom allergen 3-like	g882.t1	2	venom allergen 3-like	51.11	9.75E-58	86
GAJC01009987.1_9	2	icarapin-like	g8451.t1	3	---NA---	81.97	1.60E-32	57

GAJC01020017.1_13	3	---NA---	g21730.t1	3	---NA---	27.72	5.95E-40	59
GAJC01018220.1_18	3	---NA---	g2320.t1	3	putative nuclease HARBI1	35.78	6.10E-49	61
GAJC01029385.1_87	3	Macrophage receptor MARCO	g27649.t1	3	hypothetical protein AVEN_6331_1	35.45	2.36E-25	99
GAJC01002522.1_21	3	---NA---	g15166.t1	3	---NA---	36.98	1.11E-74	88
GAJC01020018.1_10	3	---NA---	g24465.t1	3	---NA---	24.28	4.25E-28	87
GAJC01012558.1_12	3	---NA---	g247.t1	3	---NA---	58.50	0	61
GAJC01017287.1_25	3	---NA---	g23839.t1	3	---NA---	24.64	3.90E-14	65
GAJC01000213.1_7	3	---NA---	g27699.t1	3	---NA---	42.70	1.06E-33	94
GAJC01009447.1_20	3	---NA---	g22025.t1	3	---NA---	35.48	1.00E-21	94

Table 4: Table of the *Lh 14* MSEV proteins that either had no homology or had limited homology to the *Lb* VLP proteins based on the cut-offs (see Methods). Class indicates assignment of protein as Conserved Eukaryotic (Class 1), Infection and Immunity Related (Class 2), or Novel (Class 3). Annotation is protein annotation as provided by OmicsBox. N/A refers to no annotation.

Proteins Unique to *Lh 14* MSEVs		
Protein	Class	Annotation
GAJC01013372.1_11	Class 1	D-arabinitol dehydrogenase 1-like
GAJC01022480.1_6	Class 1	thioredoxin domain-containing protein
GAJC01029414.1_9	Class 1	MICOS complex subunit MIC27 isoform X2
GAJC01012748.1_61	Class 1	ATP synthase subunit gamma, mitochondrial
GAJC01030130.1_52	Class 1	UDP-glucuronosyltransferase 1-9-like
GAJC01027332.1_14	Class 1	minor histocompatibility protein HA-1 isoform X2
GAJC01002026.1_8	Class 1	rac GTPase-activating protein 1-like isoform X3
GAJC01027751.1_9	Class 1	rho GTPase-activating protein 9 isoform X1
GAJC01010743.1_26	Class 1	rac GTPase-activating protein 1-like
GAJC01031099.1_25	Class 1	charged multivesicular body protein 6
GAJC01011097.1_122	Class 1	N/A
GAJC01022382.1_37	Class 1	sideroflexin-1
GAJC01020023.1_41	Class 1	sideroflexin-2
GAJC01027753.1_25	Class 1	lipase 3
GAJC01009624.1_17	Class 1	acyl-CoA Delta(11) desaturase-like
GAJC01030141.1_27	Class 1	succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial
GAJC01009424.1_7	Class 1	succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial
GAJC01009315.1_44	Class 1	alcohol dehydrogenase class-3
GAJC01010809.1_70	Class 1	chitinase-like protein Idgf4 isoform X2
GAJC01012412.1_24	Class 1	actin-5C
GAJC01009527.1_6	Class 1	muscle LIM protein 1 isoform X3
GAJC01012103.1_17	Class 1	farnesol dehydrogenase-like
GAJC01028672.1_28	Class 1	50S ribosome-binding GTPase family protein
GAJC01012525.1_26	Class 1	50S ribosome-binding GTPase
GAJC01013519.1_8	Class 1	cytochrome c oxidase subunit 6C
GAJC01028170.1_18	Class 1	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 10
GAJC01030906.1_27	Class 1	myb domain-containing protein
GAJC01010058.1_13	Class 1	protein NPC2 homolog
GAJC01012804.1_6	Class 1	cytochrome b5
GAJC01011893.1_31	Class 1	cytochrome b5-like

GAJC01013595.1_37	Class 1	membrane-associated progesterone receptor component 1
GAJC01011949.1_55	Class 1	mitochondrial 2-oxoglutarate/malate carrier protein-like
GAJC01009838.1_20	Class 1	transmembrane protein 214-A
GAJC01009175.1_13	Class 1	mitochondrial import inner membrane translocase subunit TIM14
GAJC01028693.1_7	Class 1	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 5, mitochondrial
GAJC01009568.1_14	Class 1	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 6
GAJC01030096.1_24	Class 1	serine/threonine-protein phosphatase Pgam5, mitochondrial
GAJC01013593.1_6	Class 1	mitochondrial hydrogen-transporting ATP synthase coupling factor F
GAJC01012408.1_7	Class 1	PREDICTED: uncharacterized protein LOC107188865
GAJC01012155.1_9	Class 1	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13
GAJC01013055.1_50	Class 1	60S ribosomal protein L7
GAJC01009770.1_24	Class 1	60S ribosomal protein L24
GAJC01010573.1_13	Class 1	very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 2
GAJC01012253.1_8	Class 1	cytochrome c oxidase subunit NDUFA4
GAJC01009482.1_8	Class 1	NADH dehydrogenase [ubiquinone] iron-sulfur protein 6, mitochondrial
GAJC01027933.1_33	Class 1	putative OPA3-like protein CG13603
GAJC01027512.1_8	Class 1	ER membrane protein complex subunit 10
GAJC01022194.1_22	Class 1	tropomyosin-1
GAJC01021790.1_43	Class 1	hypothetical protein TSAR_006518
GAJC01017428.1_36	Class 1	major royal jelly protein 1
GAJC01013300.1_83	Class 1	protein NipSnap
GAJC01012472.1_23	Class 1	ER membrane protein complex subunit 8/9 homolog
GAJC01011326.1_68	Class 1	MICOS complex subunit Mic60 isoform X1
GAJC01011232.1_6	Class 1	MICOS complex subunit MIC19
GAJC01010865.1_16	Class 1	plasminogen activator inhibitor 1 RNA-binding protein isoform X1
GAJC01010250.1_8	Class 1	hypothetical protein SINV_10737, partial
GAJC01009928.1_24	Class 1	tropomyosin isoform X8
GAJC01009529.1_21	Class 1	PDZ and LIM domain protein 3 isoform X2
GAJC01009366.1_63	Class 1	protein lethal(2)essential for life
GAJC01002392.1_58	Class 1	electron transfer flavoprotein-ubiquinone oxidoreductase, mitochondrial isoform X1
GAJC01001939.1_7	Class 1	neural/ectodermal development factor IMP-L2 isoform X3
GAJC01028989.1_23	Class 1	V-type proton ATPase subunit D
GAJC01017061.1_11	Class 1	endocuticle structural protein SgAbd-6-like
GAJC01029363.1_47	Class 1	NADH dehydrogenase [ubiquinone] iron-sulfur protein 4, mitochondrial
GAJC01011550.1_14	Class 1	NADH-ubiquinone oxidoreductase 49 kDa subunit
GAJC01011169.1_8	Class 1	NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial
GAJC01013802.1_37	Class 1	NADH-cytochrome b5 reductase 2 isoform X1

GAJC01011256.1_29	Class 1	peroxiredoxin-5, mitochondrial
GAJC01028123.1_9	Class 1	V-type proton ATPase subunit d
GAJC01027894.1_58	Class 1	NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial
GAJC01027824.1_14	Class 1	electron transfer flavoprotein subunit alpha, mitochondrial
GAJC01013495.1_10	Class 1	cytochrome c1, heme protein, mitochondrial
GAJC01013333.1_21	Class 1	fatty acid-binding protein, muscle isoform X1
GAJC01013427.1_7	Class 1	NADH-quinone oxidoreductase subunit B 2-like
GAJC01009149.1_29	Class 1	NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial
GAJC01012129.1_15	Class 1	probable NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12
GAJC01013770.1_23	Class 1	cytochrome b-c1 complex subunit Rieske, mitochondrial
GAJC01010260.1_26	Class 1	cytochrome b-c1 complex subunit 8
GAJC01009190.1_6	Class 1	endoplasmin
GAJC01013185.1_14	Class 1	T-complex protein 1 subunit zeta
GAJC01029192.1_25	Class 1	ATP-binding cassette sub-family D member 3
GAJC01029025.1_65	Class 1	ATPase family AAA domain-containing protein 3
GAJC01011881.1_45	Class 1	60 kDa heat shock protein, mitochondrial
GAJC01011586.1_19	Class 1	leucine-rich repeat, immunoglobulin-like domain and transmembrane domain-containing protein 2
GAJC01010049.1_6	Class 1	PREDICTED: titin-like
GAJC01002804.1_4	Class 1	fibronectin type III domain protein
GAJC01002803.1_8	Class 1	usherin isoform X1
GAJC01000705.1_23	Class 1	Down syndrome cell adhesion molecule
GAJC01017999.1_17	Class 1	angiopoietin-1 receptor isoform X3
GAJC01006067.1_15	Class 1	tenascin-R isoform X1
GAJC01012002.1_19	Class 1	LETM1 and EF-hand domain-containing protein anon-60Da, mitochondrial
GAJC01010984.1_6	Class 1	troponin C
GAJC01012137.1_24	Class 1	myosin light chain alkali isoform X2
GAJC01011653.1_5	Class 1	myosin regulatory light chain 2
GAJC01010919.1_17	Class 1	cytochrome P450 4g15
GAJC01028414.1_19	Class 1	transmembrane and coiled-coil domain-containing protein 1
GAJC01024725.1_36	Class 1	Trimeric intracellular cation channel type B
GAJC01012529.1_87	Class 1	probable pyruvate dehydrogenase E1 component subunit alpha, mitochondrial isoform X2
GAJC01013157.1_73	Class 1	Proline dehydrogenase 1, mitochondrial
GAJC01029325.1_93	Class 1	2-oxoglutarate dehydrogenase, mitochondrial isoform X5
GAJC01009455.1_18	Class 1	uncharacterized family 31 glucosidase KIAA1161
GAJC01027925.1_11	Class 1	isocitrate dehydrogenase [NAD] subunit beta, mitochondrial
GAJC01030132.1_27	Class 1	glycerol-3-phosphate dehydrogenase, mitochondrial isoform X1
GAJC01028891.1_55	Class 1	dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial
GAJC01011558.1_41	Class 1	dihydrolipoyl dehydrogenase, mitochondrial

GAJC01012046.1_10	Class 1	cytochrome c oxidase subunit 6A1, mitochondrial
GAJC01009552.1_6	Class 1	cytochrome c oxidase subunit 5A, mitochondrial
GAJC01010481.1_3	Class 1	cytochrome c oxidase subunit 5B, mitochondrial
GAJC01009320.1_17	Class 1	PREDICTED: uncharacterized protein LOC100115623 isoform X1
GAJC01009336.1_20	Class 1	asparagine synthetase [glutamine-hydrolyzing]
GAJC01009563.1_35	Class 1	adenylate kinase
GAJC01012769.1_5	Class 1	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 8, mitochondrial
GAJC01010822.1_3	Class 1	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 7
GAJC01011968.1_37	Class 1	succinyl-CoA ligase [ADP/GDP-forming] subunit alpha, mitochondrial
GAJC01011361.1_59	Class 1	non-specific lipid-transfer protein
GAJC01012240.1_36	Class 1	very long-chain-fatty-acid--CoA ligase bubblegum isoform X2
GAJC01028161.1_116	Class 1	pyruvate dehydrogenase E1 component subunit beta, mitochondrial
GAJC01012232.1_17	Class 1	malate dehydrogenase, cytoplasmic
GAJC01012151.1_60	Class 1	maltase 1-like
GAJC01011146.1_49	Class 1	maltase 1-like
GAJC01009176.1_51	Class 1	maltase 1-like
GAJC01010173.1_113	Class 1	pyruvate carboxylase, mitochondrial isoform X1
GAJC01024852.1_14	Class 1	lysosomal alpha-mannosidase isoform X2
GAJC01012771.1_15	Class 1	glutamine synthetase 2 cytoplasmic isoform X1
GAJC01012526.1_17	Class 1	mitochondrial-processing peptidase subunit beta
GAJC01030266.1_67	Class 1	myosin-IA
GAJC01011807.1_10	Class 1	Protein disulfide-isomerase
GAJC01013205.1_12	Class 1	60S ribosomal protein L23
GAJC01012249.1_18	Class 1	60S ribosomal protein L11
GAJC01003331.1_6	Class 1	60S ribosomal protein L12
GAJC01016301.1_14	Class 1	60S ribosomal protein L13
GAJC01009895.1_5	Class 1	40S ribosomal protein S27
GAJC01009365.1_14	Class 1	60S ribosomal protein L22-like
GAJC01011271.1_5	Class 1	40S ribosomal protein S13
GAJC01011454.1_14	Class 1	40S ribosomal protein S7
GAJC01011236.1_30	Class 1	60S ribosomal protein L10
GAJC01011006.1_24	Class 1	single-stranded DNA-binding protein, mitochondrial
GAJC01011927.1_54	Class 1	putative ATP-dependent RNA helicase me31b
GAJC01012565.1_23	Class 1	pyruvate kinase-like isoform X2
GAJC01009776.1_111	Class 1	sodium/potassium-transporting ATPase subunit alpha isoform X5
GAJC01010278.1_4	Class 1	CDGSH iron-sulfur domain-containing protein 2 homolog
GAJC01011782.1_10	Class 1	protein RER1
GAJC01028013.1_45	Class 1	facilitated trehalose transporter Tret1-like isoform X1

GAJC01012275.1_45	Class 1	protein l(2)37Cc
GAJC01011407.1_9	Class 1	prohibitin-2 isoform X1
GAJC01007074.1_63	Class 1	sodium/potassium-transporting ATPase subunit beta-2
GAJC01009972.1_32	Class 1	B-cell receptor-associated protein 31
GAJC01012068.1_6	Class 1	cytochrome b-c1 complex subunit 7-like
GAJC01028052.1_49	Class 1	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial
GAJC01013187.1_17	Class 1	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 8
GAJC01030396.1_14	Class 1	mitochondrial pyruvate carrier 1
GAJC01013335.1_5	Class 1	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 7-like
GAJC01011105.1_9	Class 1	mitochondrial import inner membrane translocase subunit TIM50-C-like
GAJC01011448.1_11	Class 1	cytochrome c oxidase subunit 6B1
GAJC01009754.1_11	Class 1	apolipophorin-III-like protein
GAJC01009958.1_42	Class 1	histone H2A.V
GAJC01013096.1_16	Class 1	ATP synthase subunit d, mitochondrial
GAJC01012746.1_16	Class 1	ATP synthase subunit g, mitochondrial
GAJC01013291.1_59	Class 1	ATP synthase subunit b, mitochondrial
GAJC01010666.1_10	Class 1	ATP synthase-coupling factor 6, mitochondrial
GAJC01010139.1_45	Class 1	ATP synthase subunit e, mitochondrial
GAJC01011419.1_7	Class 1	kielin/chordin-like protein
GAJC01018015.1_12	Class 2	Endothelin-converting enzyme 1
GAJC01018250.1_20	Class 2	GTPase RsgA
GAJC01010415.1_14	Class 2	protein Diedel-like
GAJC01002499.1_10	Class 2	cysteine-rich venom protein 6-like
GAJC01010274.1_32	Class 2	50S ribosome-binding GTPase family protein
GAJC01012524.1_31	Class 2	50S ribosome-binding GTPase
GAJC01012610.1_15	Class 2	50S ribosome-binding GTPase
GAJC01022240.1_9	Class 2	protein Diedel-like
GAJC01011813.1_4	Class 2	AMP1_MESCRRecName: Full=Antimicrobial peptide 1; Flags: Precursor
GAJC01029655.1_38	Class 2	50S ribosome-binding GTPase
GAJC01027746.1_33	Class 2	50S ribosome-binding GTPase
GAJC01021576.1_36	Class 2	ATP-binding protein
GAJC01017322.1_30	Class 2	sarcalumenin isoform X2
GAJC01011181.1_31	Class 2	50S ribosome-binding GTPase family protein
GAJC01010915.1_9	Class 2	50S ribosome-binding GTPase
GAJC01009308.1_11	Class 2	50S ribosome-binding GTPase family protein
GAJC01010572.1_196	Class 2	protein bark beetle isoform X1
GAJC01016782.1_12	Class 2	membrane metallo-endopeptidase-like 1
GAJC01020722.1_12	Class 2	neprilysin isoform X1
GAJC01013609.1_18	Class 2	endothelin-converting enzyme 1 isoform X2

GAJC01028413.1_13	Class 2	endothelin-converting enzyme 1-like
GAJC01028412.1_11	Class 2	membrane metallo-endopeptidase-like 1
GAJC01024282.1_10	Class 2	membrane metallo-endopeptidase-like 1
GAJC01021959.1_38	Class 2	membrane metallo-endopeptidase-like 1
GAJC01020721.1_11	Class 2	membrane metallo-endopeptidase-like 1
GAJC01020720.1_15	Class 2	endothelin-converting enzyme 1 isoform X1
GAJC01020718.1_27	Class 2	membrane metallo-endopeptidase-like 1 isoform X1
GAJC01017489.1_17	Class 2	membrane metallo-endopeptidase-like 1
GAJC01013610.1_18	Class 2	endothelin-converting enzyme 1 isoform X1
GAJC01003564.1_5	Class 2	endothelin-converting enzyme 1-like
GAJC01001743.1_44	Class 2	Membrane metallo-endopeptidase-like 1
GAJC01000870.1_33	Class 2	endothelin-converting enzyme 1-like
GAJC01000282.1_4	Class 2	membrane metallo-endopeptidase-like 1
GAJC01020719.1_22	Class 2	endothelin-converting enzyme 2-like isoform X1
GAJC01000093.1_11	Class 3	N/A
GAJC01030678.1_22	Class 3	N/A
GAJC01030658.1_4	Class 3	N/A
GAJC01030220.1_17	Class 3	N/A
GAJC01029271.1_8	Class 3	N/A
GAJC01028966.1_1	Class 3	N/A
GAJC01028202.1_11	Class 3	N/A
GAJC01027902.1_6	Class 3	N/A
GAJC01027680.1_7	Class 3	N/A
GAJC01027679.1_5	Class 3	N/A
GAJC01027678.1_12	Class 3	N/A
GAJC01027452.1_12	Class 3	N/A
GAJC01027372.1_5	Class 3	N/A
GAJC01027336.1_3	Class 3	N/A
GAJC01025142.1_10	Class 3	N/A
GAJC01024744.1_6	Class 3	N/A
GAJC01024743.1_12	Class 3	N/A
GAJC01024669.1_3	Class 3	N/A
GAJC01023574.1_5	Class 3	N/A
GAJC01022409.1_24	Class 3	N/A
GAJC01021659.1_29	Class 3	N/A
GAJC01021632.1_11	Class 3	N/A
GAJC01018847.1_12	Class 3	N/A
GAJC01017940.1_12	Class 3	N/A
GAJC01017051.1_6	Class 3	N/A
GAJC01015272.1_9	Class 3	N/A
GAJC01013472.1_4	Class 3	N/A

GAJC01013394.1_5	Class 3	“p40”
GAJC01013355.1_10	Class 3	N/A
GAJC01012325.1_25	Class 3	N/A
GAJC01011994.1_8	Class 3	N/A
GAJC01011465.1_5	Class 3	N/A
GAJC01011340.1_11	Class 3	N/A
GAJC01011137.1_32	Class 3	N/A
GAJC01010930.1_16	Class 3	N/A
GAJC01010848.1_5	Class 3	N/A
GAJC01010776.1_6	Class 3	N/A
GAJC01010745.1_7	Class 3	N/A
GAJC01010581.1_4	Class 3	N/A
GAJC01010373.1_18	Class 3	N/A
GAJC01010050.1_21	Class 3	N/A
GAJC01007489.1_11	Class 3	N/A
GAJC01007380.1_16	Class 3	N/A
GAJC01006954.1_8	Class 3	N/A
GAJC01006038.1_15	Class 3	N/A
GAJC01003173.1_5	Class 3	N/A
GAJC01002819.1_26	Class 3	N/A
GAJC01001957.1_22	Class 3	N/A
GAJC01001115.1_93	Class 3	N/A
GAJC01000878.1_4	Class 3	N/A
GAJC01000654.1_10	Class 3	N/A
GAJC01000653.1_18	Class 3	N/A
GAJC01000652.1_26	Class 3	N/A
GAJC01000651.1_24	Class 3	N/A
GAJC01000527.1_1	Class 3	N/A
GAJC01000524.1_2	Class 3	N/A
GAJC01000522.1_3	Class 3	N/A
GAJC01000520.1_5	Class 3	N/A
GAJC01000429.1_10	Class 3	N/A
GAJC01000367.1_14	Class 3	N/A
GAJC01000363.1_6	Class 3	N/A
GAJC01000213.1_10	Class 3	N/A
GAJC01010619.1_3	Class 3	phage tail protein
GAJC01002381.1_6	Class 3	UPF0389 protein CG9231
GAJC01030870.1_3	Class 3	N/A
GAJC01030768.1_8	Class 3	N/A
GAJC01015330.1_18	Class 3	N/A
GAJC01013671.1_10	Class 3	N/A

GAJC01013214.1_14	Class 3	N/A
GAJC01013054.1_50	Class 3	N/A
GAJC01012567.1_7	Class 3	N/A
GAJC01012252.1_5	Class 3	N/A
GAJC01011966.1_13	Class 3	N/A
GAJC01011873.1_7	Class 3	N/A
GAJC01011863.1_13	Class 3	N/A
GAJC01011635.1_5	Class 3	N/A
GAJC01011463.1_48	Class 3	N/A
GAJC01011462.1_48	Class 3	N/A
GAJC01011343.1_5	Class 3	N/A
GAJC01011342.1_8	Class 3	N/A
GAJC01011338.1_25	Class 3	N/A
GAJC01010353.1_14	Class 3	N/A
GAJC01010137.1_5	Class 3	N/A
GAJC01010012.1_3	Class 3	N/A
GAJC01009713.1_25	Class 3	N/A
GAJC01009493.1_4	Class 3	N/A
GAJC01009480.1_7	Class 3	N/A
GAJC01009309.1_26	Class 3	N/A
GAJC01009196.1_11	Class 3	N/A
GAJC01005406.1_20	Class 3	N/A
GAJC01002124.1_43	Class 3	N/A
GAJC01010444.1_106	Class 3	hypothetical protein TSAR_010554
GAJC01009485.1_6	Class 3	PREDICTED: uncharacterized protein LOC100119416
GAJC01011457.1_12	Class 3	signal recognition particle receptor subunit alpha-like protein
GAJC01012741.1_7	Class 3	N/A
GAJC01027562.1_9	Class 3	N/A

Table 5: Table of the *Lb G* VLP proteins that either had no homology or had limited homology to *Lh 14* MSEV proteins, as defined in the Methods. Equivalent Accession column provides accession numbers for the available *Lb 17* transcripts which, when translated, match to analyzed protein sequence according to the criteria specified in the Methods. NF refers to absence of equivalent accession number within available transcriptomes.

