

Title:

Bioinformatic analysis reveals mechanisms underlying parasitoid venom evolution and function

Author names and affiliations:

Gloria Alvarado, Sarah R. Holland, Jordan DePerez-Rasmussen, Brice A. Jarvis, Tyler Telander, Nicole Wagner, Ashley L. Waring, Anissa Anast, Bria Davis, Adam Frank, Katelyn Genenbacher, Josh Larson, Corey Mathis, A. Elizabeth Oates, Nicholas A. Rhoades, Liz Scott, Jamie Young, and Nathan T. Mortimer.

School of Biological Sciences, Illinois State University, Normal, IL, USA

Corresponding author:

Nathan T. Mortimer
Email: ntmorti@ilstu.edu
Phone: (309) 438-8597

Introduction

Parasitoid wasps are a numerous and diverse group of insects that obligately infect other arthropod species. These wasps lay their eggs either on the surface or within the body cavity of their hosts, and the resulting offspring exploit the hosts' resources to complete their development [1]. Many parasitoids also introduce venom gland derived proteins or polydnaviruses into the host during infection. These factors act through a variety of mechanisms to manipulate host biology in order to increase the fitness of the developing parasitoid offspring [2–4]. Many hosts mount immune responses to parasitoid infection, and accordingly, parasitoids have evolved multiple venom genes that encode immunomodulatory factors. These factors target host cellular and humoral immune mechanisms, to allow the developing parasitoid to evade the killing activity of host immunity [5–11]. These parasitoid immunomodulatory activities are strikingly diverse, and include venoms that block host immune cell signaling [5,6], alter host gene expression [11,12], or even cause immune cell death [13]. Wasp venoms also contain proteins that manipulate host physiology and metabolism [14–16], including factors that alter host development [17,18], lipid metabolism [19], and nutrient flux [20–22]. These manipulations are thought to improve the nutritional content of the hosts' internal environment in order to benefit the developing wasp [2]. Along with these physiological targets, many parasitoids can also use venom factors to alter their hosts' behavior [23–28]. These modified behaviors promote parasitoid survival and development; often at the expense of the hosts' own fitness.

Most molecular studies aimed at uncovering the functions of parasitoid venom proteins have been focused on the characterization of a single protein. However, with the advent of high throughput "venomics", venoms can now be considered as a whole in order to gain insight into higher order venom properties. These venomic studies have used an approach that combines genomic or transcriptomic sequencing with the purification and mass spectrometry based identification of venom proteins. This approach allows for the entire venom protein content to be

identified in a single experiment, and has been completed on diverse parasitoid wasp species [5,29–38]. As the number of venomomics projects has increased, understanding the composition of venoms has become an active area of research. This includes studies focused on exploring both the evolution of venom content and venom protein function [15,39].

The evolution of venom content is an especially interesting question in evolutionary biology. It has been hypothesized that proteins may be recruited into the venom proteome through a variety of processes, including: gene multifunctionality, in which genes are active in parasitoids in both cellular and venom contexts [39]; the co-option of cellular genes into venom specific expression and function [40]; the evolution of alternative splicing products of cellular genes that result in venom specific protein isoforms [5,41]; gene duplication events that create venom specific paralogs [29–31]; the evolution of *de novo* genes [39]; and the horizontal transfer of genes between organisms [42]. The ability to study a whole venom proteome will allow us to address venom evolution from a unique perspective, and to test whether these different mechanisms may collectively function to drive the evolution of venom composition within a single parasitoid species. Similarly, while the “gene-by-gene” approach has revealed the molecular details of venom protein function, the consideration of predicted functions of the venom as a whole can provide additional insight into understanding venom function.

Due to the wide range of encoded activities, understanding the composition and function of parasitoid venoms is also a subject of great interest in applied biology. Parasitoids are important biological control agents for integrated pest management programs [18,43], and as a result of their ability to disrupt the development of insect pests, parasitoid venom factors may be useful as the basis for the synthesis of novel bio-insecticides [2,15,18]. Furthermore, many parasitoid venoms specifically target conserved signaling pathways in their hosts, including the Toll and JAK-STAT signaling pathways, intracellular calcium signaling, and Rho family GTPases [5,6,12]. Since homologues of these pathways play conserved roles in human health, these

venom proteins may have the potential to be developed as pharmaceutical agents [2,15,44].

We are particularly interested in characterizing the venoms of the *Drosophila*-infecting parasitoid wasps. These wasps are endoparasites that infect *Drosophila* species during the larval stage, and complete their own development during the fly larval and pupal stages [1]. Since they develop within the larval body cavity, these parasitoid species encounter the fly's robust cellular immune response, and accordingly have evolved virulence proteins that allow them to evade fly immunity [5–7,45–48]. We have previously identified 166 venom protein encoding genes from the *Drosophila*-infecting parasitoid *Ganaspis sp.1* [5]. Here we take a whole venom bioinformatic approach to analyze the evolution as well as putative functions of *Ganaspis sp.1* venom.