Protein Unique to *Lb G* VLPs			
Protein	Class	Annotation	Equivalent Accession numbers for *Lb 17* transcripts
g1465.t1	Class 1	short-chain specific acyl-CoA dehydrogenase, mitochondrial	GITC01170655,GISX01120676,GAJA01001371.1
g24769.t1	Class 1	protein 5NUC-like	GITC01099956,GAJA01005448.1
g11746.t1	Class 1	putative fatty acyl-CoA reductase CG8306	GITC01065215,GISX01090367,GAJA01007151.1
g1069.t1	Class 1	probable citrate synthase 2, mitochondrial	GITC01026697,GAJA01020546.1
g408.t1	Class 1	dnaJ homolog subfamily A member 1	GITC01053845,GAJA01012937.1
g19132.t1	Class 1	arginine kinase	GITC01022209,GAJA01020556.1
g11592.t1	Class 1	protein disulfide-isomerase	GITC01160504,GISX01095206,GAJA01018442.1
g2224.t1	Class 1	transketolase-like protein 2 isoform X2	GITC01119652,GISX01102650,GAJA01019937.1
g3443.t1	Class 1	annexin B9	GITC01010951,GISX01061896,GAJA01019905.1
g14011.t1	Class 1	aldehyde dehydrogenase, mitochondrial	GITC01048965,GAJA01019471.1
g15500.t1	Class 1	sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like	GITC01026657,GAJA01006995.1
g5201.t1	Class 1	catalase isoform X1	GITC01051492,GISX01143135,GAJA01020919.1
g14952.t1	Class 1	Protein Dok-7	GITC01200504,GAJA01005216.1
g12611.t1	Class 1	renin receptor isoform X2	GITC01150552,GAJA01005954.1
g5500.t1	Class 1	acetyl-coenzyme A transporter 1	GITC01199066,GAJA01000834.1
g26546.t1	Class 1	alaserpin isoform X5	GITC01198069,GAJA01019393.1
g5600.t1	Class 1	peptidyl-alpha-hydroxyglycine alpha-amidating lyase 1	GITC01103963,GAJA01016066.1
g14212.t1	Class 1	3-ketoacyl-CoA thiolase, mitochondrial	GITC01033377,GISX01158239,GAJA01006839.1
g3358.t1	Class 1	beta-mannosidase	GITC01094070,GISX01102084,GAJA01002610.1
g469.t1	Class 1	protein lethal(2)essential for life-like	GITC01052564,GISX01038005,GAJA01019474.1
g1869.t1	Class 1	multidrug resistance protein 1A	GITC01110448,GISX01006229,GAJA01011165.1
g3507.t1	Class 1	peptidylglycine alpha-hydroxylating monooxygenase	GITC01023314,GISX01010886,GAJA01015755.1
g11675.t1	Class 1	sulfhydryl oxidase 2-like	GITC01158188,GISX01007172,GAJA01014272.1
g15893.t1	Class 1	protein transport protein Sec24C	GITC01133363,GISX01111185,GAJA01007520.1
g13184.t1	Class 1	alpha-tocopherol transfer protein-like	GITC01160980,GISX01010018,GAJA01015109.1
g18935.t1	Class 1	dnaJ homolog shv	GITC01044410,GAJA01015175.1
g5307.t1	Class 1	proteasomal ubiquitin receptor ADRM1 homolog isoform X2	GITC01022665,GAJA01019012.1
g10473.t1	Class 1	60S ribosomal protein L27a	GITC01093705,GAJA01019244.1
g8001.t1	Class 1	vesicle transport protein GOT1B	GITC01039099,GAJA01006852.1
g20934.t1	Class 1	lipase 3-like	GITC01045500,GISX01029643,GAJA01018823.1

g5997.t1	Class 1	juvenile hormone epoxide hydrolase 1-like	GITC01204935,GISX01057930,GAJA01006104.1
g12024.t1	Class 1	probable beta-hexosaminidase fdl isoform X1	GITC01046075,GISX01007846,GAJA01017821.1
g4353.t1	Class 1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	GAJA01006801.1
g11414.t1	Class 1	Hermansky-Pudlak syndrome 4 protein	GITC01086439,GISX01052876,GAJA01002104.1
g26437.t1	Class 1	histone H4-like	GITC01147019,GISX01052793,GAJA01018362.1
g5514.t1	Class 1	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	GITC01138609,GISX01034114,GAJA01006925.1
g7129.t1	Class 1	aldehyde dehydrogenase, dimeric NADP-preferring isoform X2	GITC01166397,GAJA01002685.1
g3059.t1	Class 1	protein wntless	GITC01000968,GISX01132444,GAJA01006542.1
g1756.t1	Class 1	sequestosome-1 isoform X1	GISX01052120,GAJA01018105.1
g21089.t1	Class 1	alaserpin isoform X5	GITC01198069,GAJA01020187.1
g3257.t1	Class 1	endoplasmic reticulum-Golgi intermediate compartment protein 3	GITC01108972,GAJA01007709.1
g332.t1	Class 1	procollagen-lysine,2-oxoglutarate 5-dioxygenase 1 isoform X1	GITC01002380,GISX01038304,GAJA01019208.1
g16717.t1	Class 1	Calreticulin	GITC01002549,GAJA01020312.1
g92.t1	Class 1	40S ribosomal protein S23	GITC01173041,GAJA01019291.1
g4759.t1	Class 1	putative beta-carotene-binding protein	GITC01034596,GISX01157145,GAJA01011141.1
g3616.t1	Class 1	probable cytochrome P450 6a13	NF
g3726.t1	Class 1	translocon-associated protein subunit delta	GITC01003513,GISX01126147,GAJA01019639.1
g4237.t1	Class 1	alkaline phosphatase	GITC01145350,GISX01050580,GAJA01014852.1
g8285.t1	Class 1	Myosin heavy chain, muscle	GITC01115643,GAJA01019881.1
g13472.t1	Class 1	uncharacterized protein LOC105834636	GITC01202287,GAJA01011613.1
g7864.t1	Class 1	Succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial	GITC01053277,GAJA01020185.1
g8412.t1	Class 1	40S ribosomal protein S3	GITC01163517,GISX01053088,GAJA01018456.1
g12533.t1	Class 1	V-type proton ATPase 116 kDa subunit a isoform X2	GITC01063707,GAJA01016358.1
g12977.t1	Class 1	elongation factor 1-beta'	NF
g732.t1	Class 1	signal peptidase complex subunit 1	GAJA01006328.1
g4242.t1	Class 1	ABC transporter G family member 20 isoform X1	GITC01142795,GISX01021523,GAJA01014091.1
g15809.t1	Class 1	putative methyltransferase TARBP1	GITC01139736,GAJA01007968.1
g9327.t1	Class 1	60S acidic ribosomal protein P2	GITC01102488,GAJA01020490.1
g329.t1	Class 1	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	GAJA01019207.1
g9718.t1	Class 1	60S ribosomal protein L23a	GITC01095446,GAJA01019658.1
g12336.t1	Class 1	protein 60A	GITC01022418,GISX01157710,GAJA01004651.1
g11435.t1	Class 1	DNA repair protein complementing XP-C cells homolog	NF
g658.t1	Class 1	peptidyl-prolyl cis-trans isomerase 5	GITC01176579,GISX01012484,GAJA01018693.1
g13251.t1	Class 1	translocation protein SEC63 homolog	GITC01147975,GAJA01021337.1
g7654.t1	Class 1	mesencephalic astrocyte-derived neurotrophic factor homolog	GITC01166479,GAJA01006568.1
g2193.t1	Class 1	V-type proton ATPase subunit B	GITC01062615,GISX01091541,GAJA01021129.1
g8216.t1	Class 1	surfeit locus protein 4 homolog	GITC01019546,GAJA01021123.1
g8500.t1	Class 1	coatomer subunit alpha	GITC01070530,GAJA01020686.1
g12892.t1	Class 1	adenylate cyclase type 10-like	NF

g14686.t1	Class 1	vesicular integral-membrane protein VIP36	GITC01175381,GISX01089820,GAJA01016135.1
g8915.t1	Class 1	40S ribosomal protein S12	GITC01103797,GISX01123001,GAJA01018094.1
g3791.t1	Class 1	fasciclin-2 isoform X2	GITC01164621,GISX01017607,GAJA01019952.1
g21501.t1	Class 1	lysosomal alpha-mannosidase isoform X1	GITC01143013,GISX01151660,GAJA01012944.1
g15153.t1	Class 1	sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like	GITC01031891,GISX01058185,GAJA01008402.1
g11874.t1	Class 1	protein YIPF1	GAJA01018743.1
g12308.t1	Class 1	coatomer subunit gamma	GITC01143614,GAJA01008807.1
g4245.t1	Class 1	protein transport protein Sec61 subunit beta	GITC01183753,GISX01013289,GAJA01021066.1
g9652.t1	Class 1	coatomer subunit beta	GITC01184187,GAJA01007247.1
g7099.t1	Class 1	Golgi reassembly-stacking protein 2	GITC01155695,GAJA01000385.1
g5466.t1	Class 1	signal peptidase complex catalytic subunit SEC11A	GITC01176630,GISX01128699,GAJA01007462.1
g11920.t1	Class 1	mannosyl-oligosaccharide glucosidase GCS1	GITC01094274,GISX01078827,GAJA01010181.1
g18649.t1	Class 1	40S ribosomal protein S19-like	GITC01170878,GAJA01018587.1
g12970.t1	Class 1	coatomer subunit zeta-1	GITC01055847,GISX01005749,GAJA01018348.1
g8281.t1	Class 1	myosin heavy chain, muscle isoform X4	GITC01019235,GISX01138967,GAJA01019882.1
g10826.t1	Class 1	heat shock 70 kDa protein cognate 5	GITC01143233,GISX01081150,GAJA01019253.1
g22586.t1	Class 1	lysosomal alpha-mannosidase isoform X1	GITC01016947,GISX01151651,GAJA01012944.1
g186.t1	Class 1	superoxide dismutase [Cu-Zn]	GITC01167390,GISX01040095,GAJA01018817.1
g15224.t1	Class 1	Polyadenylate-binding protein 1	GITC01140487,GISX01158596,GAJA01018206.1
g16252.t1	Class 1	pol polyprotein	GITC01062974
g12249.t1	Class 1	heat shock protein 83	GITC01034157,GAJA01020548.1
g8711.t1	Class 1	translocon-associated protein subunit gamma	GAJA01020036.1
g10248.t1	Class 1	FK506-binding protein 2 isoform X2	GITC01095223,GAJA01020242.1
g14457.t1	Class 1	CD63 antigen	GITC01050580,GISX01143362,GAJA01007761.1
g11819.t1	Class 1	cofilin/actin-depolymerizing factor homolog	GITC01039102,GISX01132084,GAJA01019586.1
g1500.t1	Class 1	40S ribosomal protein SA	GAJA01020035.1
g8024.t1	Class 1	aminoacylase-1	GITC01150154,GISX01072045,GAJA01015784.1
g16496.t1	Class 1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit DAD1	GAJA01020840.1
g5064.t1	Class 1	enolase isoform X2	GAJA01020476.1
g11203.t1	Class 1	probable medium-chain specific acyl-CoA dehydrogenase, mitochondrial	GITC01148053,GISX01098103,GAJA01005888.1
g9735.t1	Class 1	coatomer subunit epsilon	GITC01135087,GAJA01021164.1
g8480.t1	Class 1	peptidyl-prolyl cis-trans isomerase	GITC01200692,GISX01017882,GAJA01018970.1
g416.t1	Class 1	60s ribosomal protein l15	GAJA01020535.1
g11425.t1	Class 1	adenosine deaminase 2-like	GITC01075799,GAJA01004793.1
g11199.t1	Class 1	transmembrane emp24 domain-containing protein eca	GITC01098662,GISX01075598,GAJA01020423.1
g13896.t1	Class 1	glucose dehydrogenase [FAD, quinone]-like	GITC01058197,GISX01058493,GAJA01006729.1
g2387.t1	Class 1	metalloproteinase inhibitor 3 isoform X2	GITC01048396,GISX01018531,GAJA01019753.1
g15643.t1	Class 1	protein toll	GITC01073170,GISX01131040,GAJA01000999.1
g9235.t1	Class 1	40S ribosomal protein S11	GITC01194517,GAJA01021186.1
g23789.t1	Class 1	histone H1/H5 family protein	GITC01028211,GISX01101988,GAJA01020108.1

g13729.t1	Class 1	histone H2B	GITC01049931,GISX01050887,GAJA01014709.1
g6787.t1	Class 1	aldose reductase	GITC01133192,GAJA01019227.1
g6088.t1	Class 1	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B isoform X1	GITC01072113,GISX01025473,GAJA01018737.1
g1671.t1	Class 1	V-type proton ATPase catalytic subunit A	GITC01203844,GAJA01020437.1
g13993.t1	Class 1	Paramyosin, short form	GITC01138272,GISX01138967,GAJA01019881.1
g5610.t1	Class 1	40S ribosomal protein S28	GITC01103134,GISX01040925,GAJA01020302.1
g566.t1	Class 1	polyadenylate-binding protein 4-like	GITC01191850,GISX01158596,GAJA01018206.1
g176.t1	Class 1	40S ribosomal protein S9	GITC01121164,GAJA01021276.1
g6386.t1	Class 1	60S acidic ribosomal protein P1	GAJA01020446.1
g3852.t1	Class 1	probable isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	GITC01106952,GISX01010734,GAJA01008534.1
g785.t1	Class 1	40S ribosomal protein S6	GITC01062107,GAJA01020461.1
g1685.t1	Class 1	poly(U)-specific endoribonuclease homolog	GITC01136711,
g2033.t1	Class 1	26S proteasome non-ATPase regulatory subunit 12	GITC01037055,GAJA01019681.1
g3398.t1	Class 1	neuroglian isoform X1	GITC01149026,GISX01078699,GAJA01000094.1
g18360.t1	Class 1	40S ribosomal protein S18	GITC01179505,GAJA01018178.1
g8413.t1	Class 1	Krueppel homolog 2	GITC01204983,GISX01040721,GAJA01020150.1
g9134.t1	Class 1	protein 5NUC-like	GITC01050565,GISX01054517,GAJA01015634.1
g5940.t1	Class 1	superoxide dismutase [Cu-Zn], chloroplastic-like isoform X1	GITC01089403,GISX01040095,GAJA01020703.1
g10249.t1	Class 1	FK506-binding protein 2	GITC01095223,GAJA01020242.1
g14706.t1	Class 1	protein ERGIC-53	GITC01133290,GAJA01006253.1
g107.t1	Class 1	serine protease inhibitor 3/4-like isoform X4	GITC01078508,GAJA01019395.1
g27585.t1	Class 1	glucose dehydrogenase [FAD, quinone]	GITC01001100,GISX01062449,GAJA01006341.1
g3442.t1	Class 1	annexin B9	GISX01061893,GAJA01008236.1
g4648.t1	Class 2	venom acid phosphatase Acph-1-like	GITC01165220,GISX01021119,GAJA01011887.1
g10681.t1	Class 2	ras-related protein Rab-14	GITC01027133,GISX01068352,GAJA01005981.1
g20818.t1	Class 2	aminopeptidase N	GITC01140264,GISX01066720,GAJA01011355.1
g11662.t1	Class 2	ras-related protein Rab-1A	GITC01067731,GISX01068352,GAJA01008437.1
g23613.t1	Class 2	N/A	GITC01120012,GISX01038173,GAJA01006288.1
g7656.t1	Class 2	Rac GTPase-activating protein 1	GITC01120006,GISX01038173,GAJA01020436.1
g19482.t1	Class 2	protein Diedel-like	GITC01029986,GISX01083966,GAJA01021338.1
g11841.t1	Class 2	neutral alpha-glucosidase AB	GITC01095440,GAJA01018312.1
g4816.t1	Class 2	GTP-binding protein REM 1	GITC01082018,
g3715.t1	Class 2	leukotriene A-4 hydrolase	GITC01063867,GAJA01018399.1
g1968.t1	Class 2	Venom dipeptidyl peptidase 4	GITC01163866,GAJA01002345.1
g23282.t1	Class 2	venom metalloproteinase 3-like	GITC01190322,GISX01033600,GAJA01006373.1
g13022.t1	Class 2	nucleobindin-2 isoform X1	GITC01032445,GAJA01008352.1
g3802.t1	Class 2	ADP-ribosylation factor 1-like 2	GITC01129126,GISX01054304,GAJA01001633.1
g8456.t1	Class 2	tissue inhibitor of metalloproteinase	GITC01048396,GISX01018531,GAJA01015719.1
g9812.t1	Class 2	atlastin isoform X3	GITC01047626,GAJA01018523.1
g9493.t1	Class 2	rac GTPase-activating protein 1	GITC01134882,GISX01038173,GAJA01020282.1

g14621.t1	Class 2	venom serine carboxypeptidase	GAJA01019126.1
g14524.t1	Class 2	ADP-ribosylation factor-like protein 1	GISX01052920,GAJA01008428.1
g2220.t1	Class 2	GTP-binding protein SAR1b	GITC01198816,GISX01054304,GAJA01021124.1
g3640.t1	Class 3	protein canopy homolog 2	GITC01025928,GAJA01014507.1
g1195.t1	Class 3	N/A	NF
g27107.t1	Class 3	Putative SopA-like secreted protein	GITC01071609,GISX01040514,GAJA01000967.1
g27935.t1	Class 3	N/A	GITC01059467,GAJA01004319.1
g7876.t1	Class 3	N/A	NF
g26195.t1	Class 3	Macrophage receptor MARCO	GITC01141109,GISX01059024,
g15481.t1	Class 3	N/A	GITC01071609,GISX01118346,GAJA01000967.1
g2479.t1	Class 3	endoplasmic reticulum resident protein 44 isoform X2	GITC01177868,GISX01019322,GAJA01007202.1
g27018.t1	Class 3	N/A	GITC01156076,GAJA01003572.1
g23411.t1	Class 3	QWxxN domain	GITC01156068,GISX01079842,GAJA01012562.1
g26802.t1	Class 3	S-antigen protein-like	GITC01129438,GAJA01002663.1
g20948.t1	Class 3	four and a half LIM domains protein 2 isoform X8	GAJA01012187.1
g26603.t1	Class 3	S-antigen protein-like	GITC01129442,GAJA01002663.1
g15349.t1	Class 3	Signal recognition particle receptor subunit beta	GITC01148522,GAJA01006494.1
g28919.t1	Class 3	N/A	GITC01141109,GISX01059024,
g6093.t1	Class 3	tumor suppressor candidate 3	GAJA01021301.1
g20411.t1	Class 3	N/A	GITC01149204,GAJA01018809.1
g26199.t1	Class 3	N/A	GITC01149197,GAJA01015839.1
g10324.t1	Class 3	N/A	GITC01190350,GISX01033592,GAJA01006377.1
g20177.t1	Class 3	N/A	GITC01071867,GISX01061533,GAJA01012569.1
g20322.t1	Class 3	N/A	GITC01071867,GISX01061533,GAJA01012568.1
g1397.t1	Class 3	scavenger receptor cysteine-rich protein	GITC01023641,GISX01102393,GAJA01020201.1
g24368.t1	Class 3	N/A	GITC01156069,GISX01118363,GAJA01017988.1
g26235.t1	Class 3	lysosomal alpha-mannosidase isoform X2	GITC01031693,GISX01151651,GAJA01012944.1
g28902.t1	Class 3	N/A	GITC01029136,GISX01118367,GAJA01016769.1
g11786.t1	Class 3	nose resistant to fluoxetine protein 6-like	GITC01078514,GAJA01004776.1
g18814.t1	Class 3	PREDICTED: uncharacterized protein LOC105569791, partial	GITC01097315,GAJA01004714.1
g27130.t1	Class 3	Putative SopA-like secreted protein	GITC01156068,GISX01118363,GAJA01017988.1
g22070.t1	Class 3	N/A	GITC01029125,GISX01018255,GAJA01011578.1
g25868.t1	Class 3	N/A	GITC01053848,GISX01079839,GAJA01012166.1
g15987.t1	Class 3	N/A	GITC01053848,GISX01079839,GAJA01012937.1
Lb_LbFV_ORF85	Class 3	Ac81-like protein	GAJA01001673.1
g19441.t1	Class 3	N/A	GITC01156063,GISX01118346,GAJA01000968.1
g19774.t1	Class 3	N/A	GISX01050576,GAJA01004072.1
g23606.t1	Class 3	N/A	GITC01071872,GISX01061522,GAJA01012569.1
g25753.t1	Class 3	N/A	GITC01156063,GISX01040514,GAJA01017988.1
g21419.t1	Class 3	N/A	GITC01105368,GISX01017935,GAJA01000968.1

g385.t1	Class 3	protein THEM6	GITC01039010,GAJA01008840.1
g25831.t1	Class 3	N/A	GITC01141109,
g10080.t1	Class 3	probable uridine nucleosidase 1 isoform X1	GITC01157759,GAJA01005947.1
g19192.t1	Class 3	Leucine-rich repeat (LRR) protein	GITC01053848,GISX01079839,GAJA01012937.1
g22788.t1	Class 3	N/A	GITC01204980,GISX01059465,GAJA01006002.1
g23729.t1	Class 3	N/A	GITC01090731,GAJA01016213.1
g3916.t1	Class 3	QWxxN domain	GITC01105368,GAJA01012562.1
g29741.t1	Class 3	N/A	GITC01153128,GISX01024528,GAJA01020180.1
g11793.t1	Class 3	receptor expression-enhancing protein 5	GAJA01021283.1
g24380.t1	Class 3	pentapeptide repeat-containing protein	GITC01105370,GISX01040514,GAJA01012562.1
g18880.t1	Class 3	N/A	GITC01156063,GISX01136942,GAJA01015791.1
g20214.t1	Class 3	N/A	GITC01044639
g9031.t1	Class 3	staphylococcal nuclease domain-containing protein 1	GITC01105858,GAJA01019200.1
g28195.t1	Class 3	N/A	GITC01029125,GISX01018255,GAJA01011578.1
g26590.t1	Class 3	N/A	GITC01190350,GISX01033592,GAJA01006377.1
g2355.t1	Class 3	tumor protein D52 isoform X2	GITC01199611,GAJA01020060.1
g24649.t1	Class 3	N/A	GITC01053848,GISX01079839,GAJA01012168.1
g465.t1	Class 3	TM2 domain-containing protein almondex	GITC01052559,GAJA01002341.1
g10682.t1	Class 3	transmembrane emp24 domain-containing protein-like	GITC01098662,GISX01075598,GAJA01020423.1
g11630.t1	Class 3	protein FAM151B isoform X2	GITC01062209,GAJA01015363.1
g2472.t1	Class 3	transferrin-like	GITC01031132,GAJA01014176.1
g7964.t1	Class 3	hydroxysteroid dehydrogenase-like protein 2	GITC01044418,GISX01050519,GAJA01006815.1
g27108.t1	Class 3	N/A	GITC01156068,GISX01118346,GAJA01020069.1
g19442.t1	Class 3	N/A	GITC01035349,GISX01136942,GAJA01012767.1
g3917.t1	Class 3	hypothetical protein	GITC01105368,GAJA01000968.1
g20130.t1	Class 3	N/A	GITC01105360,GISX01017935,GAJA01000967.1
g1887.t1	Class 3	N/A	GITC01010741,GAJA01018131.1
g27228.t1	Class 3	S-antigen protein-like	GITC01129437,GAJA01002660.1
g11642.t1	Class 3	PRA1 family protein 3	GITC01051020,GAJA01020291.1
g25079.t1	Class 3	AGAP000929-PA-like protein	GITC01026659,GAJA01006995.1
g25011.t1	Class 3	N/A	GITC01168882,GAJA01020320.1
g15068.t1	Class 3	N/A	GITC01156054,GISX01079840,GAJA01009572.1
g25395.t1	Class 3	N/A	GITC01023910,GISX01033600,GAJA01006374.1
g20781.t1	Class 3	N/A	GITC01071613,GISX01079839,GAJA01012562.1
g3999.t1	Class 3	14-3-3 protein zeta isoform X1	GAJA01021146.1
g25297.t1	Class 3	N/A	GITC01156063,GISX01040514,GAJA01003572.1
g23005.t1	Class 3	Leucine-rich repeat (LRR) protein	GITC01197868,GISX01079839,GAJA01012937.1
g22206.t1	Class 3	N/A	GITC01129447,GAJA01002663.1
g28873.t1	Class 3	N/A	GITC01105360,GISX01017935,GAJA01000968.1
g6078.t1	Class 3	N/A	GITC01127826,GISX01133594,GAJA01019220.1
g899.t1	Class 3	N/A	NF

g26022.t1	Class 3	N/A	GITC01016685,GISX01059471,GAJA01004736.1
g25331.t1	Class 3	N/A	GAJA01015285.1
g19696.t1	Class 3	N/A	GITC01053838,GISX01079839,GAJA01012937.1
g19027.t1	Class 3	N/A	GITC01029125,GISX01018255,GAJA01011578.1
g21443.t1	Class 3	N/A	GITC01022450,GAJA01000837.1
g9643.t1	Class 3	uncharacterized protein LOC103318133	GAJA01014376.1
g17673.t1	Class 3	Leucine-rich repeat (LRR) protein	GITC01053839,GISX01079839,GAJA01012937.1
g9726.t1	Class 3	double-stranded RNA-specific editase 1 isoform X3	GITC01197713,GAJA01021244.1
g17565.t1	Class 3	Leucine-rich repeat (LRR) protein	GITC01105368,GISX01017935,GAJA01000968.1
g25522.t1	Class 3	N/A	GITC01156069,GISX01118346,GAJA01000968.1
g17667.t1	Class 3	hypothetical protein CsmBAC4b19.5 [Cotesia sesamiae]	GITC01061878,GAJA01010387.1
g13578.t1	Class 3	N/A	GITC01144757,GAJA01018638.1
g19652.t1	Class 3	N/A	GITC01028117,GISX01084597,GAJA01011758.1

Table 6: Listing of proteins with both a signal peptide and one or more transmembrane domain found in the *Lh 14* MSEV and *Lb G* VLP sequences. Reciprocal BLASTP results with accession number, Class, E value, %ID, % query coverage and annotation are shown. Criteria for score cutoff detailed in Methods.

High-Scoring Signal Peptide and Transmembrane Domain Lh and Lb Proteins					
Species: Protein	Class	E-value	%ID	Quer. Cov	Annotation
Lh: GAJC01030581.1_43	1	1E-140	94.3	93	translocon-associated protein subunit beta
Lb: g5494.t1	1				translocon-associated protein subunit beta
Lh: GAJC01021107.1_29	1	1E-149	85.14	93	malectin-A
Lb: g12151.t1	1				malectin-A
Lh: GAJC01009924.1_35	1	2E-145	89.47	100	transmembrane emp24 domain-containing protein bai isoform X2
Lb: g9705.t1	1				transmembrane emp24 domain-containing protein bai isoform X2
Lh: GAJC01024669.1_3	3	4.00E-04	48.39	8	N/A
Lb: g12611.t1	1				renin receptor isoform X2
Lh: GAJC01022480.1_6	1	5.00E-04	23.53	10	thioredoxin domain-containing protein
Lb: g11675.t1	1				sulfhydryl oxidase 2-like

Table 7: Results of BLASTP searches using the *Lb* VLP proteins as query (column 1) against polydnavirus protein database (nr: PDV), or the unrestricted non-redundant (nr) database. Column 2 indicates whether the search was performed against nr: PDV or against the unrestricted nr database. Subject indicates the subject accession numbers obtained. Description indicates annotation of subject sequence associated with the BLAST result. The %ID, E-value and Quer. Cov. columns indicate the corresponding results from these searches. The nr searches were done at the same time as the PDV searches.

PDV Results						
Query	Database	Subject	Description	%ID	E-value	Quer Cov
g11702.t1	nr: PDV	AFN42311.1	cytochrome P450 [Cotesia sesamiae Mombasa bracovirus]	32.258	1.39E-33	57.6
	nr	XP_033216126.1	cytochrome P450 4g15-like [Belonocnema treatae]	70.945	0	87.17
g26437.t1	nr: PDV	AAV98010.1	hypothetical protein ORF3006 [Cotesia plutellae polydnavirus]	84.127	1.04E-34	92.06
	nr	XP_033230431.1	histone H4-like [Belonocnema treatae]	100	7.36E-36	100
g3616.t1	nr: PDV	AFN42312.1	cytochrome P450 [Cotesia sesamiae Mombasa bracovirus]	33.477	1.73E-83	56.8
	nr	XP_033229661.1	uncharacterized protein LOC117181207 [Belonocnema treatae]	68.162	0	83.97
g2320.t1	nr: PDV	AGO14415.1	hypothetical protein CsmBAC4b19.6 [Cotesia sesamiae]	37.313	9.60E-26	54.48
	nr	XP_024891430.1	putative nuclease HARBI1 [Temnothorax curvispinosus]	48.175	4.72E-34	66.42
g13146.t1	nr: PDV	YP_009665791.1	putative histone 4 [Cotesia glomerata bracovirus]	85.227	2.37E-51	93.18
	nr	EDW44574.1	GM19319 [Drosophila sechellia]	98.058	7.53E-64	100
g5016.t1	nr: PDV	AEE09528.1	conserved hypothetical protein [Cotesia vestalis bracovirus]	30.159	3.70E-05	53.97
	nr	XP_033225200.1	ribonuclease Oy-like [Belonocnema treatae]	56.376	9.64E-60	79.87
g17667.t1	nr: PDV	AGO14414.1	hypothetical protein CsmBAC4b19.5 [Cotesia sesamiae]	56.94	9.20E-124	77.58
	nr	XP_021202049.1	uncharacterized protein LOC105841287 isoform X1 [Bombyx mori]	60.513	2.82E-81	79.49

Table 8: Results of BLASTP searches using *Lb* VLP proteins as query (column 1) against the Viridae protein database (nr: Viridae) or the unrestricted non-redundant (nr) database. Database in column 2 indicates whether the search was performed against the nr: Viridae or the unrestricted nr database. Subject indicates subject accession numbers obtained. Description indicates annotation of the subject sequences associated with the BLAST result. The %ID, E-value and Quer Cov columns indicate the corresponding results from these searches. The nr searches were done at the same time as the Viridae searches.

Viridae Results						
Query	Database	Subject	Description	%ID	E-value	Quer Cov
g20818.t1	nr: Viridae	AYV75523.1	hypothetical protein Terrestrivirus2_31 [Terrestrivirus sp.]	31.564	3.12E-42	51.12
	nr	XP_033224230.1	aminopeptidase Q-like [Belonocnema treatae]	50.44	6.04E-107	65.1
g15500.t1	nr: Viridae	YP_009143334.1	Secreted complement binding protein C3b/C4b [Raccoonpox virus]	29.104	1.69E-04	48.51
	nr	XP_026689472.1	neurogenic locus notch homolog protein 1-like [Ciona intestinalis]	26.556	1.14E-16	43.98
g12336.t1	nr: Viridae	ARF02669.1	SWPV1-055 [Shearwaterpox virus]	31.068	1.55E-05	50.49
	nr	XP_033220088.1	protein 60A-like [Belonocnema treatae]	43.137	1.46E-103	64.71
g8480.t1	nr: Viridae	ARF11440.1	cyclophilin type peptidyl-prolyl cis-trans isomerase [Klosneuvirus KNV1]	60.825	3.54E-36	77.32
	nr	XP_014278363.1	peptidyl-prolyl cis-trans isomerase [Halyomorpha halys]	84.314	6.78E-95	95.42
g10248.t1	nr: Viridae	CAB4130158.1	FKBP-type peptidyl-prolyl cis-trans isomerase [uncultured Caudovirales phage]	46.341	1.73E-04	63.41
	nr	XP_014220978.1	FK506-binding protein 2 [Trichogramma pretiosum]	80.519	1.79E-85	89.61
g17667.t1	nr: Viridae	AGO14414.1	hypothetical protein CsmBAC4b19.5 [Cotesia sesamiae]	56.94	2.48E-120	77.58
	nr	XP_025836086.1	uncharacterized protein LOC112906351 [Agrilus planipennis]	54.93	2.08E-114	78.17
g25630.t1	nr: Viridae	AFX92824.1	glucose-methanol-choline oxidoreductase [Megavirus courdo11]	64.865	4.69E-07	83.78
	nr	XP_033220716.1	glucose dehydrogenase [FAD, quinone]-like isoform X1 [Belonocnema treatae]	70.82	8.04E-161	84.92
g4910.t1	nr: Viridae	CAB4164752.1	FabG Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases) [uncultured Caudovirales phage]	30.864	0.001	55.56
	nr	XP_015519192.1	PREDICTED: inactive hydroxysteroid dehydrogenase-like protein 1 isoform X1 [Neodiprion lecontei]	73.481	1.61E-95	91.16
g13578.t1	nr: Viridae	QBK87221.1	translation initiation factor 2, alpha subunit [Marseillevirus LCMAC201]	46.154	1.78E-16	67.03
	nr	XP_033218990.1	eukaryotic translation initiation factor 2 subunit 1 [Belonocnema treatae]	96.33	0	98.78
g4237.t1	nr: Viridae	AAX62124.1	manganese superoxide dismutase [Bacillus thuringiensis phage MZTP02]	42.365	9.77E-42	58.62
	nr	XP_033230380.1	alkaline phosphatase, tissue-nonspecific isozyme-like isoform X2 [Belonocnema treatae]	81.76	0	91.42
g26546.t1	nr: Viridae	YP_009408579.1	Serpin 123 [NY_014 poxvirus]	46.774	0.001	61.29
	nr	ACQ83466.1	venom serpin precursor [Leptopilina boulardi]	68.478	9.48E-26	77.17
g15014.t1	nr: Viridae	YP_009282876.1	eev type-1 membrane glycoprotein [Skunkpox virus]	27.35	7.21E-04	46.15
	nr	XP_022783568.1	sushi domain-containing protein 2-like isoform X3 [Stylophora pistillata]	27.826	5.04E-07	51.3
g10794.t1	nr: Viridae	AGV99426.1	3-oxoacyl-[acyl-carrier protein] reductase [Bacillus phage proCM3]	39.474	2.21E-04	50
	nr	XP_033223437.1	3-hydroxyacyl-CoA dehydrogenase type-2 [Belonocnema treatae]	78.824	3.85E-148	89.8

g6787.t1	nr: Viridae	ATW69262.1	glyoxal reductase [Staphylococcus phage UPMK_1]	41.438	1.49E-70	59.25
	nr	XP_033223159.1	aldo-keto reductase family 1 member B1-like [Belonocnema treatae]	85.22	0	91.82
g12817.t1	nr: Viridae	AYV82031.1	Ras-like protein Rab-11A [Homavirus sp.]	55.882	1.13E-62	72.35
	nr	XP_002434427.1	ras-related protein Rab-2A [Ixodes scapularis]	93.367	5.35E-132	96.43
g11662.t1	nr: Viridae	AYV75487.1	GTP-binding protein YPTC1 [Terrestrivirus sp.]	69.048	3.04E-79	81.55
	nr	XP_033218245.1	ras-related protein ORAB-1 [Belonocnema treatae]	97.044	8.91E-145	99.51
g18935.t1	nr: Viridae	YP_003970008.1	putative DnaJ/Hsp40 [Cafeteria roenbergensis virus BV-PW1]	36.765	5.67E-08	69.12
	nr	XP_011309936.1	PREDICTED: dnaJ homolog subfamily B member 11 [Fopius arisanus]	80.352	0	90.03
g2172.t1	nr: Viridae	QFG74169.1	aspartyl protease [Megaviridae environmental sample]	39.329	9.55E-61	54.88
	nr	XP_033215802.1	lysosomal aspartic protease [Belonocnema treatae]	82.271	0	90.58
g8500.t1	nr: Viridae	AYV84508.1	WD domain-containing protein [Hyperionvirus sp.]	44.828	1.24E-05	68.97
	nr	XP_033210231.1	coatomer subunit alpha [Belonocnema treatae]	93.008	0	97.05
g7569.t1	nr: Viridae	QDH87640.1	hypothetical protein H1Bulk3010505_000003, partial [Mitovirus sp.]	71.545	1.72E-51	78.86
	nr	NP_001295472.1	elongation factor 1-alpha [Athalia rosae]	96.009	0	98.23
g1465.t1	nr: Viridae	QDY52194.1	acyl-CoA dehydrogenase [Mimiviridae sp. ChoanoV1]	23.308	3.82E-12	43.36
	nr	XP_033212879.1	short-chain specific acyl-CoA dehydrogenase, mitochondrial [Belonocnema treatae]	83.416	0	93.56
g12249.t1	nr: Viridae	QFG73724.1	Hsp90 protein [Megaviridae environmental sample]	65.455	1.17E-87	81.82
	nr	XP_029670400.1	heat shock protein 83 [Formica exsecta]	91.061	0	97.07
g8740.t1	nr: Viridae	YP_009213665.1	putative fructose-1,6-bisphosphate aldolase [Prochlorococcus phage P-TIM68]	51.955	3.62E-119	65.92
	nr	XP_011329542.2	fructose-bisphosphate aldolase isoform X2 [Ooceraea biroi]	90.137	0	95.07
g18649.t1	nr: Viridae	AEW87518.1	non-structural protein, partial [Hepatitis E virus]	57.895	5.74E-05	71.05
	nr	XP_015588029.1	40S ribosomal protein S19 [Cephus cinctus]	86.154	1.00E-78	94.62
g2224.t1	nr: Viridae	CAB4133185.1	COG3959 Transketolase, N-terminal subunit [uncultured Caudovirales phage]	33.566	1.83E-11	48.95
	nr	XP_033214935.1	transketolase-like protein 2 isoform X1 [Belonocnema treatae]	88.136	0	95.42
g658.t1	nr: Viridae	ARF11440.1	cyclophilin type peptidyl-prolyl cis-trans isomerase [Klosneuvirus KNV1]	62.245	1.37E-34	76.53
	nr	XP_011698873.1	PREDICTED: peptidyl-prolyl cis-trans isomerase 5 [Wasmannia auropunctata]	78.261	1.49E-99	88.04
g3616.t1	nr: Viridae	AFN42312.1	cytochrome P450 [Cotesia sesamiae Mombasa bracovirus]	33.477	4.67E-80	56.8
	nr	XP_033229661.1	uncharacterized protein LOC117181207 [Belonocnema treatae]	68.162	0	83.97
g13146.t1	nr: Viridae	YP_009665791.1	putative histone 4 [Cotesia glomerata bracovirus]	85.227	6.39E-48	93.18
	nr	EOB06745.1	Histone H4, partial [Anas platyrhynchos]	98.058	5.72E-64	100
g186.t1	nr: Viridae	YP_003517850.1	superoxide dismutase [Lymantria xylina nucleopolyhedrovirus]	56.291	5.72E-56	76.16
	nr	AET83766.1	cytoplasmic superoxide dismutase 1 [Leptopilina boulardi]	100	9.14E-103	100
g7964.t1	nr: Viridae	QIH04859.1	hypothetical protein [Dasineura jujubifolia toursvirus 2a]	27	1.03E-08	46.5
	nr	XP_033214489.1	hydroxysteroid dehydrogenase-like protein 2 [Belonocnema treatae]	79.208	0	90.84
g2220.t1	nr: Viridae	YP_009163818.1	ORF_057R [Scale drop disease virus]	61.017	3.21E-73	76.27
	nr	KYN40003.1	GTP-binding protein SAR1b [Trachymyrmex septentrionalis]	93.258	6.73E-119	96.63
g11675.t1	nr: Viridae	AYV75779.1	putative thioredoxin domain-containing protein 5 [Terrestrivirus sp.]	31.343	5.59E-06	56.72
	nr	XP_033207328.1	sulfhydryl oxidase 1-like [Belonocnema treatae]	60.429	0	77.76
g6078.t1	nr: Viridae	QKU33982.1	putative heat shock 70 kDa protein [Tupanvirus deep ocean]	35.849	2.00E-20	57.23