Methods

Sequence Data

The *Ganaspis sp.1* sequence data has been deposited under Transcriptome Shotgun Assembly (TSA) project accession GAIW00000000. Other sequence accession numbers were EU482092.1 (*Blastocystis* 18S rRNA), AB038366.1 (*Arsenophonus* 16S rRNA), AB915783.1 (*Sodalis* 16S rRNA), and as given in the text.

Transcript Abundance

The relative abundances of venom and cellular mRNA transcripts in our RNAseq data were compared using the RSEM software [49]. RNAseq reads were aligned to our *de novo* transcriptome assembly to generate transcript counts in transcripts per million (TPM). TPM values were then ranked and the venom and cellular transcript groups were compared by Wilcoxon signed-rank test (implemented in R).

Protein Properties

We first generated a random list of cellular proteins using a custom Perl script. The

molecular weight (MW) and isoelectric point (pI) of each protein was then calculated using the ExPASy Compute pI/Mw tool [50]. Values were then compared by Welch Two Sample t-test (implemented in R).

Homology Finding

To find homologous sequences, each *Ganaspis sp. 1* venom protein was aligned to the non-redundant database using blastp (accessed February 2016) [51]. Conserved domains were identified using the Batch CD-Search tool to query the Conserved Domain Database (CDD), and the numbers of occurrences of each CDD accession number were counted [52,53]. Domain enrichment was determined using Fisher's exact test, and domain numbers were compared by Welch Two Sample t-test (implemented in R).

Phylogeny

To build phylogenies for investigating potential horizontal gene transfers, homologous sequences were identified by homology finding and then processed using MEGA [54]. These sequences were first aligned and the alignments were used to construct Maximum Likelihood phylogenetic trees, with 500 Bootstrap replications and JTT substitution model.

Results and Discussion

Characterization of *Ganaspis sp. 1* venom encoding genes

We predict that *Ganaspis sp. 1* venom protein encoding genes are distinct from cellular (non-venom) protein encoding genes. We expect that when compared to a subset of cellular genes, venom genes would have distinct properties that include expression level and the properties of the encoded venom proteins. We hypothesize that venom encoding transcripts will be among the most abundant, due to the crucial role of the venom in facilitating successful parasitization [40]. To test this hypothesis, we compared the abundance of venom protein encoding transcripts with wasp cellular gene transcripts. We found that venom gene transcripts

have significantly higher expression levels, with a mean abundance of 145.15 transcripts per million (TPM), compared to 35.89 TPM for cellular transcripts (Supplementary Table 1; Wilcoxon rank sum test: $W = 4020800$, $p < 2.2e-16$).

To test whether the venom proteins from *Ganaspis sp.1* also show distinct physical properties, we compared the predicted molecular weight (MW) and isoelectric point (pI) of *Ganaspis sp.1* venom proteins with a randomly derived list composed of an equal number of *Ganaspis sp.1* cellular proteins. We found that, on average, *Ganaspis sp.1* venom proteins are significantly smaller (MW of 44.3 kDa for venom proteins compared to 62.1 kDa for non-venom proteins, $p = 0.004397$, Figure 1A), and have a significantly lower pI (6.99 for venom proteins compared to 8.06 for non-venom proteins, $p = 1.112e-06$, Figure 1B) than *Ganaspis sp.1* cellular proteins. Interestingly, *Ganaspis sp.1* venom lacks the abundant class of small, highly charged peptides that characterize the venoms of many social hymenopteran species [55–59]. These data suggest that the composition of parasitoid wasp venom may be distinct from both cellular proteins and also from the venoms of non-parasitoid hymenopterans.

Origin of *Ganaspis sp.1* venom encoding genes

Our data can provide insight into the origins of the *Ganaspis sp.1* venom proteome. To investigate the origins of *Ganaspis sp.1* venom proteins, we first used BLAST to identify the closest homolog of each venom. We determined that for 117 venom proteins the closest homolog is found within another hymenopteran species, and that other insect species account for the nearest homolog of a further 11 venom proteins. These findings suggest that the majority of *Ganaspis sp.1* venom proteins are either multifunctional or have evolved from cellular genes (Figure 1C). Interestingly, we also detected 37 venom proteins that do not have homology to any sequence found in the NCBI database (Figure 1C), suggesting that these venom proteins (Supplementary Table 2) may have evolved as *de novo* protein encoding genes. This finding is in accordance with the venoms of other parasitoid species, notably *Nasonia vitripennis*, in which

23 of the 79 identified venom proteins have no sequence similarity to known proteins [2]. The proteins encoded by these putative *Ganaspis sp.1 de novo* venom genes are significantly smaller than the conserved venom proteins (24.8 kDa vs 49.7 kDa, $p = 1.345e-05$, Figure 1D), but we do not find a difference in pI (6.97 vs 7.00, $p = 0.9447$, Figure 1E). This smaller protein size is consistent with *de novo* proteins in a wide range of other species, including other insects [60,61], humans [62,63], and plants [64,65].