	nr	XP_033207075.1	hypoxia up-regulated protein 1 isoform X1 [Belonocnema treatae]	77.544	0	89.38
g13994.t1	nr: Viridae	4XA3_A	Chain A, Gp7-MYH7(1361-1425)-Eb1 chimera protein	48.649	1.70E-12	68.92
	nr	XP_033226047.1	paramyosin, long form [Belonocnema treatae]	96.339	0	98.63
Lb_LbFV_ORF85	nr: Viridae	QKN22513.1	putative Ac81-like protein [Drosophila-associated filamentous virus]	32.589	1.01E-26	51.79
	nr	QKN22513.1	putative Ac81-like protein [Drosophila-associated filamentous virus]	32.589	1.01E-24	51.79
g12.t1	nr: Viridae	AEQ60451.1	Dnaj-like protein [Acanthamoeba castellanii mamavirus]	38.806	1.76E-06	70.15
	nr	XP_033225841.1	dnaJ homolog subfamily C member 3 [Belonocnema treatae]	79.528	4.84E-140	90.94
g3260.t1	nr: Viridae	CAB4137972.1	AhpC Peroxiredoxin [uncultured Caudovirales phage]	31.098	6.79E-20	51.83
	nr	XP_033209095.1	peroxiredoxin-like [Belonocnema treatae]	78.512	3.86E-132	90.91
g5016.t1	nr: Viridae	ARF09950.1	T2 family ribonuclease [Indivirus ILV1]	34.127	1.10E-12	53.17
	nr	XP_033225200.1	ribonuclease Oy-like [Belonocnema treatae]	56.376	1.01E-59	79.87
g5735.t1	nr: Viridae	ADX06546.1	putative thioredoxin family protein [Organic Lake phycodnavirus]	40.278	4.52E-09	63.89
	nr	XP_011062424.1	PREDICTED: protein disulfide-isomerase A6 [Acromyrmex echinatior]	79.469	0	88.16
g2479.t1	nr: Viridae	YP_009904145.1	thioredoxin [Burkholderia phage BcepSaruman]	33.735	3.32E-06	55.42
	nr	XP_015590504.1	endoplasmic reticulum resident protein 44 isoform X2 [Cephus cinctus]	76.302	0	90.1
g5466.t1	nr: Viridae	QHN71161.1	Peptidase domain protein [Mollivirus kamchatka]	31.25	9.85E-20	57.5
	nr	XP_033223899.1	signal peptidase complex catalytic subunit SEC11A [Belonocnema treatae]	92.222	2.43E-120	98.33
g11523.t1	nr: Viridae	QKF93534.1	peroxisomal acyl-CoA oxidase [Fadolivirus 1]	22.792	1.28E-05	39.89
	nr	XP_033210196.1	very long-chain specific acyl-CoA dehydrogenase, mitochondrial [Belonocnema treatae]	82.12	0	91.93
g15153.t1	nr: Viridae	NP_536444.1	D14L [Monkeypox virus Zaire-96-I-16]	40.964	7.50E-05	50.6
	nr	KTG07343.1	hypothetical protein cypCar_00007315, partial [Cyprinus carpio]	33.333	3.07E-08	52.27
g11793.t1	nr: Viridae	QFG73569.1	TB2/DP1, HVA22 family protein [Megaviridae environmental sample]	25.143	4.47E-09	48
	nr	XP_033214298.1	receptor expression-enhancing protein 5-like isoform X1 [Belonocnema treatae]	81.111	2.55E-101	92.22
g12150.t1	nr: Viridae	AEX62781.1	putative heat shock 70 kDa protein [Moumouvirus Monve]	77.157	2.95E-103	86.29
	nr	XP_011259986.1	heat shock 70 kDa protein cognate 4 [Camponotus floridanus]	95.33	0	98.23
g465.t1	nr: Viridae	ALF01610.1	hypothetical protein BMBtp4_24 [Bacillus phage vB_BtS_BMBtp16]	47.826	7.64E-04	63.04
	nr	XP_033211095.1	TM2 domain-containing protein almondex [Belonocnema treatae]	82.553	1.04E-144	91.91
g11592.t1	nr: Viridae	CAB4129576.1	TrxA Thiol-disulfide isomerase and thioredoxins [uncultured Caudovirales phage]	39.062	1.04E-04	67.19
	nr	XP_024884258.1	protein disulfide-isomerase isoform X1 [Temnothorax curvispinosus]	67.432	0	84.97
g13729.t1	nr: Viridae	YP_009482227.1	Histone H2B domain containing protein [Pandoravirus neocaledonia]	70.968	1.31E-43	88.17
	nr	CAB0028520.1	unnamed protein product [Trichogramma brassicae]	95.161	7.28E-68	98.39
g21089.t1	nr: Viridae	YP_004821530.1	serine protease inhibitor-like protein SPI-1 [Yokapox virus]	33.799	2.77E-37	53.07
	nr	ACQ83466.1	venom serpin precursor [Leptopilina boulardi]	72.634	0	82.1
g3715.t1	nr: Viridae	ALJ98964.1	membrane alanine aminopeptidase N [Acinetobacter phage Ab105-3phi]	28.902	3.38E-24	46.53
	nr	XP_033208757.1	leukotriene A-4 hydrolase isoform X3 [Belonocnema treatae]	81.5	0	91.5
g1968.t1	nr: Viridae	QBK88429.1	S9 family peptidase [Mimivirus LCMiAC01]	34.014	1.27E-07	51.02
	nr	XP_033227064.1	venom dipeptidyl peptidase 4 isoform X1 [Belonocnema treatae]	88.889	1.29E-95	91.98
g11786.t1	nr: Viridae	ADZ29534.1	serine protease inhibitor-like protein [Cowpox virus]	24.503	8.84E-14	48.01
	nr	XP_033222638.1	uncharacterized protein LOC117176493 [Belonocnema treatae]	63.971	1.32E-43	72.06

g14621.t1	nr: Viridae	API81738.1	hypothetical protein [Cafeteria roenbergensis]	27.602	1.20E-36	44.8
	nr	KAF7392498.1	hypothetical protein HZH66_008331 [Vespula vulgaris]	73.154	0	85.01
g937.t1	nr: Viridae	CAB5514185.1	CPXV199 protein [Cowpox virus]	44.643	2.29E-04	60.71
	nr	XP_033221269.1	apolipophorins [Belonocnema treatae]	69.723	0	84.19
g8282.t1	nr: Viridae	4XA1_A	Chain A, Gp7-MYH7(1173-1238)-EB1 chimera protein	67.442	4.43E-10	83.72
	nr	XP_033213193.1	myosin heavy chain, muscle isoform X1 [Belonocnema treatae]	90.342	0	93.16
g12024.t1	nr: Viridae	ATZ80657.1	hypothetical protein BMW23_0611 [Bodo saltans virus]	28.488	8.41E-10	44.77
	nr	XP_033210032.1	probable beta-hexosaminidase fdl isoform X1 [Belonocnema treatae]	81.25	0	92.31
g1259.t1	nr: Viridae	ADX06678.1	putative disulphide isomerase [Organic Lake phycodnavirus]	41.86	2.19E-11	65.12
	nr	XP_033210967.1	protein disulfide-isomerase A3 [Belonocnema treatae]	72.537	0	88.26
g5064.t1	nr: Viridae	YP_009003341.1	enolase [Streptococcus phage 20617]	49.074	1.82E-132	66.9
	nr	XP_029053482.1	enolase [Osmia bicornis bicornis]	94.7	0	98.39
g11933.t1	nr: Viridae	QKF93897.1	Rab-related GTPase protein [Fadolivirus 1]	29.06	1.31E-06	52.99
	nr	XP_024873570.1	ADP-ribosylation factor 2 [Temnothorax curvispinosus]	93.889	3.71E-123	98.33
g19482.t1	nr: Viridae	YP_003422448.1	hypothetical protein PsunGV_gp109 [Pseudalatia unipuncta granulovirus]	50	2.15E-08	68.75
	nr	XP_017869253.1	PREDICTED: protein Diedel [Drosophila arizonae]	48	3.14E-08	74
g9812.t1	nr: Viridae	QFG73773.1	guanylate-binding protein [Megaviridae environmental sample]	29.762	1.61E-31	46.73
	nr	XP_015113556.1	atlastin isoform X5 [Diachasma alloeum]	83.456	0	93.38
g11746.t1	nr: Viridae	VBB18527.1	hypothetical protein YASMINEVIRUS_990 [Yasminevirus sp. GU-2018]	29.909	1.48E-28	47.13
	nr	XP_033222189.1	putative fatty acyl-CoA reductase CG8306 [Belonocnema treatae]	77.603	0	89.59
g4242.t1	nr: Viridae	ATW69167.1	lantibiotic immunity protein F [Staphylococcus phage UPMK_1]	32.026	3.38E-12	57.52
	nr	XP_033209457.1	ABC transporter G family member 23 isoform X2 [Belonocnema treatae]	88.345	0	93.24
g16252.t1	nr: Viridae	YP_009507248.1	ORF B [Trichoplusia ni TED virus]	50	2.14E-06	64.81
	nr	XP_012235034.1	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC105679540 [Linepithema humile]	55.556	8.46E-29	73.5
g1748.t1	nr: Viridae	AYV76390.1	mitochondrial carrier [Terrestrivirus sp.]	21.081	0.001	46.49
	nr	XP_015181057.1	PREDICTED: phosphate carrier protein, mitochondrial [Polistes dominula]	87.006	0	93.22
g13332.t1	nr: Viridae	AYV75146.1	plasma membrane ATPase [Terrestrivirus sp.]	34.804	3.10E-24	55.39
	nr	XP_011336454.1	calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type isoform X1 [Ooceraea biroi]	93.77	0	98.1
g5940.t1	nr: Viridae	AYV81989.1	superoxide dismutase [Homavirus sp.]	52.033	3.02E-38	69.11
	nr	AET83767.1	extracellular superoxide dismutase 3 [Leptopilina boulardi]	100	1.76E-101	100
g3398.t1	nr: Viridae	YP_009618565.1	putative tail protein [Shigella phage Sf14]	29.787	5.15E-06	41.84
	nr	XP_033217713.1	neuroglian isoform X4 [Belonocnema treatae]	79.659	0	88.74
g10681.t1	nr: Viridae	AYV81890.1	putative ras-related protein Rab-13 [Harvfovirus sp.]	37.755	1.94E-18	67.35
	nr	XP_033227775.1	ras-related protein Rab-14 [Belonocnema treatae]	86.636	2.44E-133	87.1
g8876.t1	nr: Viridae	QDH91349.1	hypothetical protein H3Bulk427771_000001, partial [Partitiviridae sp.]	73.312	7.37E-156	85.53
	nr	XP_011345547.1	transitional endoplasmic reticulum ATPase TER94 isoform X1 [Ooceraea biroi]	96.766	0	98.71
g408.t1	nr: Viridae	CAB4221060.1	CbpA DnaJ-class molecular chaperone [uncultured Caudovirales phage]	38.71	2.16E-08	77.42
	nr	XP_033229918.1	dnaJ homolog subfamily A member 1 [Belonocnema treatae]	87.626	0	94.44

g27228.t1	nr: Viridae	NP_671543.1	EVM025 [Ectromelia virus]	28.682	8.32E-06	44.57
	nr	XP_033212427.1	uncharacterized PE-PGRS family protein PE_PGRS54-like [Belonocnema treatae]	34.94	0.001	49.4
g2320.t1	nr: Viridae	AGO14415.1	hypothetical protein CsmBAC4b19.6 [Cotesia sesamiae]	37.313	2.59E-22	54.48
	nr	XP_024891430.1	putative nuclease HARBI1 [Temnothorax curvispinosus]	48.175	4.94E-34	66.42
g21977.t1	nr: Viridae	AAC58526.1	enhancin [Lymantria dispar multiple nucleopolyhedrovirus]	22.013	4.63E-04	40.25
	nr	WP_014227995.1	MULTISPECIES: viral enhancin protein [Klebsiella]	23.958	4.86E-14	41.04
g26437.t1	nr: Viridae	AAV98010.1	hypothetical protein ORF3006 [Cotesia plutellae polydnavirus]	84.127	2.80E-31	92.06
	nr	EDW52671.1	GM22498 [Drosophila sechellia]	100	4.35E-36	100
g3124.t1	nr: Viridae	QDH76139.1	putative gag protein, partial [Porcine endogenous retrovirus]	98.23	9.40E-79	100
	nr	AGR44868.1	actin-4 [Bombyx mori]	99.734	0	100
g6700.t1	nr: Viridae	AYV85778.1	rab family small GTPase [Satyrvirus sp.]	40.206	1.15E-22	67.01
	nr	TGZ47008.1	Uncharacterized protein DBV15_03015, partial [Temnothorax longispinosus]	95.169	7.51E-145	99.03
g3802.t1	nr: Viridae	YP_009163818.1	ORF_057R [Scale drop disease virus]	34.454	1.88E-16	57.98
	nr	KRZ51011.1	ADP-ribosylation factor 1-like 2 [Trichinella nativa]	76.339	9.66E-117	82.14
g8002.t1	nr: Viridae	QKF93583.1	mitochondrial carrier protein [Fadolivirus 1]	25.556	8.83E-08	45
	nr	XP_001606673.1	ADP,ATP carrier protein 2 [Nasonia vitripennis]	95.333	0	98.67
g8024.t1	nr: Viridae	YP_001883378.1	aminoacylase [Musca domestica salivary gland hypertrophy virus]	29.28	1.06E-36	47.39
	nr	XP_033221955.1	aminoacylase-1-like [Belonocnema treatae]	75.661	0	87.04
g3791.t1	nr: Viridae	P00545.2	RecName: Full=Tyrosine-protein kinase transforming protein fms	23.024	4.00E-05	40.21
	nr	XP_033219612.1	fasciclin-2-like isoform X1 [Belonocnema treatae]	68.112	0	82.38
g729.t1	nr: Viridae	NP_598374.1	ubiquitin-like protein [Murine osteosarcoma virus]	53.03	1.07E-31	64.39
	nr	XP_033220510.1	uncharacterized protein LOC117175076 [Belonocnema treatae]	81.061	1.71E-74	88.64
g7824.t1	nr: Viridae	QFG73724.1	Hsp90 protein [Megaviridae environmental sample]	53.704	8.59E-65	71.76
	nr	XP_003493872.1	endoplasmin [Bombus impatiens]	80.79	0	91.14
g1869.t1	nr: Viridae	AGV99422.1	ABC transporter permease/ATP-binding protein [Bacillus phage proCM3]	43.396	3.97E-06	60.38
	nr	XP_033224972.1	ATP-dependent translocase ABCB1-like isoform X1 [Belonocnema treatae]	74.16	0	85.87
g329.t1	nr: Viridae	QBK89144.1	2OG-Fe(II) oxygenase [Mimivirus LCMiAC02]	37.949	5.60E-36	59.49
	nr	PBC29644.1	Procollagen-lysine,2-oxoglutarate 5-dioxygenase [Apis cerana cerana]	75	7.89E-180	85.19
g13391.t1	nr: Viridae	AYV80835.1	SDR family NAD(P)-dependent oxidoreductase [Harvfovirus sp.]	23.834	1.42E-05	46.63
	nr	XP_033207630.1	peroxisomal multifunctional enzyme type 2 isoform X1 [Belonocnema treatae]	83.588	0	92.21
g3528.t1	nr: Viridae	ARF11106.1	serine/threonine protein kinase [Hokovirus HKV1]	32.759	1.79E-15	62.07
	nr	XP_033210455.1	titin homolog isoform X2 [Belonocnema treatae]	80.323	7.71E-155	88.39
g26603.t1	nr: Viridae	CAD0300740.1	hypothetical protein [Enterococcus phage 156]	25.455	2.60E-08	48.18
	nr	XP_031659529.1	S-antigen protein-like [Oncorhynchus kisutch]	24.82	5.43E-14	47.12
g15152.t1	nr: Viridae	AYV76458.1	putative short-chain type dehydrogenase/reductase [Terrestrivirus sp.]	30.612	1.20E-04	53.06
	nr	XP_033217196.1	farnesol dehydrogenase-like [Belonocnema treatae]	68.05	6.73E-117	80.5
g27585.t1	nr: Viridae	AFX92824.1	glucose-methanol-choline oxidoreductase [Megavirus courdo11]	55	2.96E-05	70
	nr	XP_033213779.1	glucose dehydrogenase [FAD, quinone]-like [Belonocnema treatae]	59.398	4.44E-48	79.7
g12041.t1	nr: Viridae	VBB19039.1	ras-related protein RABB1b [Yasminevirus sp. GU-2018]	45.963	8.76E-42	67.08
	nr	XP_033227464.1	ras-related protein Rab-5C [Belonocnema treatae]	99.061	5.76E-156	99.06

g5201.t1	nr: Viridae	QKU35287.1	catalase [Tupanvirus soda lake]	41.763	2.81E-105	58.24
	nr	XP_033224601.1	catalase-like [Belonocnema treatae]	84.553	0	93.29
g12490.t1	nr: Viridae	AGF85322.1	hypothetical protein glt_00513 [Moumouvirus goulette]	70.707	6.92E-91	85.86
	nr	XP_015179269.1	PREDICTED: heat shock 70 kDa protein cognate 3 isoform X2 [Polistes dominula]	93.016	0	97.14
g332.t1	nr: Viridae	QDP49545.1	hypothetical protein GOVbin3695_74 [Prokaryotic dsDNA virus sp.]	34.634	1.30E-27	56.59
	nr	XP_033224286.1	procollagen-lysine,2-oxoglutarate 5-dioxygenase isoform X3 [Belonocnema treatae]	70.883	0	87.11
g8281.t1	nr: Viridae	5CJ1_A	Chain A, Gp7-MYH7-(1526-1571) chimera protein	67.347	5.16E-09	85.71
	nr	XP_033215377.1	myosin heavy chain, muscle isoform X30 [Belonocnema treatae]	98.897	0	99.31
g107.t1	nr: Viridae	1C8O_A	Chain A, 2.9 A Structure Of Cleaved Viral Serpin Crma	30.128	1.74E-41	55.13
	nr	XP_033215197.1	alaserpin-like isoform X5 [Belonocnema treatae]	63.529	5.74E-155	81.76
g11702.t1	nr: Viridae	AFN42311.1	cytochrome P450 [Cotesia sesamiae Mombasa bracovirus]	32.258	3.75E-30	57.6
	nr	XP_033216126.1	cytochrome P450 4g15-like [Belonocnema treatae]	70.945	0	87.17
g13896.t1	nr: Viridae	AGF84934.1	oxidoreductase [Moumouvirus goulette]	44.444	6.39E-05	57.14
	nr	XP_033213779.1	glucose dehydrogenase [FAD, quinone]-like [Belonocnema treatae]	48.086	2.00E-129	66.51
g14524.t1	nr: Viridae	AKI79945.1	putative rab-related GTPase [Acanthamoeba polyphaga mimivirus]	33.058	6.16E-09	54.55
	nr	XP_033216294.1	ADP-ribosylation factor-like protein 1 [Belonocnema treatae]	97.222	4.44E-127	99.44
g10826.t1	nr: Viridae	CAB4220686.1	DnaK Molecular chaperone [uncultured Caudovirales phage]	61.745	0	78.02
	nr	XP_017875606.1	heat shock 70 kDa protein cognate 5 isoform X1 [Ceratina calcarata]	87.5	0	96.08
g13251.t1	nr: Viridae	CAB4241741.1	chaperone protein DnaJ [uncultured Caudovirales phage]	42.857	5.04E-07	76.79
	nr	XP_033224436.1	translocation protein SEC63 homolog [Belonocnema treatae]	88.111	0	93.79

REFERENCES

1. C.-H. Lue, A. C. Driskell, J. Leips, M. L. Buffington, Review of the genus Leptopilina (Hymenoptera, Cynipoidea, Figitidae, Eucoilinae) from the Eastern United States, including three newly described species. *Journal of Hymenoptera Research* **53**, 35 (2016).
2. M. Schilthuizen, G. Nordlander, R. Stouthamer, J. Van Alphen, Morphological and molecular phylogenetics in the genus Leptopilina (Hymenoptera: Cynipoidea: Eucoilidae). (1998).
3. T. A. Schlenke, J. Morales, S. Govind, A. G. Clark, Contrasting infection strategies in generalist and specialist wasp parasitoids of Drosophila melanogaster. *PLoS Pathog* **3**, 1486-1501 (2007).
4. T. M. Rizki, R. M. Rizki, Y. Carton, Leptopilina heterotoma and L. boulardi: strategies to avoid cellular defense responses of Drosophila melanogaster. *Exp Parasitol* **70**, 466-475 (1990).
5. E. S. Keebaugh, T. A. Schlenke, Insights from natural host-parasite interactions: The Drosophila model. *Developmental and Comparative Immunology* **42**, 111-123 (2014).
6. D. Colinet *et al.*, Extensive inter- and intraspecific venom variation in closely related parasites targeting the same host: the case of Leptopilina parasitoids of Drosophila. *Insect Biochem Mol Biol* **43**, 601-611 (2013).
7. J. Goecks *et al.*, Integrative approach reveals composition of endoparasitoid wasp venoms. *PloS one* **8**, e64125 (2013).
8. M. E. Heavner *et al.*, Novel Organelles with Elements of Bacterial and Eukaryotic Secretion Systems Weaponize Parasites of Drosophila. *Current biology : CB* **27**, 2869-2877.e2866 (2017).
9. D. Di Giovanni *et al.*, A behavior-manipulating virus relative as a source of adaptive genes for Drosophila parasitoids. *Mol Biol Evol*, (2020).
10. B. Wey *et al.*, Immune Suppressive Extracellular Vesicle Proteins of Leptopilina heterotoma Are Encoded in the Wasp Genome. *G3 (Bethesda)* **10**, 1-12 (2020).
11. J. Huang *et al.*, Two novel venom proteins underlie divergent parasitic strategies between a generalist and a specialist parasite. *Nat Commun* **12**, 234 (2021).
12. C. Small, I. Paddibhatla, R. Rajwani, S. Govind, An introduction to parasitic wasps of Drosophila and the antiparasite immune response. *J Vis Exp*, e3347 (2012).
13. A. Conesa *et al.*, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676 (2005).
14. S. Götz *et al.*, High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* **36**, 3420-3435 (2008).
15. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
16. P. Jones *et al.*, InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).
17. U. Consortium, UniProt: a hub for protein information. *Nucleic Acids Res* **43**, D204-212 (2015).
18. S. Carbon *et al.*, The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research* **47**, D330-D338 (2019).
19. H. Kalra *et al.*, Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation. *PLoS Biol* **10**, e1001450 (2012).
20. M. Pathan *et al.*, A novel community driven software for functional enrichment analysis of extracellular vesicles data. *J Extracell Vesicles* **6**, 1321455 (2017).
21. M. Pathan *et al.*, Vesiclepedia 2019: a compendium of RNA, proteins, lipids and metabolites in extracellular vesicles. *Nucleic Acids Res* **47**, D516-D519 (2019).
22. W. Li *et al.*, The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res* **43**, W580-584 (2015).
23. L. Kall, A. Krogh, E. L. Sonnhammer, A combined transmembrane topology and signal peptide prediction method. *Journal of molecular biology* **338**, 1027-1036 (2004).
24. L. Kall, A. Krogh, E. L. L. Sonnhammer, Advantages of combined transmembrane topology and signal peptide prediction - the Phobius web server. *Nucleic Acids Research* **35**, W429-W432 (2007).
25. H. Nielsen, Predicting Secretory Proteins with SignalP. *Methods Mol Biol* **1611**, 59-73 (2017).
26. J. J. Almagro Armenteros *et al.*, SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420-423 (2019).