Our BLAST data further identified 3 venom proteins that only show significant homology to non-insect sequences, suggesting that the genes encoding these proteins may have been introduced into the *Ganaspis sp.1* genome, and subsequently into the venom proteome, by horizontal gene transfer (HGT; [39]). HGT is characterized by the transfer of DNA between species, and HGT events can be seen as phylogenetic incongruities [66,67]. The first of these non-insect derived genes is *comp112* which shows homology to the AV274_2271 protein from *Blastocystis sp. NandII* (Figure 2A). *Blastocystis spp.* are single celled intracellular protozoan parasites that infect a wide range of species from humans to insects [68,69]. The other two non-insect derived venom genes, *comp3581* and *comp11645*, are most similar to genes from insect symbiotic enterobacteria species [70,71]. *comp3581* is homologous to the WP 032114503.1 gene from *Arsenophonus sp.* (an endosymbiont of *Nilaparvata lugens*; Figure 2B) and the *comp11645* gene is homologous to the *SopA-like* gene from *Sodalis sp.* (Figure 2C).

It has recently been noted that many proposed examples of HGT events may instead represent sequencing artefacts [72,73]. However, due to our workflow, we have isolated these 3 putative horizontally transferred genes from distinct biological samples; the transcripts were sequenced from our mRNA library, and the proteins were identified from purified venom by mass spectrometry [5]. This provides important support for the hypothesis of HGT in *Ganaspis sp.1* venom. Additionally, we did not find significant homology to either the 18S (*Blastocystis*) or 16S (*Arsenophonus* and *Sodalis*) rRNA sequences of these species, suggesting that they are

not present in *Ganaspis sp.1* wasps.

HGT into eukaryotes is a comparatively rare event, however, there are examples of horizontally transferred genes being expressed in eukaryotic transcriptomes [74], and eukaryotic HGT is most commonly observed between intracellular microbes and their hosts [75–80]. The intracellular localization of *Blastocystis spp.*, *Arsenophonus spp.* and *Sodalis spp.* therefore suggests a possible mechanism for the transfer of these genes into the *Ganaspis sp.1* genome. While the functions of these genes are unknown, their presence in *Ganaspis sp.1* venom suggests they may play a role in successful parasitization. There are multiple examples of horizontally transferred genes in arthropod genomes that are expressed and play important functional roles in metabolism and immunity [81–84]. Of particular relevance is the GH19 family of glycoside hydrolase chitinases, which is widespread in bacteria and plants but only found in one metazoan lineage, the chalcid parasitoid wasps [42]. GH19 chitinase is expressed in the venom glands of multiple chalcid parasitoid species, including the model parasitoid wasp *N. vitripennis*, where it acts to manipulate the transcription of host immune genes following infection [42].

Paralog evolution

Along with the recruitment of *de novo* or horizontally transferred genes, venom proteins can also arise as a result of duplications of cellular genes [29–31]. The duplication of either entire or partial genes can result in the production of novel genes via paralog formation or exon shuffling [85]. We predict that these processes are likely to contribute to the evolution of the composition of *Ganaspis sp.1* venom. A venom gene formed from a genome duplication would show a high degree of similarity to an extant cellular gene. Additionally, we would predict that if the genome duplication event was specific to *Ganaspis sp.1* and/or closely related parasitoids, we would expect that the resulting genes would have only a single homolog in more distantly related species. In order to identify venom encoding genes that may have arisen as a result of

recent genome duplications, we searched for venom proteins with significant homology to a cellular protein, and with only a single homolog in the *N. vitripennis* and *D. melanogaster* genomes.

We found 10 venom genes that meet these criteria and identified their paralogs among the cellular proteins (Table 1; Supplementary Table 3). Interestingly, the *D. melanogaster* homologs of 9 of these 10 venom paralogs are important for organismal viability or fertility (Table 1). Several of these *D. melanogaster* genes are involved in energy production processes including glycolysis (*Ald*), the TCA cycle (*Nc73EF*), and the electron transport chain (*Cyt-c-d* and both the alpha and beta subunits of ATP synthase, *blw* and *ATPsynB*). The remaining genes include the peptidyl-prolyl isomerase *Cyp1*, the annexin family member *AnxB9*, the aldo-keto reductase *CG10638*, the M13 family metallopeptidase *Nep2*, and the myosin light chain *Mlc2*.

The vital roles of these genes in organismal development and fertility suggest that their functions are likely to be highly maintained by selection. If these important roles are conserved between *D. melanogaster* and *Ganaspis sp.1*, we would predict that the cellular paralogs would also be subject to purifying selection. Conversely, the venom paralog would be free to diversify through the process of neofunctionalization [85], perhaps acquiring dominant negative properties that would allow *Ganaspis sp.1* venom to manipulate these important processes in *D. melanogaster* cells. These venom proteins could thereby inhibit host immunity or alter host metabolism to confer a benefit to the developing parasitoid offspring.