27. E. L. Sonnhammer, G. von Heijne, A. Krogh, A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**, 175-182 (1998).
28. A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* **305**, 567-580 (2001).
29. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410 (1990).
30. M. Johnson *et al.*, NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**, W5-9 (2008).
31. C. Camacho *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
32. T. Schallus *et al.*, Malectin: A Novel Carbohydrate-binding Protein of the Endoplasmic Reticulum and a Candidate Player in the Early Steps of Protein N-Glycosylation. *Molecular Biology of the Cell* **19**, 3404-3414 (2008).
33. C. Galli, R. Bernasconi, T. Soldà, V. Calanca, M. Molinari, Malectin Participates in a Backup Glycoprotein Quality Control Pathway in the Mammalian ER. *PloS one* **6**, e16304 (2011).
34. Q.-P. Yang *et al.*, Subcellular distribution of endogenous malectin under rest and stress conditions is regulated by ribophorin I. *Glycobiology* **28**, 374-381 (2018).
35. R. D. Fons, B. A. Bogert, R. S. Hegde, Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *Journal of Cell Biology* **160**, 529-539 (2003).
36. S. Pfeffer *et al.*, Dissecting the molecular organization of the translocon-associated protein complex. *Nat Commun* **8**, 14516 (2017).
37. R. Aber, W. Chan, S. Mugisha, L. A. Jerome-Majewska, Transmembrane emp24 domain proteins in development and disease. *Genetics Research* **101**, e14 (2019).
38. S. Mathivanan, H. Ji, R. J. Simpson, Exosomes: extracellular organelles important in intercellular communication. *J Proteomics* **73**, 1907-1920 (2010).
39. E. Nolte-'t Hoen, T. Cremer, R. C. Gallo, L. B. Margolis, Extracellular vesicles and viruses: Are they close relatives? *Proceedings of the National Academy of Sciences* **113**, 9155-9161 (2016).
40. G. van Niel, G. D'Angelo, G. Raposo, Shedding light on the cell biology of extracellular vesicles. *Nature Reviews Molecular Cell Biology* **19**, 213-228 (2018).
41. H. Stenmark, V. M. Olkkonen, The Rab GTPase family. *Genome biology* **2**, REVIEWS3007-REVIEWS3007 (2001).
42. M. E. Heavner *et al.*, Partial venom gland transcriptome of a Drosophila parasitoid wasp, Leptopilina heterotoma, reveals novel and shared bioactive profiles with stinging Hymenoptera. *Gene* **526**, 195-204 (2013).
43. D. R. Hoffman, Allergens in bee venom. III. Identification of allergen B of bee venom as an acid phosphatase. *J Allergy Clin Immunol* **59**, 364-366 (1977).
44. T. Grunwald *et al.*, Molecular cloning and expression in insect cells of honeybee venom allergen acid phosphatase (Api m 3). *Journal of Allergy and Clinical Immunology* **117**, 848-854 (2006).
45. P. R. d. Lima, M. R. Brochetto-Braga, Hymenoptera venom review focusing on Apis mellifera. *Journal of Venomous Animals and Toxins including Tropical Diseases* **9**, 149-162 (2003).
46. M. P. Dani, J. P. Edwards, E. H. Richards, Hydrolase activity in the venom of the pupal endoparasitic wasp, Pimpla hypochondriaca. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **141**, 373-381 (2005).
47. J.-Y. Zhu, G.-Y. Ye, S.-Z. Dong, Q. Fang, C. Hu, Venom of Pteromalus puparum (Hymenoptera: Pteromalidae) induced endocrine changes in the hemolymph of its host, Pieris rapae (Lepidoptera: Pieridae). *Archives of Insect Biochemistry and Physiology* **71**, 45-53 (2009).
48. D. C. De Graaf *et al.*, Insights into the venom composition of the ectoparasitoid wasp Nasonia vitripennis from bioinformatic and proteomic studies. *Insect Molecular Biology* **19**, 11-26 (2010).
49. O. Lamiable *et al.*, Cytokine Diedel and a viral homologue suppress the IMD pathway in <em>Drosophila</em&gt. *Proceedings of the National Academy of Sciences* **113**, 698 (2016).
50. C. Labrosse, P. Eslin, G. Doury, J. M. Drezen, M. Poirié, Haemocyte changes in D. Melanogaster in response to long gland components of the parasitoid wasp Leptopilina boulardi: a Rho-GAP protein as an important factor. *J Insect Physiol* **51**, 161-170 (2005).

51. C. Labrosse *et al.*, A RhoGAP protein as a main immune suppressive factor in the Leptopilina boulardi (Hymenoptera, Figitidae) - Drosophila melanogaster interaction. *Insect Biochem Molec* **35**, 93-103 (2005).
52. D. Colinet, A. Schmitz, D. Depoix, D. Crochard, M. Poirié, Convergent use of RhoGAP toxins by eukaryotic parasites and bacterial pathogens. *PLoS Pathog* **3**, e203 (2007).
53. F. Dong, J. Wang, R. Deng, X. Wang, Autographa californica multiple nucleopolyhedrovirus gene ac81 is required for nucleocapsid envelopment. *Virus Research* **221**, 47-57 (2016).

CHAPTER 2

Transgenic and RNA Interference studies of a generalist Drosophila parasitic wasp protein p40 reveals key virulence function

ABSTRACT:

Parasitic wasps *Leptopilina heterotoma* produce spiked particles, called Mixed Strategy Extracellular Vesicles (MSEVs), that are 300 nm in diameter and produced in a venom gland. Previous molecular profiling suggested a complex vesicular nature with more than 400 proteins and an absence of nucleic acid in these particles. However, these particles are full of proteins of unknown function. One such protein is a spike/surface protein called “p40”. Previous work showed that although p40 is encoded in a typical eukaryotic gene in the wasp genome, it is predicted to be a transmembrane MSEV protein. Modeling studies predicted that p40 folds similarly to the Type III secretion system protein IpaD of the Gram-negative bacteria *Shigella*. Previous studies also showed that pre-treating MSEVs with anti-p40 antibodies prevents the venom’s action on host lamellocytes, suggesting that p40 plays a role in *L. heterotoma*’s strategy to lyse host lamellocytes. Here we show that transgenic expression of a full length p40-RFP fusion protein localized to plasma membranes of the larval host’s fat body, plasmatocytes, and lamellocytes. Furthermore, a p40 mRFP-fusion construct with the putative transmembrane domain deleted caused loss of membrane localization; secretion of this protein is evidenced by sequestering of the mRFP signal within the larval pericardial cells. Expression of the secreted p40 transgenic construct in hosts prevented encapsulation of the less successful *L. victoriae* wasps. p40 transcripts and protein are expressed in the venom gland prior to wasp eclosion. Knockdown of p40 in developing female wasps did not affect normal wasp development but increased parasite death rate through encapsulation when compared to untreated or GFP knockdown controls. These results confirm that p40 can act on host hemocytes independently of other MSEV proteins and underscore the importance of p40’s significant role in *L. heterotoma*’s unique method of active immune suppression in their fly hosts.

INTRODUCTION:

During their work on *Leptoplilina heterotoma (Lh)*, Rizki and Rizki first examined the action of *Lh* venom *on D. melanogaster* larval hemocytes (*1*). They hypothesized that a factor called lamellolysin was responsible for causing lamellocytes, normally large and flat cells, to turn into a long, thin, and spindly morphology called 'bipolar' but they never identified the protein in *Lh* venom that possessed the lamellolysin activity (*2*). This same activity, present in *Lh*, was also found in the venom of its sister species *L. victoriae* (*Lv*) but not in the related species *L. boulardi* (*Lb*) (*3, 4*). Like Lh, the Lv venom gland also produces spiked VLPs (*4*). The *Lh* and *Lv* venom extracts also promote apoptosis of plasmatocytes (*5*).

The lamellolysin activity was localized to particles within *Lh* venom that the Rizkis called virus-like particles (VLPs) (*6*). Proteomic analysis of VLPs in our lab revealed an abundance of proteins involved in cellular processes and immune suppression with an enrichment profile similar to extracellular vesicles (*7, 8*). This analysis did not reveal the presence of any known viral coat proteins (*7, 8*). Based on these analyses, we proposed renaming *Lh* VLPs to mixed-strategy extracellular vesicles (MSEVs) (see Chapter 1 for more details).

Associated with *Lh* MSEVs is p40, a highly abundant protein (*9*). Originally identified in extracts of *Lh* and *Lv* venom gland, p40 was found to be secreted from the secretory cells of the long gland portion of the venom gland (*4, 9, 10*). However, transcription and translation of p40 in pre-eclosion wasp stages has not been studied to date.

In *ex vivo* assays using lamellocytes from mutant animals, venom gland extracts from both *Lh* and *Lv* promote lamellocyte lysis and bipolar morphology (*2, 5*) although the effects of the *Lv* venom are weaker (*4*). In the same *ex vivo* bipolar assays, pre-treating venom extracts with anti-p40 antibodies protects lamellocytes from the lytic effects of *Lh* and *Lv* MSEVs (*9*). Inmmuno-electron microscopy studies with *Lh* and *Lv* MSEVs showed that p40 is associated with the MSEV membrane and spike tips (*9*). Based on its localization to spike surfaces and tips and MSEV spike interaction with lamellocyte surfaces, it was proposed that MSEV-associated p40 mediates MSEV interactions with lamellocytes and is hence may be important to *Lh* and *Lv* virulence (*9*). Significantly, p40 is absent in the genomes and abdominal transcriptomes of *Lb* (*7, 11*), suggesting that it contributes to *Lh*'s unique lamellocyte lysis activity.

p40's domain analysis and homology modeling suggested it to be bound to the MSEV membrane as its predicted sequence structure contains (a) a signal sequence, (b) a central domain important for p40 function, and (c) a transmembrane domain (*7*). Homology modeling the protein's central domain (without its N-terminal signal sequence and C-terminal trans-membrane domain) revealed that at the 3D level, p40 is structurally folded in a way that is similar to IpaD/SipD proteins, even though no significant similarity is found in their primary sequences (*7*). Modeling results also showed that like SipD/IpaD, p40 shares structural similarities to the actin binding proteins, spectrin and plectin (*7*). With these structural similarities, we hypothesized that p40 may function similarly to bacterial IpaD. IpaD can cause apoptosis of macrophages and B-lymphocytes in humans (*12, 13*). IpaD also interacts with macrophage and epithelial cell membranes(*14, 15*). Mutations in key residues in IpaD prevent its interaction with host cells, secretion of effector proteins, or bacterial entry into host cells (*14*). Here we test the hypotheses that (a) transgenic expression of p40 in wild type flies might alter host hemocyte behavior and inhibit encapsulation induced by less virulent wasps; and; (b) knockdown of p40 in wasps would negatively affect the success rate of parasitism.

Utilizing the UAS-GAL4 system in *D. melanogaster*, we found that a full length p40 construct localizes to cell membranes similarly to a membrane-targeted myristoylated mRFP protein and a membrane-associated mCD8-GFP, whereas a p40 construct without its transmembrane domain appears to be secreted into hemolymph and sequestered in cells responsible for filtering hemolymph. Fly larvae expressing transgenic *Lh* p40 showed reduced encapsulation of the related wasp, *Lv*. Unlike its host, *Lh* does not offer forward genetic tools for in vivo studies. However, reverse genetic approaches such as RNA interference (RNAi) have been shown successful for "loss-of-gene function" studies. Experiments with other wasps have shown that dsRNA administration through microinjection or feeding is effective in this regard (*16-25*). We show that knockdown of p40 increased encapsulation of parasite eggs significantly compared to GFP dsRNA and uninjected controls confirming its important role in *Lh* virulence. These studies illustrate how a novel immune suppression factor can greatly change the success of a parasite, thus broadening its niche and success compared to its evolutionary cousins.

MATERIALS AND METHODS:

Insects: Isogenized *Lh* 14 wasps (*26*) and *Lv (4)*, were raised on the *y w* strain of *D. melanogaster* that were reared on standard cornmeal, yeast, and agar fly food at 25°C as described (*27*). Adult wasps were collected from parasitized hosts, 25 days after infection. Male and female wasps were stored on fly food with 70% honey on "buzz" plugs. Drosophila crosses were performed at 27 °C to express transgene at high levels.

Transgenic fly lines: Constructs for full-length p40 containing predicted p40 domains and variations with domain deletions were cloned into the pENTR Gateway Vector (Invitrogen). Clones were then used for transferring constructs into destination vectors from the Drosophila Gateway Destination Vector collection (Cornell, https://emb.carnegiescience.edu/drosophila-gateway-vector-collection (*28-31*)) to express constructs using the UAS-GAL4 system and to tag the C-terminal with mRFP. Primers used to develop each construct are listed in Table 1. Recombination between entry and destination vector was performed using the LR-Clonase II kit (Invitrogen) following included protocols. Recombination products were transformed into TOP10B competent *E. coli* cells (Invitrogen) using heat shock at 42 °C for 30 seconds, recovery at 37 °C for 1 hour, and then plated on Luria-Bertani Agar with 100 µg/mL ampicillin overnight at 37 °C. Colonies were screened with polymerase chain reaction and restriction digest using PstI to verify successful insertion of construct in destination vector. Inserts in selected plasmids were sequenced using primers in Table 1 (GENEWIZ) to confirm the proper reading frame. Plasmids were prepared for injection using Qiagen Midi Prep kit (Qiagen). Injection of vectors with P-element helper plasmid into strain *y w D. melanogaster* embryos was performed by Rainbow Gene Transgenic Flies. Transgenic insertions were mapped to fly chromosomes and a balanced or homozygous strain was produced using standard crosses. Recombinant lines containing the *Cg-GAL4* P-element transgenic insert (*32*) and all four *UAS-p40* constructs or *UAS-myr-mRFP* were made using standard genetic crosses (Table 2).

Verification of transgenic p40 expression: Transcription: *Cg-Gal4* and the *UAS-p40* construct larvae were washed twice in deionized water prior to processing. Total RNA was extracted using 100 µL of Trizol (Invitrogen) following manufacturer's protocols. RNA was resuspended in 0.1% DEPC treated water and treated with DNase I to remove contaminating DNA (Thermo Fisher Scientific). The RNA concentration

was determined by NanoDrop (Thermo Fisher). cDNA was synthesized using Proto-Script First Strand cDNA Synthesis Kit (New England Biolabs).

PCRs were performed with Taq polymerase (gift of C. Li lab, CCNY), the cDNA templates, PCR buffer (300 mM Tris HCl pH 9.5, 75 mM $(NH_4)_2SO_4$, 10 mM $MgCl_2$) and deoxyribonucleotides (0.2 mM; Thermo Fisher Scientific). The PCR products were resolved on a 1% agarose gel in Tris acetic acid EDTA buffer (40 mM Tris HCl pH 7.6, 20 mM acetic acid, 1mM EDTA pH 8.0). Ethidium bromide (Sigma Aldrich)-stained gels were visualized on an ultraviolet Trans-Illuminator (UVP) and gel images were taken using the DigiDocIt Imaging System (UVP). Gel images were processed in Adobe Photoshop for clarity only. Primers are listed in Table 2.

Translation: Protein extracts were made from third instar larval progeny of *da-GAL4* (*33*) flies crossed to *UAS-p40* construct flies. Larvae were washed in water and 70% ethanol before crushing in insect lysis buffer (1x PBS pH 7.4, 1% Triton-X 100, 25 mM Tris-HCl pH 7.4, 1x mammalian proteinase inhibitor (VWR)). Samples were centrifuged for 15 min at 13,000 X revolutions per minute (RPM), 4°C. The supernatant was pipetted into clean microfuge tubes and used for analysis. Samples were stored at -80°C until use.

SDS page gels were run using 6% stacking gel and 12% resolving gel. For Western blot, proteins were transferred onto a nitrocellulose membrane with Tris-Glycine buffer for 1 hr at 70V. Membrane was blocked with 3% (w/v) bovine serum albumin (BSA) with 0.1% Tween 20 in PBS for 1 hr. Anti-p40 (*9*) or anti-RFP (Rockland Immuno) antibodies were diluted 1:1000 in block and incubated with membrane for 1 hr with shaking. Membrane was washed before incubation with alkaline phosphatase conjugated anti-mouse or anti-rabbit secondary antibody diluted 1:200 in block for 1 hr. Color development and band detection was performed with 125 µg/mL 5-bromo-4-chloro-3-indolyl phosphate (BCIP) with 250 µg/mL nitro blue tetrazolium chloride (NBT) in NTM 9.5 (1µM NaCl, 1µM TRIS pH 9.5, 5nM $MgCl_2$). Color development was stopped by washing membrane in deionized water.

Subcellular localization of p40-RFP and in vivo effects: Tissues were dissected from animals with the following genotypes: *w; Cg-GAL4* (called *Cg*-GAL4), *w; Cg-GAL4, UAS-p40^{FL}* (Cg>p40^{FL}), *w; Cg-GAL4, UAS-p40^{noTM}* (Cg>p40^{noTM}), *w; Cg-GAL4, UAS-p40^{CD}* (Cg>p40^{CD}), *w; Cg-GAL4, UAS-p40^{noSP}/CyO-GFP* (Cg>p40^{noSP}). These genotypes and designations are listed in Table 2.

Larvae were washed twice in deionized water and kept in phosphate buffered saline (PBS) prior to dissection. Lymph gland and hemolymph smears were allowed to dry for 1 hr at 25°C before fixation with 4% paraformaldehyde for 15 min. Tissues were rinsed three times using PBS before staining actin with Alexa Fluor 488 for 30 min and Hoechst for 15 min. Tissues were rinsed three times with PBS before application of VectaShield mounting media. Fat body tissue samples were dissected into PBS and kept hydrated while following the same fixation, staining, and mounting procedures listed for other tissues. Confocal microscopy was performed on LSM 710 confocal microscope (Zeiss). Images were processed using Zen Blue.

Transgenic p40 encapsulation and cell aggregate assay: *Lv* wasps were provided Cg>p40^{noTM}, Cg>p40^{FL}, and Cg>myr-mRFP larvae as hosts and allowed to parasitize at 27°C. Three days later, larvae were dissected to determine presence/absence of wasp embryos/larvae, encapsulation of wasp embryos/larvae, and to count number of melanized cell aggregates (i.e., hemocytes which were not surrounding a wasp embryo/larva). Percentage encapsulation was calculated as the total number of infected larvae that had at least one encapsulated wasp embryo/larva divided by the total number of infected larvae. Percentage presence of cell aggregates was calculated as the total number of infected larvae that had at least one melanized cell aggregate divided by the total number of infected larvae. Welch's one-tailed t-test assuming unequal variances was performed to determine statistical significance between conditions.

***In situ* hybridization:** Sense and anti-sense digoxigenin (DIG)-labeled RNA probes were made using the DIG RNA labeling kit from ROCHE and SP6 RNA polymerase from Thermo Fisher following provided protocols.

Protocol for hybridization of probe was modified from protocols specified in (*34*). To prepare tissue for hybridization, *L. heterotoma* venom glands were dissected in 1X Phosphate Buffered Saline pH 7.4 (PBS) and then fixed in 4% paraformaldehyde. Samples were then permeabilized for 20 min using 0.3% Triton-X in PBS (PBT). Samples were washed in 0.1% PBT before incubation at 42°C for 1 hr in hybridization buffer (50% Formamide, 5X SSC, 1 mg/mL ARN torula, 50 µg/mL Heparine, 2% Roche Blocking Reagent, 5 µM EDTA, 0.1% CHAPS, 0.1% Tween20, DEPC treated H_2O). Samples were incubated with DIG-labeled RNA probes at 65°C overnight. Samples were then washed in hybridization

buffer and 0.1% PBT to the following ratios: 3:1, 1:1, 1:3. Samples were then blocked in 10 mg/mL BSA with 0.1% Tween in PBS (PTW) for 1 hr. Anti-DIG antibody from DIG-labeling kit (ROCHE) was diluted 1:2000 in block and incubated with samples for 2 hrs. Samples were washed in 0.1% PBT and PBS before color development with 125 µg/mL BCIP with 250 µg/mL NBT in NTM 9.5 (1µM NaCl, 1µM TRIS pH 9.5, 5nM $MgCl_2$). Color development was stopped by washing samples in 1X PBS. Samples were then mounted in 50% (v/v) glycerol in PBS and imaged at 20x magnification on an Axioskop 2 Plus (Zeiss).

Microinjection of dsRNA: Constructs for double-stranded RNA for *cinnabar* and *p40* were prepared using primers in Table 1 and cloned into pCRII TOPO (Invitrogen) prior to cloning into L4440 expression vector (Caenorhabditis Genetics Center). L4440:GFP was used for expression of GFP dsRNA as negative control (gift from C. Li, CCNY). Constructs were cloned into *E. coli* strain HT115(DE3) (Caenorhabditis Genetics Center) for expression.

Overnight bacterial cultures were diluted 1:100 in 50 mL of 2x Yeast Tryptone media and grown to OD 0.4 at 37°C with shaking. Cultures were then induced with 0.4mM isopropyl β-D-1-thiogalactopyranoside for 4 hrs. Cultures were pelleted at 6000 RPM for 5 min at 4°C. Extraction and purification of dsRNA was performed as described in (*16*) and then resuspended in 0.1% diethyl pyrocarbonate treated water. dsRNA concentration was determined via NanoDrop (ThermoFisher). dsRNA aliquots were stored at -80°C until use.

For microinjection, wasp larvae were dissected out of host pupal cases at 10 days-post infection (dpi) and placed on 2% agar plates made with 1x PBS and supplemented with 1% Methyl 4-hydroxybenzoate (Sigma). Injections were performed using a Pico-Injector (Harvard Apparatus). All injections were approximately 1 nL. Knockdown of *cn* was done with injections of 700 ng/µL *cinnabar* dsRNA. Negative control was 700 ng/µL *GFP* dsRNA. Knockdown of *p40* was done using injections of 350, 700, or 1000 ng/µL *p40* dsRNA. Negative dsRNA control was 350, 700, or 1000 ng/µL of *GFP* dsRNA. dsRNA was denatured at 95°C for 5 min before reannealing at room temperature before use in injections. Wasps that were dissected from pupal case but not injected with any dsRNA (uninjected) were used as wounding/stress negative control. 1% blue food coloring (final concentration in solution, McCormick brand) was added to dsRNA and used to trace injection. Injected wasps were incubated at

25°C until adulthood. Knockdown of *cinnabar* was scored via microscope observation on a Leica MKFLIII (Leica). Survival statistics for injection of *p40* dsRNA, *GFP* dsRNA, and uninjected control are listed in Table 3.

Encapsulation assay: dsRNA injected and uninjected wasps were used to infect *y w D. melanogaster* larvae. The number of female wasps used in each infection repeat was recorded in order to calculate encapsulations per female. An average of three females were used per egglay. Each egglay used for a set of “knockdown” (KD) females was counted as a repeat. Infection repeats were incubated at 25°C. All emerging flies and wasps from each repeat were counted approximately ten day’s post injection/dissection. Adult flies were scored for presence/absence of capsules. Representative images of affected larvae and flies were taken on a Leica MKFLIII (Leica). Encapsulations were normalized to number of encapsulations per female. F-test for variances was performed to confirm if variance between conditions was unequal prior to statistical significance testing. Welch’s one-tailed t-test assuming unequal variances was performed to determine statistical significance between conditions. If variance is equal, Welch’s t-test performs as if Student’s t-test.

RESULTS:

Transgenic expression of p40 in flies

In previous experiments, the central domain of p40 (p40 CD) was expressed in bacteria to test if exposure to target hemocyte cells ex vivo might alter their shape and viability (*35*). The rationale was that as IpaD interacts with macrophage and epithelial cell membranes to allow the Type III Secretion System needle to inject proteins (*12, 14, 15*), a similar 3D structure, even in solution, may help lyse target host cell membranes, when presented from the outside (*35*). This bacterially expressed p40 CD was able to induce significant lamellocyte shape change (*35*).

To verify these results from bacterially derived p40 in an *ex vivo* system, we turned to transgenic flies whose cells would add post-translational modifications to p40, and also allow testing its predicted membrane localization. For this, we designed four monomeric RFP-tagged (mRFP-tagged) p40 constructs for expression in Drosophila under the UAS-GAL4 system: full length p40 ($p40^{FL}$), p40 without transmembrane domain ($p40^{noTM}$), p40 central domain ($p40^{CD}$), and p40 without signal peptide ($p40^{noSP}$) (Fig 1A). Transgenic fly lines were checked for expression of constructs. While all constructs expressed transcripts detectable with construct specific primers (Fig 1B), only two transgenes ($p40^{FL}$ and $p40^{noTM}$) expressed detectable protein level in Western blot analysis (Fig 1C).

Fusion constructs were first expressed in the larval fat body (Fig 1E-G) using *Collagen* (*Cg)-GAL4* as the fat body is large and made up of a single layer of endopolyploid cells, where it is easy to observe intracellular protein localization. (The fat body is not a known target of wasp venom or p40 and we did not expect specific biological effects of p40 expression in it.) Fat bodies from *Cg-GAL4* animals without the UAS-transgene were used as negative control (Fig 1D). Expression of myristoylated mRFP (myr-mRFP) was used to label membrane with mRFP as another control (Fig 1E) (*36*). We observed that $p40^{FL}$ clearly localized to the plasma membrane, as indicated by the mRFP signal outlining the cell boundary. Punctate staining in the fat body cell cytoplasm is likely to be vesicular. Colocalization of $p40^{FL}$ mRFP signal with mCD8-GFP (which is known to insert into membranes due to mCD8's transmembrane domain (*37, 38*)) also occurred in lymph glands when expressed with *Hemese (He)-GAL4*, which further supports p40's membrane association (Fig 2B, Fig 3E).

Finally, fat body cells and plasmatocytes expressing $p40^{FL}$ appear to have denser F-actin signal (Fig 1F, Fig 3C-C'') compared to controls (Fig 1D and E, Fig 3B-B''). F-actin and the associated modeling protein spectrin are associated with cell membranes and involved in several cellular processes including motility and signaling (*39-43*). While we do hypothesize that p40 CD may fold similarly to spectrin, we cannot confirm the orientation of membrane associated $p40^{FL}$ and if the p40 CD is within the cytoplasm to interfere with actin modeling (Fig 3A). However, the density of the actin staining may indicate that a cytoplasmic region of p40 is able to interact with actin.

As expected from the Western blot results (Fig 1C), we failed to detect an mRFP signal in the fat bodies of Cg>$p40^{CD}$ and Cg>$p40^{noSP}$ animals, possibly due to their rapid degradation (not shown). In Cg>$p40^{noTM}$ fat body cells the mRFP signal was also not detected (Fig 1G), however, pericardial cells (PCs) in these animals showed a clear, but punctate, mRFP signal which was not detected in the Cg>myr-mRFP control PCs ($p40^{noTM}$ total n = 31/33 larvae, myr-mRFP n = 0/25 across three rounds of dissections, Fig 1E''' and 1G'''). This result is consistent with the model that p40 lacking the transmembrane domain is secreted and taken up by PCs, as PCs are pinocytotic (*Cg-GAL4* is not expressed within PCs ;Fig 1E''' (*44-47*)). Thus, sequestration of $p40^{noTM}$ within these cells is possible if it is secreted.

We were surprised to find that as in the Cg> $p40^{noTM}$ larval pericardial cells, the Cg>$p40^{FL}$ animals also showed uptake of the RFP signal into their PCs (Fig 1F'''). However, unlike the fat body cells, the Cg>$p40^{FL}$ p40 signal is not localized to the pericardial cell membrane, but instead it is punctate, just like the $p40^{noTM}$ signal. This uptake in Cg>$p40^{FL}$ was observed multiple times in three separate dissections ($p40^{FL}$ total n = 23/28 across three dissections).

To observe if there were any differences in localization in hemocytes (plasmatocytes and lamellocytes) compared to fat body localization, we expressed p40 constructs in lamellocytes using *misshapen (msn)-GAL4* and in plasmatocytes via *Cg-GAL4* (Fig 2B-E, expression of $p40^{noTM}$ not shown). As seen in fat body cells, $p40^{FL}$ continued to localize to cell membranes of both plasmatocytes and lamellocytes. The $p40^{noTM}$ proteins were not detected from the cytoplasm and membrane of expressing cells (data not shown). Given the expected expression patterns and subcellular localization of the $p40^{FL}$ and $p40^{noTM}$ proteins in transgenic fat bodies and hemocytes, we performed functional assays.

Transgenic p40 proteins appear to be active

We hypothesized that p40 may interact with hemocytes differently depending on presentation (Fig 2A). Effects may occur when $p40^{FL}$ is internalized during endocytosis when presented on the cell membrane. Secreted $p40^{noTM}$ may also be taken up by the very same plasmatocytes that produce the construct and cause a biological effect. Other effects may arise from interaction of individual cells (plasmatocyte-plasmatocyte, lamellocyte-lamellocyte, plasmatocyte-lamellocyte, Fig 3A). To observe effects of p40 expression on lamellocytes, we expressed constructs using *msn*-GAL4 and via *Cg*-GAL4 (Fig 3B-E, expression of $p40^{noTM}$ not shown). We also utilized an over-proliferative hemocyte background ($hop^{Tum\text{-}l}$) in order to ensure that both plamatocytes and lamellocytes were present. The $hop^{Tum\text{-}l}$ background causes melanized aggregates of hemocytes which are referred to as tumors. The production of these tumors during larval stages decreases viability of the adult fly. We hypothesized that if transgenic p40 was active against hemocytes, it should reduce tumor burden and/or increase viability. A significant effect on tumors and viability was seen when p40 was expressed in plasmatocytes or lamellocytes with decreased tumor burden and increased viability (Sevilla, Vashist, and Govind, unpublished work). These effects of p40 expression appear to be cell-specific as no deleterious effects were observed when p40 was expressed in the eye using *eyeless*-GAL4 and *GMR*-GAL4 (data not shown).

Expression of transgenic p40 reduces encapsulation of *L. victoriae*

Given the potential effects on tumors and viability in an over-proliferative hemocyte background, we hypothesized that transgenic p40 expression would reduce the encapsulation of the *Leptopilina* spp. *L. victoriae*, which is known to have p40, but is less successful leading to higher incidence of encapsulation in comparison to *Lh (4)*. After allowing *L. victoriae* to parasitize hosts expressing either myr-mRFP, $p40^{FL}$, or $p40^{noTM}$ under the control of *Cg*-GAL4, we found hosts expressing $p40^{noTM}$ had no encapsulations of *L. victoriae* embryos or larvae, a stark contrast to myr-mRFP control and $p40^{FL}$ which still had an average of encapsulation occurrence of 36% and 61% respectively (Fig 4A). Expression of myr-mRFP or $p40^{FL}$ created no significant difference in percentage of larvae with encapsulations despite the difference in averages (Fig 4A).

Despite a striking difference between $p40^{noTM}$ and myr-mRFP controls, expression of $p40^{noTM}$ did not prevent activation of the host immune response. Instead, it significantly reduced the percentage of

hosts with melanized cell aggregates compared to myr-mRFP control (p-value = 2.57E-5, Fig 4B). No significant difference was observed between $p40^{FL}$ and myr-mRFP controls.

Knockdown of p40 increases parasite encapsulation

Colinet et al. (*23*) have shown that RNA interference (RNAi) is possible within the *L. boulardi* system using microinjection of double strand RNA (dsRNA). Using this as an example, we performed knockdown of the eye color gene *cinnabar* (*cn*) as a proof of concept.

Timing for injection of *cn* dsRNA was performed on wasps aged 10 days post infection (dpi) as performed in Colinet et al. as at this stage, eyes have not yet developed nor has any pigmentation developed (*23*). However, timing for p40 expression had not yet been determined. To find the optimal time point for injection of p40 dsRNA for knockdown, we performed *in situ* hybridization in venom glands to time expression of p40 mRNA. In venom glands from wasps (13-14 days post infection (dpi)) until up to 2 weeks dpi) we found p40 mRNA transcribed and translated even at the earliest time point (Fig 5 and 6). We were also concerned about the duration of knockdown and wondered if knockdown of p40 would persist long enough for wasps to reach adulthood and parasitize hosts. Colinet et al. reported that in their experiments, as detected by Southern blots of *L. boulardi* Rho-GAP (also known as LbGAP), knockdown lasts well into adulthood of female wasps (*23*). We thus expected that inducing knockdown of p40 early would lead to lasting effects into adulthood to see effects in attempts to suppress fly response to parasitism.

Injection of *cn* dsRNA caused a dark red eye color compared to the black/gray GFP dsRNA (Fig 7A and B) and thus we moved forward with p40 dsRNA injections. Despite attempts to minimize lethality, dissection from pupal case and injection of dsRNA caused wasp lethality or diapause (a stop in development) due to injury. Dissection from pupal case without injection (uninjected) had survival rates of 51% to 71% across all experiments (Table 3; total attempts n = 409 wasps). However, injection of p40 dsRNA or GFP dsRNA reduced survival rates ranging from 23% to 57% across all three dsRNA dosages (total p40 dsRNA attempts n = 534, total GFP dsRNA attempts n = 500, Table 3) and there was larger variability in survival. These numbers are similar to injection survival observed by Colinet et al. with for

Luciferase dsRNA control (65-68.7% survival, total n = 140) and for LbGAP dsRNA (58.7-60% survival, total n = 140) (*23*).

To examine the effects of p40 KD, we injected p40 dsRNA at 10 dpi, at three different concentrations (350 ng/μL, 700 ng/μL, and 1 μg/μL) prior to development of a venom gland in order to prevent early expression and studied encapsulation rates in hosts infected with “KD” wasps (Fig 7C-F). The rate of encapsulation per surviving female was determined as the total number of encapsulations in a vial per the number of females used in each vial. GFP dsRNA control did not affect encapsulation compared to uninjected control which was expected. When comparing injection of p40 dsRNA to uninjected wasps or GFP dsRNA injected wasps, we found significantly more encapsulations per female in all three concentrations of dsRNA for injection.

DISCUSSION AND CONCLUSION:

The history and previous work on *Lh*'s methods of host immune suppression are focused around understanding the mechanism of host hemocyte death and prevention of initiating the immune response (*1, 3, 4, 6-9, 26, 48*). In fact, the fat body immune response requires Spätzle secreted from hemocytes (*49*) and so destruction of host hemocytes may serve two functions: prevention of immune response and prevention of encapsulation. So far, no single protein in the *Lh* venom has been indicated to possess this "lamellolysin" activity proposed by the Rizkis (*1, 2, 6*). Here, we investigated the highly abundant protein p40 and establish its importance to the parasite strategy in hopes of determining if p40 is important or related to lamellolysin activity. Consistent with antibody inhibition assays with p40 (*9*), p40 knockdown in wasps protects lamellocytes from lysis.

In both hemocytes and fat body cells, we found our $p40^{FL}$-mRFP signal localized to cell membranes, supporting a membrane association. This signal overlapped with our membrane-associated actin stain as well as mCD8-GFP (Fig 1F-F'' and Fig 2A-A'') and is the same as the membrane localization of myr-mRFP (Fig 1G-G'' and Fig 2B-B''). The overlap between actin staining, mCD8-GFP, and $p40^{FL}$ may also be because the mRFP tag is on the cytoplasmic side while p40 CD may be presented to the outside of the cell. Contrary to this, we also found that expression of $p40^{FL}$ appears to cause denser actin staining around the cell membrane. While this does not directly support the modeling of p40 CD as similar to spectrin/plectin, it does potentially indicate that a domain of p40 that is within the cytoplasm is influencing actin localization. Although immune-gold EM of MSEVs for p40 supports external presentation of p40 on the outside of MSEVs (*9*), we do not have any evidence to support that the orientation of the transgenic construct will match that of MSEVs.

Domain architecture analysis of p40 is further supported by the presence of $p40^{noTM}$ in pericardial cells (Fig 1G'''). Pericardial cells are also known to act as the filtration system of the larval hemolymph (*45-47, 50*) and Cg-GAL4 does not cause expression of UAS constructs within pericardial cells (Fig 1E''' and (*32*)). Thus, without the transmembrane domain, the tagged $p40^{noTM}$ construct is unable to remain associated with membrane and appears to be secreted into the hemolymph only to be filtered out by the pericardial cells. However, we were particularly surprised to find $p40^{FL}$ within pericardial cells (Fig 1F'''),

especially considering the membrane association. Interestingly, pericardial cells have been shown to take up macromolecules and colloids via pinocytosis (*46*). This may indicate that $p40^{FL}$ may be shed with membrane or cellular debris by cells. Follow up experiments will check for presence of additional membrane markers, such as mCD8-GFP, from hemocytes and fat body cells in combination and separately through use of different GAL4 drivers specific to those tissues. We hypothesize that if $p40^{FL}$ is being shed with membrane due to cell death, blebbing, or other reasons, pericardial cells should show signal for both p40 and membrane marker.

p40-expressing hemocytes decreased tumor burden and increased viability of $hop^{Tum\text{-}l}$ flies (Sevilla, Vashist, and Govind, unpublished work). These data support the idea that the $p40^{FL}$ and $p40^{noTM}$ constructs, encoding the central domain, possess a biochemical activity, which when presented to target hemocytes' membranes, alter their function. The strongest effects were however observed when secreted p40 expressing hosts were infected with *Lv* (36-61% with myr-mRFP and $p40^{FL}$ compared to 0% with $p40^{noTM}$) (Fig 4A). The effectiveness of the secreted p40 on hemocytes could be due to the presence of other immune-suppressive factors in the *Lv* venom and VLPs that also worked in conjunction with Lv's p40 variant (*9*). Additionally, it was possible to clearly visualize the neutralizing effects of the transgene-derived secreted p40 in the context of a natural encapsulation reaction which retains the homeostatic resolution mechanisms intact (*51, 52*), whereas the $hop^{Tum\text{-}l}$ mutants exhibit constitutive cellular immunity and is unable to resolve hemocyte proliferation and differentiation due to loss of homeostasis (*53*).

Knockdown of p40 increases encapsulation (0.89-1.5 encapsulations per p40 dsRNA injected female on average compared to 0.2-0.5 encapsulations per GFP dsRNA injected or uninjected female) (Fig 7F). However, encapsulation results were quite variable. This variability may be due to technical and biological reasons arising from differences in volume of dsRNA administered, the precise site of delivery affecting the degree of knockdown, and inherent variation in wasps' infectivity in different biological replicates. Even though MSEVs contain a multitude of other proteins including infection and immunity proteins (*7, 8*), strong effects are seen with just p40 knockdown. This observation suggests that these other effector proteins might suppress other aspects of cellular or humoral immunity. The combined use of protein modeling, transgenic expression assays coupled with RNA interference adopted for p40 can also be used to study select candidate effector proteins. The absence of p40-like protein in the *Lb* venom

correlates with its inability to lyse host lamellocytes. This result suggests that the p40's activity on hemocytes might facilitate or hasten the effects of other factors to rapidly suppress host immunity. Like bacterial IpaD/SipD proteins, it may be able to interact directly with host membranes, mediating its own efficient entry and also of other effector proteins, although this remains to be shown. Like IpaD, *Lh* p40 may be able to promote hemocyte death, a result that is supported in preliminary results with bacterially expressed p40 in our lab. These results need to be confirmed in larvae expressing the secreted p40. If p40 alone is able to induce cell shape change and lysis of lamellocytes, it would be a candidate for the lamellolysin activity hypothesized by the Rizkis. Additional experiments with transgene derived proteins in *Drosophila* would confirm this function.

Our results suggest that p40 is an important arsenal in the *Lh*/*Lv* immune suppressive armament. Its presence and high activity in *Lh* may help explain *Lh*'s ability to succeed on a wide range of species. BLAST searches of *L. claivpes* genomic and transcriptomic sequences did not reveal the existence of a p40-like protein in *L. clavipes*, and like *Lb* its host range is limited, if not preferential (*54*), Characterization of the p40 variants in different *Lv* strains will expand our understanding of the evolution of this important protein in the *Leptopilina* clade.

ACKNOWLEDGEMENTS:

Initial cloning and primer design for p40 was performed by Dr. Johnny Ramroop. Gateway destination vectors for transgenic flies were obtained from the Drosophila Genomics Resource Center, supported by NIH grant 2P40OD010949-10A1. Western blot in Figure 1 was performed by Carlo Sevilla. Tumor burden assay and viability assay performed by Carlo Sevilla and Kushagra Vashist. Asif Siddiq helped with the preparation of figures. This work was supported by the following grants: NASA (NNX15AB42G), NSF (IOS-1121817, IOS-2022235).

FIGURES

Figure 1: Expression of p40 constructs follows domain architecture predictions. (A) Illustration of p40 domains within each designed construct. (B) cDNA of p40 fusion constructs expressed in transgenic fly larvae (*Cg-GAL4*) amplified by PCR and compared to controls amplified from corresponding plasmid templates. Lanes are as follows 0- Molecular weight ladder, 1 – $p40^{FL}$ plasmid control, 2 – $p40^{FL}$ cDNA, 3 – $p40^{noTM}$ plasmid control, 4 – $p40^{noTM}$ cDNA, 5 – $p40^{CD}$ plasmid control, 6 – $p40^{CD}$ cDNA, 7 – $p40^{noSP}$ plasmid control, 8 – $p40^{noSP}$ cDNA. (C) Western blot to detect mRFP-tagged p40 in fly larvae extracts. (D-G) Confocal microscopy of fly larval fat body showing localization of p40 mRFP signal with myristoylated mRFP as positive control and *Cg-GAL4* only as negative control. Scale bars for panels indicate 50 µm. Panel C: Courtesy Carlo Sevilla

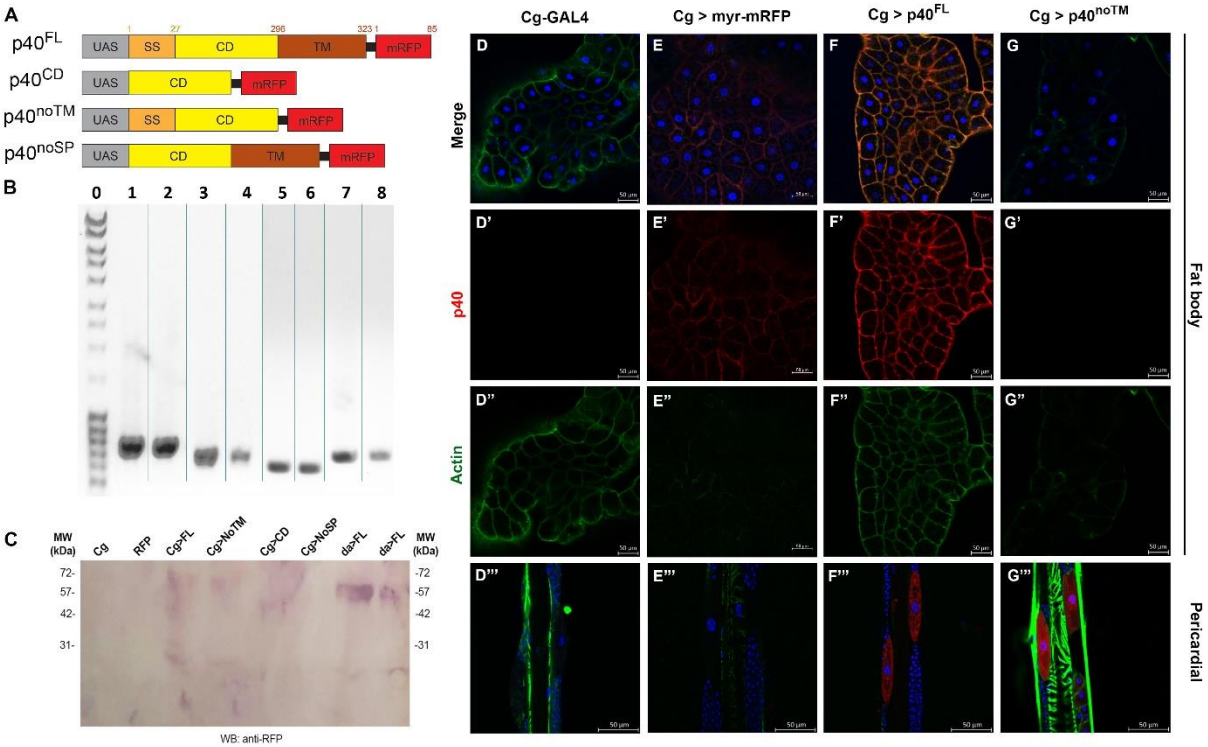


Figure 2: Expression of p40$^{FL.}$ In lymph gland hemocytes with *He-GAL4*. (A-A'') Control *He-GAL4*>mCD8-GFP lymph gland showing mCD8-GFP expression. (B-B'') Full length p40-mRFP fusion protein shows similar localization with membrane-associated mCD8-GFP. Dashed line indicates selected zoom and split channel regions. Scale bar in indicates 50 µm (A and B) and 20 µm (A', A'', B', B'').