In support of this hypothesis, we find that *comp1724*, the venom paralog of oxoglutarate dehydrogenase (OGDH), is lacking the dimerization interface normally found within the Transketolase PYR domain (pfam02779; Figure 3A). This dimerization interface is present in the sequences of *comp9906*, the *Ganaspis sp.1* cellular OGDH paralog, and OGDH family members in other species. It allows OGDH to interact with the other subunits of the oxoglutarate

dehydrogenase complex (OGDC), and the formation of this complex is required for enzymatic function [86,87]. This suggests that venom OGDH would not be capable of forming an enzymatically active complex. However, venom OGDH does encode the necessary sequences for binding to thiamine pyrophosphate (TPP; Figure 3B), an obligate cofactor for OGDC function [88,89], and also an essential cofactor for the pyruvate dehydrogenase complex (PDC) [90,91]. Based on these sequence data, we would predict that in host cells, venom OGDH would have a dominant negative effect by sequestering available TPP into nonfunctional interactions. This would likely result in altered host metabolism through inhibiting the function of the OGDC and PDC complexes.

Conserved proteins and venom specific isoforms

In addition to the presence of venom specific proteins, it has previously been demonstrated that multiple venom protein encoding genes are also expressed outside the venom gland [29,40]. This suggests that they encode multifunctionalized proteins that likely act both within the venom and in non-venom cellular processes. We have found that one venom encoding gene, *Ganaspis sp.1* SERCA, encodes two protein isoforms (SERCA₁₀₀₂ and SERCA₁₀₂₀). We demonstrated that SERCA₁₀₂₀ is found throughout the wasp while SERCA₁₀₀₂ is found only in the venom [5], suggesting the presence of a venom-specific SERCA isoform. A similar phenomenon is seen in the parasitoid wasp *Pteromalus puparum*, the venom of which contains a venom-specific isoform of the SERPIN domain protein *PpS1V* [41]. In our current analysis we identified 31 venom proteins that are encoded by genes with multiple isoforms in the *Ganaspis sp.1* transcriptome (Supplementary Table 4). We would hypothesize that the identified venom transcripts may correspond to venom-specific isoforms (as in the case of SERCA and *PpS1V*), however, this hypothesis will require additional testing.

Functional Annotation of *Ganaspis sp.1* Venom Proteins

Parasitoid venom proteins are capable of altering host physiology, metabolism and

immunity, and while there have been numerous studies into the roles of individual venom proteins [5–7,92,93], we wanted to explore the predicted activity of the *Ganaspis sp.1* venom proteome taken as a whole. Interestingly, we find that *Ganaspis sp.1* venom genes encode a wide range of putative functions, with 863 NCBI-curated conserved domain families and superfamilies and an additional 241 PFAM domains identified from the 166 venom proteins. Notably, we find that 14 of these conserved domains are significantly enriched in the venom proteome when compared with our random subset of cellular proteins (Table 2). This suggests that the *Ganaspis sp.1* venom proteome may be composed of a distinct functional set of proteins.

The list of enriched functional domains provides insight into the immunomodulatory activity of *Ganaspis sp.1* venom. We have previously shown that *Ganaspis sp.1* venom blocks the activation of host immune cells by antagonizing intracellular calcium signaling via activity of the venom specific SERCA₁₀₀₂ isoform [5]. SERCA is a calcium ATPase that transports calcium ions out of the cytosol [94], and the addition of *Ganaspis sp.1* venom significantly decreased calcium ion levels in *D. melanogaster* immune cells [5]. Our venom analysis suggests that *Ganaspis sp.1* venom may contain additional proteins with calcium regulatory potential, as revealed by the enrichment of the CALCOCO1, Calponin Homology, and Annexin calcium interacting domains [95–97]. The roles of these proteins in *Ganaspis sp.1* are unknown, but raise the possibility that multiple mechanisms act to antagonize host calcium signaling.

Additionally, insects make extensive use of serine protease cascades in triggering immune responses [98–100], and in turn, these responses are negatively regulated by serine protease inhibitor (SERPIN) domain containing proteins [101,102]. Many previously investigated parasitoid wasps have been found to contain SERPIN domain proteins in their venoms [7,29,31,41,103] suggesting that these parasitoids may use SERPIN domain proteins to negatively regulate host immunity. Interestingly, we also found multiple SERPIN domain

containing proteins in *Ganaspis sp.1* venom, suggesting that this might be a shared feature of parasitoid venoms.

Along with these known venom functions, our proteome-wide analysis has revealed potential novel venom functions. The most abundant of the enriched domains are predicted to mediate protein-ligand interactions. These domains include two distinct forms of Leucine Rich Repeats (LRR; LRR8 repeats [pfam13855] and LRR_RI repeats [cd00116]) and Filamin Repeats (FR; pfam00630). LRRs are 20-29 amino acid long motifs that form discrete helical structures and allow for binding to specific partners in both homotypic and heterotypic interactions [104,105]. Similarly, FR motifs mediate a wide range of protein-protein interactions through the formation of an immunoglobulin-like beta sandwich structure [106–108]. This suggests that the *Ganaspis sp.1* venom proteins containing these domains may be able to interact with specific host proteins following infection.