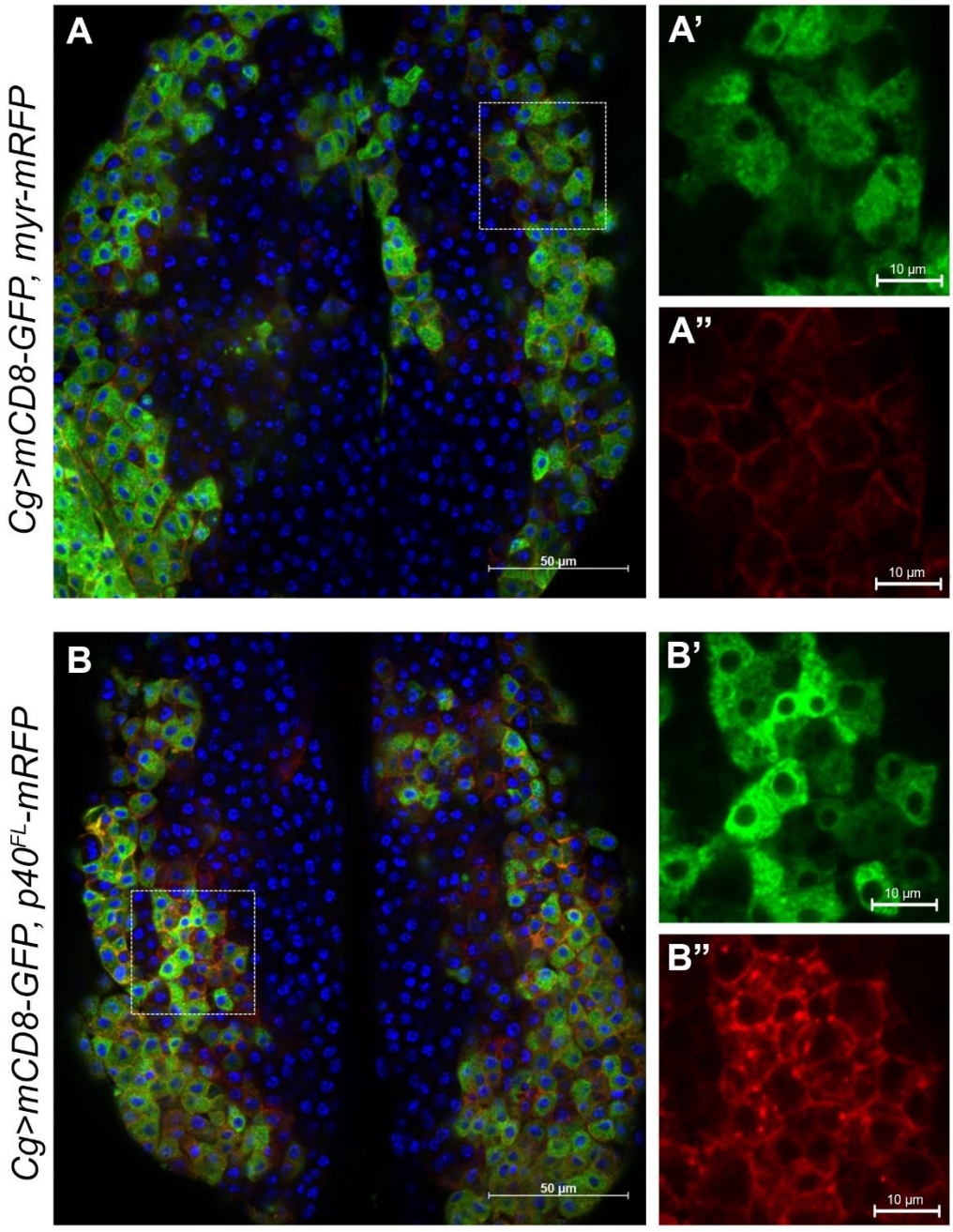


Figure 3: Presentation of p40 with and to hemocyte types. (A) p40 was expressed either in plasmatocytes (*Cg-GAL4*) or lamellocytes (*msn-GAL4*). This expression may cause autonomous (steps 1a-1c), or non-autonomous (steps 2-4) effects in hemocytes (that are not themselves producing the p40). In this experimental design, p40 is also expressed in the fat body via *Cg-GAL4*. (B-E) Localization of $p40^{FL}$ in plasmatocytes (C) and lamellocytes (E) shows cytoplasmic and membrane association of the protein. Scale bars indicate 20 μm.

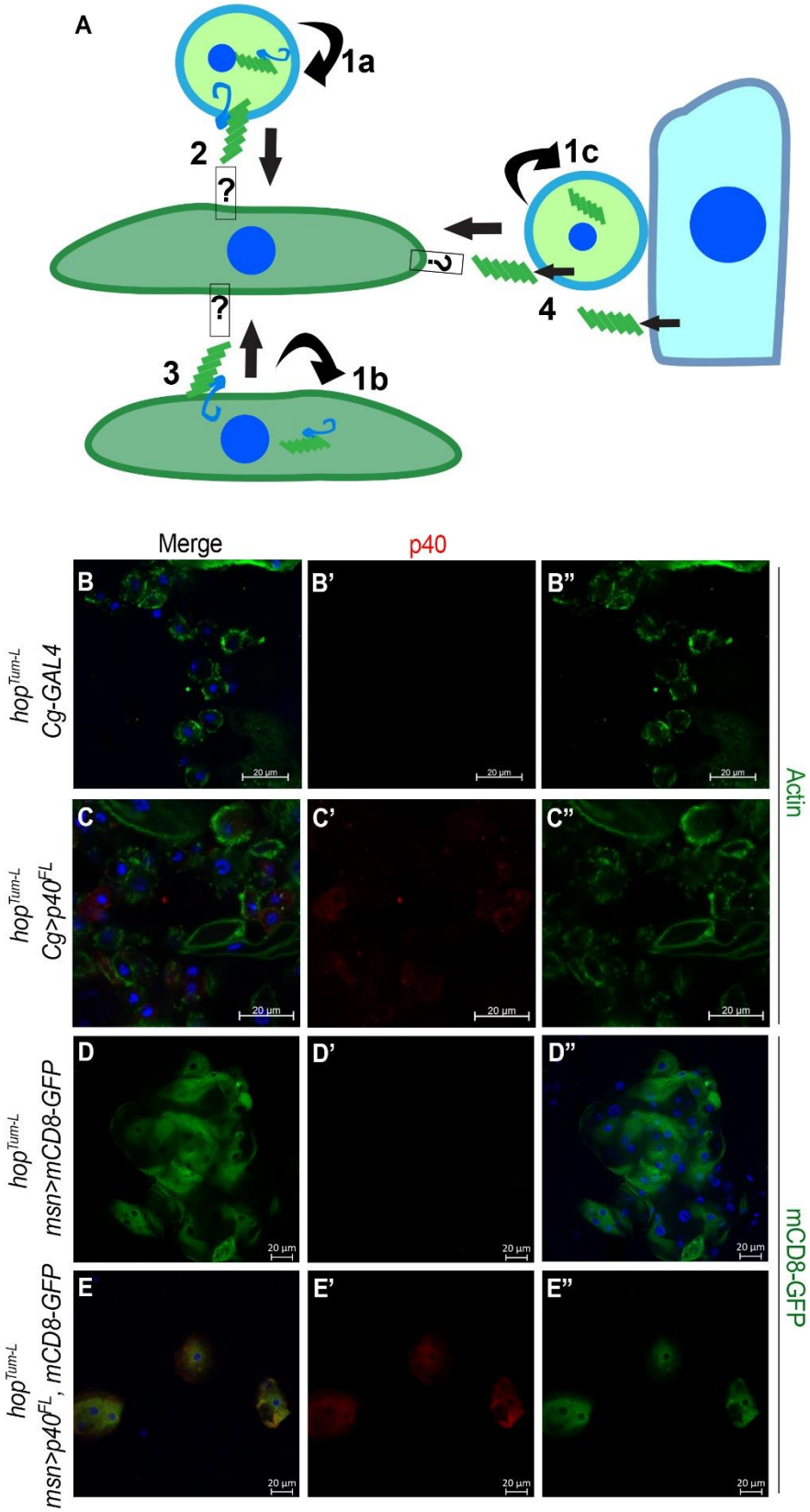

A
1a
2
?
1c
?
4
3
1b
Merge
p40
B
B'
B"
C
C'
C"
D
D'
D"
E
E'
E"
hopTum-L Cg-GAL4
hopTum-L Cg>p40FL
hopTum-L msn>mCD8-GFP
hopTum-L msn>p40FL, mCD8-GFP
Actin
mCD8-GFP
20 µm

Figure 4: Transgenic expression of secreted p40 reduces encapsulation of *Lv* but does not prevent differentiation of lamellocytes. (A) Expression of $p40^{noTM}$ significantly reduced encapsulation of *Lv* embryos/larvae compared to myr-mRFP control, p-value = 0.004. (B) Expression of $p40^{noTM}$ significantly reduced presence of melanized cell aggregates in *Lv* parasitized hosts compared to myr-mRFP control, p-value = 2.57E-5.

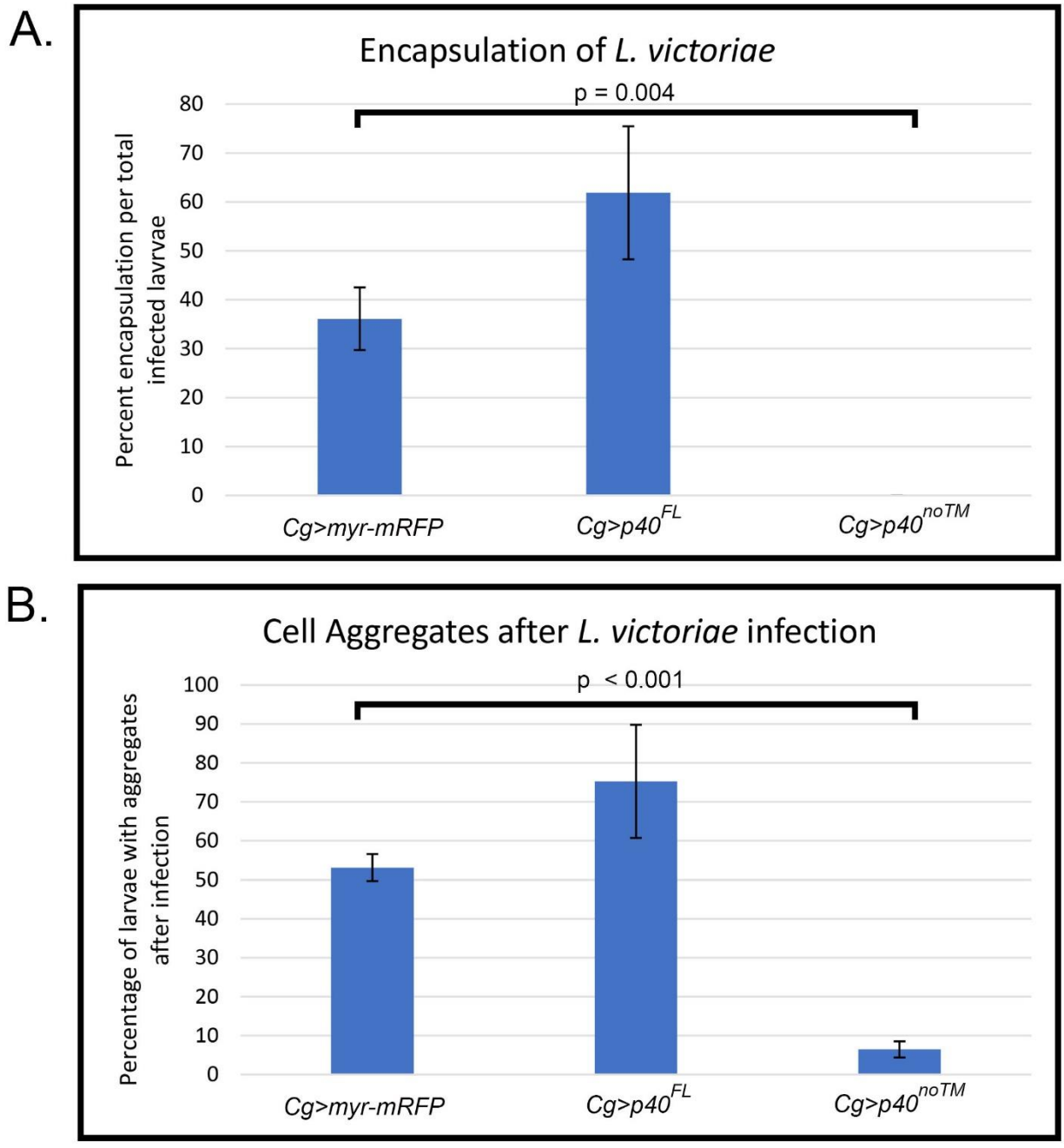


Figure 5: *Lh* venom gland *in situ* with DIG-labeled probes to detect *p40* transcription. Venom glands are from wasps 3 days before eclosion (A-A''), newly eclosed wasps (B-B''), 1-week post eclosion (C-C''), or 2 weeks post eclosion (D-D''). Presence of transcript is detected as purple color from alkaline phosphatase staining reaction. Scale bar indicates 100 µm.

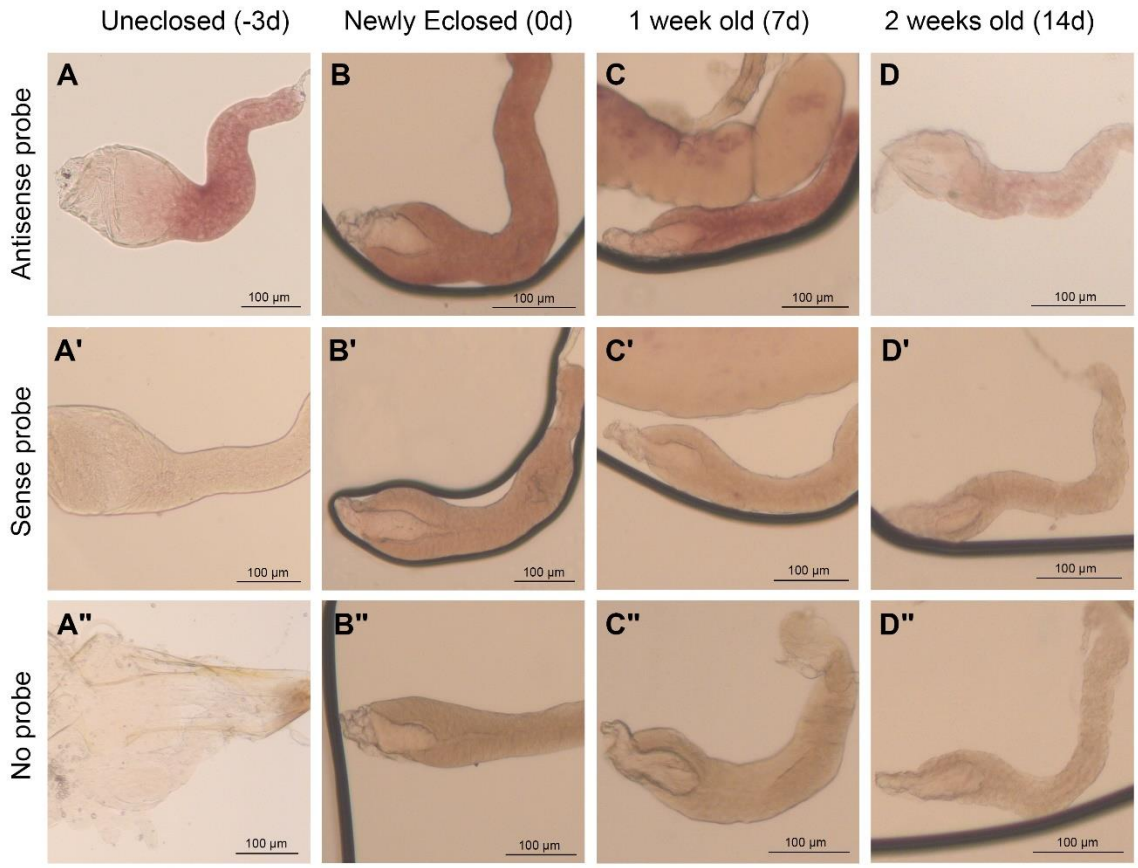


Figure 6: Immunohistochemistry detection of p40 protein in wasp venom glands. Wasps at four developmental time points (A) uneclosed (3 days before eclosion), (B) newly eclosed, (C) 1 week old, (D) 2 weeks old. Images were taken at 40x magnification. Scale bar indicates 50 µm.

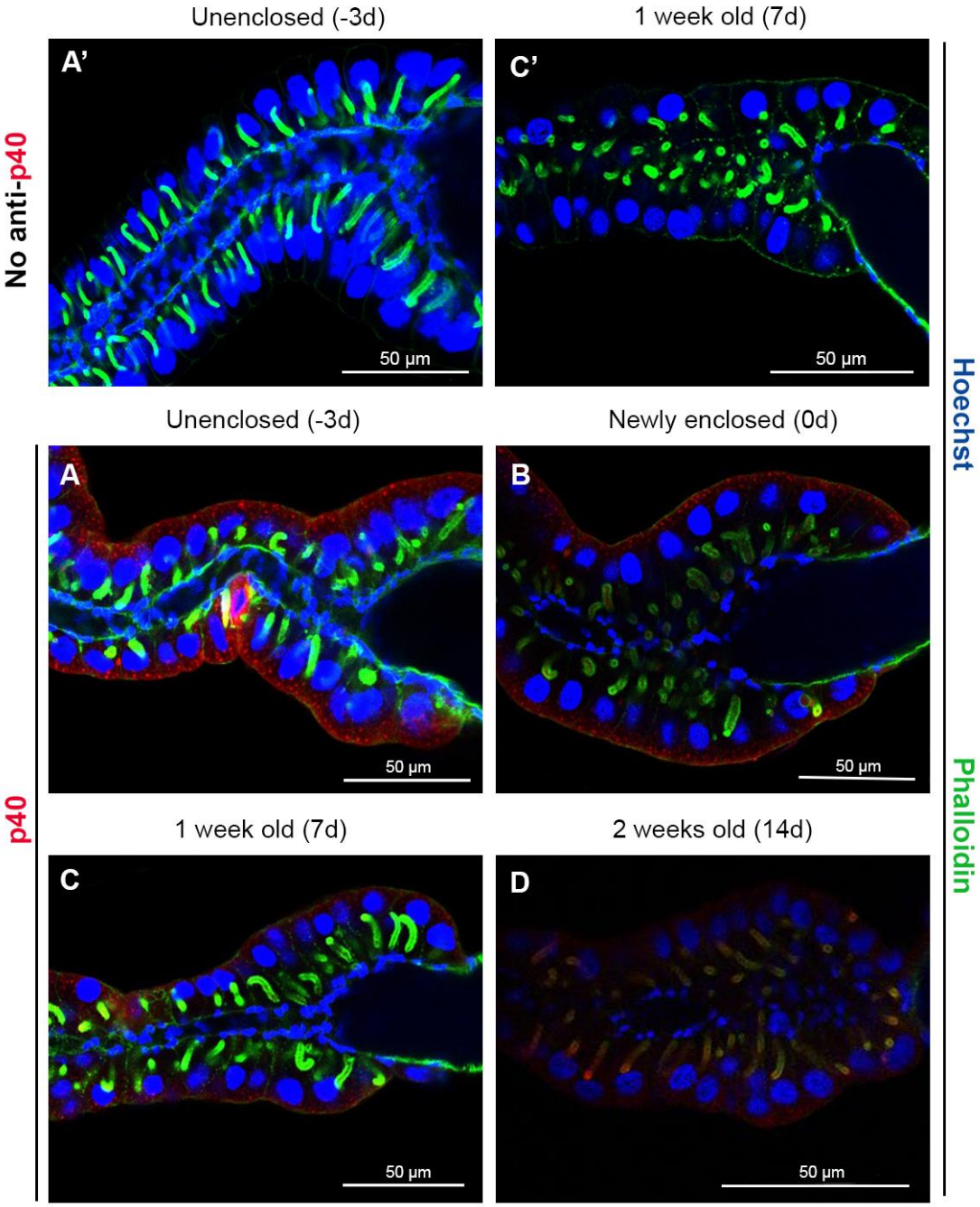
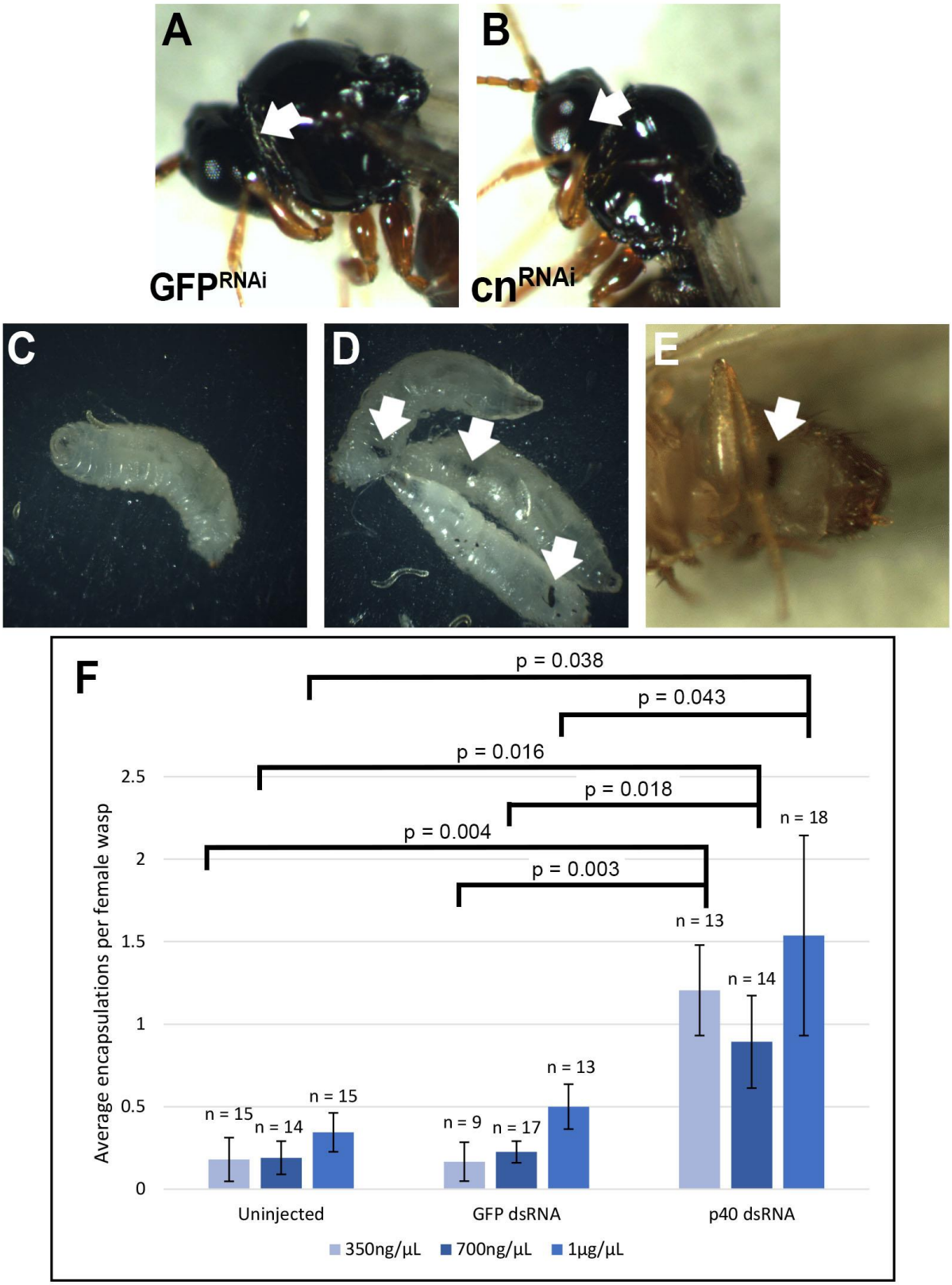