Evolutionary minimization of *Ganaspis sp.1* venom proteins

Protein minimization is a concept originally described in synthetic biology, in which proteins are artificially reduced to the minimum sequence that will support a selected function [109,110]. In our analysis, we found a subset of 17 venom proteins that appear to represent evolutionarily minimized proteins (Table 3). In comparison with the homologous proteins in *D. melanogaster*, these "minimized" venom proteins are smaller (with an average size of 37.3 kDa for the venom proteins compared to an average size 127.9 kDa for the homologs, $p= 3.559e-04$) and encode a decreased number of total conserved domains (3.0 vs. 10.9 total conserved domains per protein, $p = 3.961e-05$). We find that 14 of these minimized proteins contain conserved LRR domains and are therefore predicted to be involved in protein-ligand interactions. The other minimized proteins contain CUB (cd00041), EGF/CA (pfam07645) or LIM (cd08368) family conserved domains, all of which are also predicted to mediate protein binding [111–114].

The minimized *Ganaspis sp.1* venom proteins also show reduced domain complexity when compared to their *D. melanogaster* homologs; the *D. melanogaster* proteins have an average of 2.1 distinct conserved domains per protein, whereas the minimized *Ganaspis sp.1* venom proteins all contain only a single type of conserved domain (Table 3; $p = 3.103e-05$). More specifically, 12 of the proteins are missing domains that are present in the host homologs, including both conserved functional domains and predicted transmembrane domains (Figure 4). These alterations in protein domain architecture are often seen in dominant negative protein isoforms [115–118]. We propose that these minimized venom proteins may act as dominant negatives by directly binding to, and thereby sequestering, host proteins. This dominant negative strategy is also widely used by bacterial and viral pathogens, where pathogen proteins inhibit host immunity by interacting with host immune factors [119–122].

Indeed, several of the *D. melanogaster* homologs are involved in immune responses (Table 3, [123–131]) and we would therefore predict that interfering with activity of these genes would aid *Ganaspis sp.1* in evading host immunity. Three of these host immune genes encode Toll-like receptors (TLRs: *Toll* [*Tl*], *Toll-7* and *18 wheeler* [*18w*]). *D. melanogaster* TLRs are transmembrane proteins, and receive signals by the binding of ligands to their extracellular LRR domains [132,133]. The signal is then transduced through the interaction of the intracellular Toll/Interleukin-1 Receptor homology domain (TIR) with the downstream adaptor protein MyD88, leading to the activation of NF κ B transcription factors and subsequent immune signaling [134,135]. The venom homologs of these TLRs have highly conserved LRRs, but lack transmembrane or TIR domains (Figure 4). We would predict that these proteins would bind to endogenous TLR ligands, but would be unable to activate downstream signaling and could act to sequester ligands and subsequently dampen host signaling in response to infection.

Conclusions

Our bioinformatic analysis reveals the complexity of the evolution and function of

parasitoid venoms. First, our results suggest that the evolution of venom composition in a single parasitoid species is driven by multiple mechanisms. *Ganaspis sp. 1* venom encoding genes show evidence of gene multifunctionality and co-option, along with neofunctionalization following gene duplication. Additionally, several of these venom genes appear to have evolved as *de novo* or horizontally transferred genes. Second, our analyses suggest several potential virulence functions of *Ganaspis sp. 1* venoms. The enriched functional domains found in these proteins suggest that venoms may act to inhibit fly immunity through antagonizing calcium signaling dependent immune cell activation or blocking serine protease cascades. Furthermore, we have uncovered evidence suggesting that the venom paralogs arising from recent gene duplication events may have evolved dominant negative traits that would interfere with essential host functions. Finally, we have identified a subset of venom proteins that appear to have undergone a minimization process resulting in smaller and less complex venom proteins that we would predict could interfere with host immune function by sequestering immune receptor ligands. Testing the predictions we have made based on our analyses will provide further insight into the function of parasitoid venoms, and open new areas of research within this exciting field.

Figure Legends

Figure 1. Venom protein characterization. (A,B) The molecular weight (A) and isoelectric point (B) of venom proteins (right) compared to a random subset of non-venom proteins (left). (C) The phylogenetic origin of the the best BLAST hit for each venom protein. *Hymenoptera* indicates that the closest homolog is from an insect within the order Hymenoptera, *Other insect* indicates that the closet homolog is from an insect from any other order, *Non-insect* indicates that the closest homolog is from a non-insect species, and *None* indicates that there were no homologous proteins found in the NCBI database. (D,E) The molecular weight (D) and isoelectric point (E) of novel evolutionarily novel venom proteins (right) compared to venom proteins showing evidence of sequence conservation (left). * indicates $p < 0.05$ by Welch Two Sample t-test.

Figure 2. Phylogenetic analysis of putative horizontally transferred genes. (A-C) Phylogenetic trees showing the closest homologs of the venom proteins from the *Non-insect* class of BLAST results: (A) *G1 venom comp112*, (B) *G1 venom comp3581*, (C) *G1 venom comp11645*. Numbers represent bootstrap values (with 1000 replicates).

Figure 3. Oxoglutarate dehydrogenase sequence alignments. (A,B) Sequence alignments for the (A) Transketolase PYR domain and (B) Thiamine pyrophosphate binding domain from *comp9906*, the non-venom (Cellular) oxoglutarate dehydrogenase paralog, and *comp1724*, the venom paralog of oxoglutarate dehydrogenase (Venom). Functional residues are highlighted, demonstrating the specific lack of conservation of functional residues in the Transketolase PYR domain in the venom paralog.

Figure 4. *Ganaspis sp.1* Toll-like venom proteins. A schematic diagram illustrating the differences between the *Ganaspis sp.1* Toll-like venom proteins, and their *Drosophila* homologs. The *Ganaspis sp.1* Toll-like venom proteins are smaller and have fewer LRR domains (green boxes). LRR domains will predicted homologous ligand affinities are shown binding to the same

hypothetical ligands (L1, L2, L3). The Toll-like venom proteins additionally do not contain predicted transmembrane domains or TIR domains (red boxes), and so would not be predicted to signal through the MyD88 protein to potentiate NF κ B activity.

Acknowledgements

We would like to thank Drs. Jeff Helms and Pennapa Manitchotpisit for their assistance. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors, and was supported by the School of Biological Sciences at Illinois State University.

References

- [1] Y. Carton, M. Bouletreau, J.J.M. Van Alphen, J.C. Van Lenteren, The *Drosophila* parasitic wasps, in: Genet. Biol. *Drosoph.*, 3rd ed., Academic Press, London, 1986.
- [2] E.L. Danneels, D.B. Rivers, D.C. De Graaf, Venom Proteins of the Parasitoid Wasp *Nasonia vitripennis*: Recent Discovery of an Untapped Pharmacopee, *Toxins*. 2 (2010) 494–516. doi:10.3390/toxins2040494.
- [3] N.E. Beckage, Modulation of immune responses to parasitoids by polydnviruses, *Parasitology*. 116 Suppl (1998) S57-64.
- [4] N.T. Mortimer, Parasitoid wasp virulence: A window into fly immunity, *Fly (Austin)*. 7 (2013) 242–248.
- [5] N.T. Mortimer, J. Goecks, B.Z. Kacsoh, J.A. Mobley, G.J. Bowersock, J. Taylor, T.A. Schlenke, Parasitoid wasp venom SERCA regulates *Drosophila* calcium levels and inhibits cellular immunity, *Proc. Natl. Acad. Sci.* 110 (2013) 9427–9432. doi:10.1073/pnas.1222351110.
- [6] D. Colinet, A. Schmitz, D. Depoix, D. Crochard, M. Poirié, Convergent Use of RhoGAP Toxins by Eukaryotic Parasites and Bacterial Pathogens, *PLOS Pathog.* 3 (2007) e203. doi:10.1371/journal.ppat.0030203.
- [7] D. Colinet, A. Dubuffet, D. Cazes, S. Moreau, J.-M. Drezen, M. Poirié, A serpin from the parasitoid wasp *Leptopilina boulardi* targets the *Drosophila* phenoloxidase cascade, *Dev. Comp. Immunol.* 33 (2009) 681–689. doi:10.1016/j.dci.2008.11.013.
- [8] E. Vass, A.J. Nappi, Developmental and immunological aspects of *Drosophila*-parasitoid relationships, *J. Parasitol.* 86 (2000) 1259–1270. doi:10.1645/0022-3395(2000)086[1259:DAIAOD]2.0.CO;2.
- [9] O. Schmidt, U. Theopold, M. Strand, Innate immunity and its evasion and suppression by hymenopteran endoparasitoids, *BioEssays News Rev. Mol. Cell. Dev. Biol.* 23 (2001) 344–351. doi:10.1002/bies.1049.
- [10] D.B. Rivers, L. Ruggiero, M. Hayes, The ectoparasitic wasp *Nasonia vitripennis* (Walker) (Hymenoptera: Pteromalidae) differentially affects cells mediating the immune response of its flesh fly host, *Sarcophaga bullata* Parker (Diptera: Sarcophagidae), *J. Insect Physiol.* 48 (2002) 1053–1064. doi:10.1016/S0022-1910(02)00193-2.
- [11] E.O. Martinson, D. Wheeler, J. Wright, Mrinalini, A.L. Siebert, J.H. Werren, *Nasonia vitripennis* venom causes targeted gene expression changes in its fly host, *Mol. Ecol.* 23 (2014) 5918–5930. doi:10.1111/mec.12967.
- [12] 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 (2007) e158. doi:10.1371/journal.ppat.0030158.
- [13] R.M. Rizki, T.M. Rizki, Parasitoid virus-like particles destroy *Drosophila* cellular immunity., *Proc. Natl. Acad. Sci. U. S. A.* 87 (1990) 8388–8392.
- [14] D.B. Rivers, D.L. Denlinger, Redirection of metabolism in the flesh fly, *Sarcophaga bullata*, following envenomation by the ectoparasitoid *Nasonia vitripennis* and correlation of metabolic effects with the diapause status of the host, *J. Insect Physiol.* 40 (1994) 207–215. doi:10.1016/0022-1910(94)90044-2.
- [15] S.J.M. Moreau, S. Asgari, Venom Proteins from Parasitoid Wasps and Their Biological Functions, *Toxins*. 7 (2015) 2385–2412. doi:10.3390/toxins7072385.
- [16] S.B. Vinson, G.F. Iwantsch, Host Regulation by Insect Parasitoids, *Q. Rev. Biol.* 55 (1980) 143–165. doi:10.1086/411731.
- [17] D.B. Rivers, D.L. Denlinger, Developmental fate of the flesh fly, *Sarcophaga bullata*, envenomated by the pupal ectoparasitoid, *Nasonia vitripennis*, *J. Insect Physiol.* 40 (1994) 121–127. doi:10.1016/0022-1910(94)90083-3.