Figure 7: RNA interference in *Lh* indicates importance of p40 in suppressing encapsulation. (A and B) knockdown of eye color gene *cinnabar* as proof of concept. GFP^{RNAi} does not change the normal black eye color (A) but cn^{RNAi} causes a deep red eye color (B). (C-E) Knockdown of p40 causes increased encapsulation. Uninjected wasps were also used as a control. Panel C shows a host without encapsulation after infection by a wasp injected with 700 ng/μL GFP^{RNAi}. Panels D and E show larval hosts with capsules and an adult fly with capsule caused by $p40^{RNAi}$ (700 ng/μL) wasps. (F) Comparison of encapsulation (per female wasp-see Methods) induced by wasps injected with different $p40^{RNAi}$ concentrations relative to uninjected wasps or GFP^{RNAi} wasps. Number of encapsulation assays (biological replicates) for each treatment is indicated as n above each bar. The number of wasps in each biological replicate ranged from 1 to 4. Error bars indicate standard error of mean. p values indicate statistical significance using Welch's one-tailed t-test. p40 dsRNA vs uninjected: 350 ng/μL p = 0.004, 700 ng/μL p = 0.015, 1 μg/μL p = 0.038. p40 dsRNA vs GFP dsRNA: 350 ng/μL p = 0.003, 700 ng/μL p = 0.018, 1 μg/μL p = 0.043. Uninjected control vs GFP dsRNA control: 350 ng/μL p = 0.94, 700 ng/μL p = 0.38, 1 μg/μL p = 0.42).


A
GFP^RNAi
B
cn^RNAi
C
D
E
F
p = 0.038
p = 0.043
p = 0.016
p = 0.018
p = 0.004
p = 0.003
Average encapsulations per female wasp
2.5
2
1.5
1
0.5
0
n = 15
n = 14
n = 15
n = 9
n = 17
n = 13
n = 13
n = 14
n = 18
Uninjected
GFP dsRNA
p40 dsRNA
350ng/µL
700ng/µL
1µg/µL

TABLES

Table 1: Primer sequences used in cloning of *L. heterotoma* p40 and *cinnabar* constructs.

Primer Sequences		
Name	**Sequence**	**Ref**
p40-F	5'-GAATCATTGTTCGTTTGCTTGAAGAAAGAATTGG-3'	Ramroop (*7, 35*)
p40-R	5'-CATTATTAATGGGCCTTTACAATAATTTTAGCC-3'	Ramroop (*7, 35*)
p40CD-F	5'-AAAGCAGAAATAAGAAAACCAACTGCAGATGA-3'	Ramroop (*35*)
p40CD-R	5'-GTAATTGTTTTCTTCCAAGGACTACTAACAATCAC-3'	Ramroop (*35*)
RFPf	5'- GCTGCGCGGCACCAACTTCC-3'	Drosophila Gateway Collection
RFPr	5'- GGACAGCTTCAAGTAGTCGG-3'	Drosophila Gateway Collection
***cn*-F**	5'-CGACCACGTTACGATTGTG-3'	This work
***cn*-F**	5'-GGAACCATTGCATGGGC-3'	This work
GW-p40-5'SP	5'-GGGGCACCATGAAAATGCATACTTCAGTTTTTGGC-3'	This work
GW-p40-3'TM	5'- ATTTGACTTGGTCATTATTGAGGCACCATA-3'	This work
GW-p40-3'noTM	5'-AATTGTTTTCTTCCAAGGACTACTAACAATCAC-3'	This work
GW-p40-5'noSP	5-GGGGCACCATGGAAATAAGAAAACCAACTGCAGATGAAATT-3'	This work

Table 2: Fly lines used in this work. References for original strains are provided where applicable.

Fly Lines		
Name	**Genotype**	**Obtained/Ref**
Collagen (*Cg*)-GAL4 (Expressed in fat body and plasmatocytes)	*w; Cg-GAL4*	Bloomington Stock Center #7011, (*32*)
daughterless (*da*)-GAL4 (Expressed globally)	*w;+;da-GAL4*	Bloomington Stock Center #55851, (*33*)
Hemese (*He*)-Gal4 (Expressed in hemocytes)	*w; He-GAL4*	Bloomington Stock Center #8699, (*55*)
UAS-mCD8-GFP	*y w; UAS-mCD8-GFP*	Bloomington Stock Center #30002, (*37*)
UAS-myr-mRFP	*w; UAS-myr-mRFP*	Bloomington Stock Center #7118 (Henry Chang lab)
15.2 (contains *misshapen (msn)-GAL4* for expression in lamellocytes)	*y v hop^{Tum-l} msn-GAL4; UAS-mCD8-GFP*	Govind lab, (*53*)
15.2/FM7 (contains *misshapen (msn)-GAL4* for expression in lamellocytes)	*y v hop^{Tum-l} msn-GAL4/FM7a; UAS-mCD8-GFP*	Govind lab, (*53*)
hop^{Tum-l}, msn-GFP	*y v hop^{Tum-l}; msnf9-83*	This work
hop^{Tum-l}/FM7, msn-GFP	*y v hop^{Tum-l}/FM7; msnf9-83*	This work
UAS-p40^{FL}	*y w; UAS-p40^{FL}*	This work
UAS-p40^{CD}	*y w; UAS-p40^{CD}*	This work
UAS-p40^{noTM}	*w; UAS-p40^{noTM}*	This work
UAS-p40^{noSP}	*w; UAS-p40^{noSP}/CyO-GFP*	This work
UAS-p40^{FL}, UAS-mCD8-GFP	*y w UAS-p40^{FL}; UAS-mCD8-GFP*	This work
Cg>p40^{FL}	*w; Cg-GAL4, UAS-p40^{FL}*	Recombinant line made for this work
Cg>p40^{CD}	*w; Cg-GAL4, UAS-p40^{CD}*	Recombinant line made for this work
Cg>p40^{noTM}	*w; Cg-GAL4, UAS- p40^{noTM}*	Recombinant line made for this work
Cg>p40^{noSP}	*w; Cg-GAL4, UAS-p40^{noSP}/CyO-GFP*	Recombinant line made for this work
Cg>myr-mRFP	*w; Cg-GAL4, UAS-myr-mRFP*	Recombinant line made for this work

Table 3: Survival statistics for double strand RNA injection. Wasps were considered to have survived if they were fully developed and motile 12 days post injection. Cultured indicates total number of wasps dissected for a given treatment (uninjected, GFP dsRNA injected (GFP), or p40 dsRNA injected (p40). The term “survived” refers to the total number of male and female wasps that survived treatment. The number of surviving females is also noted.

dsRNA Injection Statistics				
Dosage	Parameter	Uninjected	GFP	p40
350ng/uL	Total Attempt	186	212	188
	Total Survived	102	53	63
	% Survived	54.83871	25	33.51064
	Total Surviving Females	48	27	27
700ng/uL	Total Attempt	130	176	182
	Total Survived	88	102	94
	% Survived	67.692308	57.95455	51.64835
	Total Surviving Females	42	53	46
1ug/uL	Total Attempt	160	203	210
	Total Survived	114	84	116
	% Survived	71.25	41.37931	55.2381
	Total Surviving Females	46	36	53

REFERENCES

1. R. M. Rizki, T. M. Rizki, Selective destruction of a host blood cell type by a parasitoid wasp. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 6154-6158 (1984).
2. R. M. Rizki, T. M. Rizki, Effects of lamellolysin from a parasitoid wasp on Drosophila blood cells in vitro. *J Exp Zool* **257**, 236-244 (1991).
3. T. M. Rizki, R. M. Rizki, Y. Carton, Leptopilina heterotoma and L. boulardi: strategies to avoid cellular defense responses of Drosophila melanogaster. *Exp Parasitol* **70**, 466-475 (1990).
4. J. Morales *et al.*, Biogenesis, structure, and immune-suppressive effects of virus-like particles of a Drosophila parasitoid, Leptopilina victoriae. *J Insect Physiol* **51**, 181-195 (2005).
5. H. Chiu, S. Govind, Natural infection of D. melanogaster by virulent parasitic wasps induces apoptotic depletion of hematopoietic precursors. *Cell death and differentiation* **9**, 1379-1381 (2002).
6. R. M. Rizki, T. M. Rizki, Parasitoid virus-like particles destroy Drosophila cellular immunity. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 8388-8392 (1990).
7. M. E. Heavner *et al.*, Novel Organelles with Elements of Bacterial and Eukaryotic Secretion Systems Weaponize Parasites of Drosophila. *Curr Biol* **27**, 2869-2877 e2866 (2017).
8. B. Wey *et al.*, Immune Suppressive Extracellular Vesicle Proteins of Leptopilina heterotoma Are Encoded in the Wasp Genome. *G3 (Bethesda)* **10**, 1-12 (2020).
9. H. Chiu, J. Morales, S. Govind, Identification and immuno-electron microscopy localization of p40, a protein component of immunosuppressive virus-like particles from Leptopilina heterotoma, a virulent parasitoid wasp of Drosophila. *J Gen Virol* **87**, 461-470 (2006).
10. R. Ferrarese, J. Morales, D. Fimiarz, B. A. Webb, S. Govind, A supracellular system of actin-lined canals controls biogenesis and release of virulence factors in parasitoid venom glands. *J Exp Biol* **212**, 2261-2268 (2009).
11. J. Goecks *et al.*, Integrative approach reveals composition of endoparasitoid wasp venoms. *PloS one* **8**, e64125 (2013).
12. O. Arizmendi, W. D. Picking, W. L. Picking, Macrophage Apoptosis Triggered by IpaD from Shigella flexneri. *Infect Immun* **84**, 1857-1865 (2016).
13. K. Nothelfer *et al.*, B lymphocytes undergo TLR2-dependent apoptosis upon Shigella infection. *J Exp Med* **211**, 1215-1229 (2014).
14. W. L. Picking *et al.*, IpaD of Shigella flexneri is independently required for regulation of Ipa protein secretion and efficient insertion of IpaB and IpaC into host membranes. *Infect Immun* **73**, 1432-1440 (2005).
15. A. D. Roehrich, E. Guillossou, A. J. Blocker, I. Martinez-Argudo, Shigella IpaD has a dual role: signal transduction from the type III secretion system needle tip and intracellular secretion regulation. *Molecular microbiology* **87**, 690-706 (2013).
16. C. Ongvarrasopone, Y. Roshorm, S. Panyim, A simple and cost effective method to generate dsRNA for RNAi studies in invertebrates. *ScienceAsia* **33**, 35-39 (2007).
17. Y. Tomoyasu *et al.*, Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in Tribolium. *Genome Biol* **9**, R10 (2008).
18. D. R. G. Price, J. A. Gatehouse, RNAi-mediated crop protection against insects. *Trends in Biotechnology* **26**, 393-400 (2008).
19. J. H. Werren, D. W. Loehlin, J. D. Giebel, Larval RNAi in Nasonia (parasitoid wasp). *Cold Spring Harb Protoc* **2009**, pdb.prot5311 (2009).
20. W. Hunter *et al.*, Large-Scale Field Application of RNAi Technology Reducing Israeli Acute Paralysis Virus Disease in Honey Bees (Apis mellifera, Hymenoptera: Apidae). *PLOS Pathogens* **6**, e1001160 (2010).
21. R. Asokan, G. S. Chandra, M. Manamohan, N. K. Kumar, Effect of diet delivered various concentrations of double-stranded RNA in silencing a midgut and a non-midgut gene of Helicoverpa armigera. *Bull Entomol Res* **103**, 555-563 (2013).
22. R. Asokan, G. S. Chandra, M. Manamohan, N. K. K. Kumar, T. Sita, Response of various target genes to diet-delivered dsRNA mediated RNA interference in the cotton bollworm, Helicoverpa armigera. *J Pest Sci* **87**, 163-172 (2014).

23. D. Colinet *et al.*, Development of RNAi in a Drosophila endoparasitoid wasp and demonstration of its efficiency in impairing venom protein production. *J Insect Physiol* **63**, 56-61 (2014).
24. K. B. Rebijith *et al.*, Diet-Delivered dsRNAs for Juvenile Hormone-Binding Protein and Vacuolar ATPase-H Implied Their Potential in the Management of the Melon Aphid (Hemiptera: Aphididae). *Environ Entomol*, (2015).
25. A. L. Siebert, D. Wheeler, J. H. Werren, A new approach for investigating venom function applied to venom calreticulin in a parasitoid wasp. *Toxicon*, (2015).
26. T. A. Schlenke, J. Morales, S. Govind, A. G. Clark, Contrasting infection strategies in generalist and specialist wasp parasitoids of Drosophila melanogaster. *PLoS Pathog* **3**, 1486-1501 (2007).
27. C. Small, I. Paddibhatla, R. Rajwani, S. Govind, An introduction to parasitic wasps of Drosophila and the antiparasite immune response. *J Vis Exp*, e3347 (2012).
28. A. H. Brand, N. Perrimon, Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401-415 (1993).
29. A. H. Brand, A. S. Manoukian, N. Perrimon, Ectopic expression in Drosophila. *Methods Cell Biol* **44**, 635-654 (1994).
30. R. E. Campbell *et al.*, A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences* **99**, 7877-7882 (2002).
31. C. Q. Huynh, H. Zieler, Construction of modular and versatile plasmid vectors for the high-level expression of single or multiple genes in insects and insect cell lines 11Edited by M. Yaniv. *Journal of molecular biology* **288**, 13-20 (1999).
32. K. J. Argue, W. S. Neckameyer, Altering the sex determination pathway in Drosophila fat body modifies sex-specific stress responses. *Am J Physiol Regul Integr Comp Physiol* **307**, R82-92 (2014).
33. A. Wodarz, U. Hinz, M. Engelbert, E. Knust, Expression of crumbs confers apical character on plasma membrane domains of ectodermal epithelia of Drosophila. *Cell* **82**, 67-76 (1995).
34. M. Crozatier, D. Valle, L. Dubois, S. Ibnsouda, A. Vincent, collier, a novel regulator of Drosophila head development, is expressed in a single mitotic domain. *Current Biology* **6**, 707-718 (1996).
35. J. Ramroop, Dissertation, Graduate Center, City University of New York, (2016).
36. R. A. J. McIlhinney, in *Protein Targeting Protocols,* R. A. Clegg, Ed. (Humana Press, Totowa, NJ, 1998), pp. 211-225.
37. T. Lee, L. Luo, Mosaic Analysis with a Repressible Cell Marker for Studies of Gene Function in Neuronal Morphogenesis. *Neuron* **22**, 451-461 (1999).
38. C. W. Liaw, R. Zamoyska, J. R. Parnes, Structure, sequence, and polymorphism of the Lyt-2 T cell differentiation antigen gene. *The Journal of Immunology* **137**, 1037-1043 (1986).
39. A. S. Sechi, J. Wehland, The actin cytoskeleton and plasma membrane connection: PtdIns(4,5)P(2) influences cytoskeletal protein activity at the plasma membrane. *Journal of Cell Science* **113**, 3685-3695 (2000).
40. J. Saarikangas, H. Zhao, P. Lappalainen, Regulation of the Actin Cytoskeleton-Plasma Membrane Interplay by Phosphoinositides. *Physiological Reviews* **90**, 259-289 (2010).
41. D. V. Köster, S. Mayor, Cortical actin and the plasma membrane: inextricably intertwined. *Current Opinion in Cell Biology* **38**, 81-89 (2016).
42. T. C. Pesacreta, T. J. Byers, R. Dubreuil, D. P. Kiehart, D. Branton, Drosophila spectrin: the membrane skeleton during embryogenesis. *Journal of Cell Biology* **108**, 1697-1709 (1989).
43. J. K. Lee, R. S. Coyne, R. R. Dubreuil, L. S. Goldstein, D. Branton, Cell shape and interaction defects in alpha-spectrin mutants of Drosophila melanogaster. *Journal of Cell Biology* **123**, 1797-1809 (1993).
44. H. Asha *et al.*, Analysis of Ras-induced overproliferation in Drosophila hemocytes. *Genetics* **163**, 203-215 (2003).
45. R. P. Mills, R. C. King, The pericardial cells of Drosophila melanogaster. *Q J Microsc Sci* **106**, 261-268 (1965).
46. D. Das, R. Aradhya, D. Ashoka, M. Inamdar, Macromolecular uptake in Drosophila pericardial cells requires rudhira function. *Experimental Cell Research* **314**, 1804-1810 (2008).
47. F. Zhang, Y. Zhao, Z. Han, An in vivo functional analysis system for renal gene discovery in Drosophila pericardial nephrocytes. *J Am Soc Nephrol* **24**, 191-197 (2013).

48. M. E. Heavner *et al.*, Partial venom gland transcriptome of a Drosophila parasitoid wasp, Leptopilina heterotoma, reveals novel and shared bioactive profiles with stinging Hymenoptera. *Gene* **526**, 195-204 (2013).
49. A. K. Shia *et al.*, Toll-dependent antimicrobial responses in Drosophila larval fat body require Spatzle secreted by haemocytes. *J Cell Sci* **122**, 4505-4515 (2009).
50. B. Rotstein, A. Paululat, On the Morphology of the Drosophila Heart. *J Cardiovasc Dev Dis* **3**, (2016).
51. R. P. Sorrentino, Y. Carton, S. Govind, Cellular immune response to parasite infection in the Drosophila lymph gland is developmentally regulated. *Developmental biology* **243**, 65-80 (2002).
52. I. Paddibhatla, M. J. Lee, M. E. Kalamarz, R. Ferrarese, S. Govind, Role for sumoylation in systemic inflammation and immune homeostasis in Drosophila larvae. *PLoS Pathog* **6**, e1001234 (2010).
53. S. Panettieri *et al.*, Discovery of aspirin-triggered eicosanoid-like mediators in a Drosophila metainflammation blood tumor model. *J Cell Sci* **133**, (2019).
54. B. A. Pannebakker, N. R. T. Garrido, B. J. Zwaan, J. J. M. Van Alphen, Geographic variation in host-selection behaviour in the Drosophila parasitoid Leptopilina clavipes. *Entomologia Experimentalis et Applicata* **127**, 48-54 (2008).
55. C.-J. Zettervall *et al.*, A directed screen for genes involved in *Drosophila* blood cell activation. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 14192-14197 (2004).