- [18] N.E. Beckage, D.B. Gelman, Wasp Parasitoid Disruption of Host Development: Implications for New Biologically Based Strategies for Insect Control, *Annu. Rev. Entomol.* 49 (2004) 299–330. doi:10.1146/annurev.ento.49.061802.123324.
- [19] D.B. Rivers, D.L. Denlinger, Venom-Induced Alterations in Fly Lipid Metabolism and Its Impact on Larval Development of the Ectoparasitoid *Nasonia vitripennis* (Walker) (Hymenoptera: Pteromalidae), *J. Invertebr. Pathol.* 66 (1995) 104–110. doi:10.1006/jipa.1995.1071.
- [20] M.P. Dani, J.P. Edwards, E.H. Richards, Hydrolase activity in the venom of the pupal endoparasitoid wasp, *Pimpla hypochondriaca*, *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 141 (2005) 373–381. doi:10.1016/j.cbpc.2005.04.010.
- [21] Y. Nakamatsu, S. Fujii, T. Tanaka, Larvae of an endoparasitoid, *Cotesia kariyai* (Hymenoptera: Braconidae), feed on the host fat body directly in the second stadium with the help of teratocytes, *J. Insect Physiol.* 48 (2002) 1041–1052.
- [22] Mrinalini, A.L. Siebert, J. Wright, E. Martinson, D. Wheeler, J.H. Werren, Parasitoid Venom Induces Metabolic Cascades in Fly Hosts, *Metabolomics Off. J. Metabolomic Soc.* 11 (2015) 350–366. doi:10.1007/s11306-014-0697-z.
- [23] F. Libersat, Wasp uses venom cocktail to manipulate the behavior of its cockroach prey, *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 189 (2003) 497–508. doi:10.1007/s00359-003-0432-0.
- [24] R. Gal, F. Libersat, A Parasitoid Wasp Manipulates the Drive for Walking of Its Cockroach Prey, *Curr. Biol.* 18 (2008) 877–882. doi:10.1016/j.cub.2008.04.076.
- [25] F. Libersat, A. Delago, R. Gal, Manipulation of host behavior by parasitic insects and insect parasites, *Annu. Rev. Entomol.* 54 (2009) 189–207. doi:10.1146/annurev.ento.54.110807.090556.
- [26] S. Korenko, S. Pekár, A Parasitoid Wasp Induces Overwintering Behaviour in Its Spider Host, *PLOS ONE*. 6 (2011) e24628. doi:10.1371/journal.pone.0024628.
- [27] M.O. Gonzaga, A.P. Loffredo, A.M. Penteado-Dias, J.C.F. Cardoso, Host behavior modification of *Achaeearanea tingo* (Araneae: Theridiidae) induced by the parasitoid wasp *Zatypota alborhombarta* (Hymenoptera: Ichneumonidae), *Entomol. Sci.* 19 (2016) 133–137. doi:10.1111/ens.12178.
- [28] T.G. Kloss, M.O. Gonzaga, L.L. de Oliveira, C.F. Sperber, Proximate mechanism of behavioral manipulation of an orb-weaver spider host by a parasitoid wasp, *PLOS ONE*. 12 (2017) e0171336. doi:10.1371/journal.pone.0171336.
- [29] J. Goecks, N.T. Mortimer, J.A. Mobley, G.J. Bowersock, J. Taylor, T.A. Schlenke, Integrative approach reveals composition of endoparasitoid wasp venoms, *PLOS One*. 8 (2013) e64125.
- [30] G.R. Burke, M.R. Strand, Systematic analysis of a wasp parasitism arsenal, *Mol. Ecol.* 23 (2014) 890–901. doi:10.1111/mec.12648.
- [31] D. Colinet, C. Anselme, E. Deleury, D. Mancini, J. Poulain, C. Azéma-Dossat, M. Belghazi, S. Tares, F. Pennacchio, M. Poirié, J.-L. Gatti, Identification of the main venom protein components of *Aphidius ervi*, a parasitoid wasp of the aphid model *Acyrtosiphon pisum*, *BMC Genomics*. 15 (2014) 342. doi:10.1186/1471-2164-15-342.
- [32] M.E. Heavner, G. Gueguen, R. Rajwani, P.E. Pagan, C. Small, S. Govind, Partial venom gland transcriptome of a *Drosophila* parasitoid wasp, *Leptopilina heterotoma*, reveals novel and shared bioactive profiles with stinging Hymenoptera, *Gene*. 526 (2013) 195–204. doi:10.1016/j.gene.2013.04.080.
- [33] J.-Y. Zhu, Q. Fang, L. Wang, C. Hu, G.-Y. Ye, Proteomic analysis of the venom from the endoparasitoid wasp *Pteromalus puparum* (Hymenoptera: Pteromalidae), *Arch. Insect Biochem. Physiol.* 75 (2010) 28–44. doi:10.1002/arch.20380.

- [34] A.M. Crawford, R. Brauning, G. Smolenski, C. Ferguson, D. Barton, T.T. Wheeler, A. McCulloch, The constituents of *Microctonus* sp. parasitoid venoms, *Insect Mol. Biol.* 17 (2008) 313–324. doi:10.1111/j.1365-2583.2008.00802.x.
- [35] D.C. De Graaf, M. Aerts, M. Brunain, C.A. Desjardins, F.J. Jacobs, J.H. Werren, B. Devreese, Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from bioinformatic and proteomic studies, *Insect Mol. Biol.* 19 (2010) 11–26. doi:10.1111/j.1365-2583.2009.00914.x.
- [36] B. Vincent, M. Kaeslin, T. Roth, M. Heller, J. Poulain, F. Cousserans, J. Schaller, M. Poirié, B. Lanzrein, J.-M. Drezen, S.J.M. Moreau, The venom composition of the parasitic wasp *Chelonus inanitus* resolved by combined expressed sequence tags analysis and proteomic approach, *BMC Genomics.* 11 (2010) 693. doi:10.1186/1471-2164-11-693.
- [37] M. Kaeslin, M. Reinhard, D. Bühler, T. Roth, R. Pfister-Wilhelm, B. Lanzrein, Venom of the egg-larval parasitoid *Chelonus inanitus* is a complex mixture and has multiple biological effects, *J. Insect Physiol.* 56 (2010) 686–694. doi:10.1016/j.jinsphys.2009.12.005.
- [38] Z. Yan, Q. Fang, L. Wang, J. Liu, Y. Zhu, F. Wang, F. Li, J.H. Werren, G. Ye, Insights into the venom composition and evolution of an endoparasitoid wasp by combining proteomic and transcriptomic analyses, *Sci. Rep.* 6 (2016) srep19604. doi:10.1038/srep19604.
- [39] Mrinalini, J.H. Werren, Parasitoid Wasps and Their Venoms, in: P. Gopalakrishnakone, A. Malhotra (Eds.), *Evol. Venom. Anim. Their Toxins*, Springer Netherlands, 2015: pp. 1–26. doi:10.1007/978-94-007-6727-0_2-1.
- [40] E.O. Martinson, Mrinalini, Y.D. Kelkar, C.-H. Chang, J.H. Werren, The Evolution of Venom by Co-option of Single-Copy Genes, *Curr. Biol.* 0 (2017). doi:10.1016/j.cub.2017.05.032.
- [41] Z. Yan, Q. Fang, Y. Liu, S. Xiao, L. Yang, F. Wang, C. An, J.H. Werren, G. Ye, A Venom Serpin Splicing Isoform of the Endoparasitoid Wasp *Pteromalus puparum* Suppresses Host Prophenoloxidase Cascade by Forming Complexes with Host Hemolymph Proteinases, *J. Biol. Chem.* 292 (2017) 1038–1051. doi:10.1074/jbc.M116.739565.
- [42] 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 (2016) 1042–1052. doi:10.1093/molbev/msv348.
- [43] K.R. Hopper, United States Department of Agriculture-Agricultural Research Service research on biological control of arthropods, *Pest Manag. Sci.* 59 (2003) 643–653.
- [44] N.A. Ratcliffe, C.B. Mello, E.S. Garcia, T.M. Butt, P. Azambuja, Insect natural products and processes: New treatments for human disease, *Insect Biochem. Mol. Biol.* 41 (2011) 747–769. doi:10.1016/j.ibmb.2011.05.007.
- [45] Y. Carton, A.J. Nappi, *Drosophila* cellular immunity against parasitoids, *Parasitol. Today.* 13 (1997) 218–227.
- [46] Y. Carton, A.J. Nappi, Immunogenetic aspects of the cellular immune response of *Drosophila* against parasitoids, *Immunogenetics.* 52 (2001) 157–164.
- [47] E.S. Keebaugh, T.A. Schlenke, Insights from natural host–parasite interactions: The *Drosophila* model, *Dev. Comp. Immunol.* 42 (2014) 111–123. doi:10.1016/j.dci.2013.06.001.
- [48] M.J. Williams, *Drosophila* hemopoiesis and cellular immunity, *J. Immunol. Baltim. Md* 1950. 178 (2007) 4711–4716.
- [49] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, *BMC Bioinformatics.* 12 (2011) 323. doi:10.1186/1471-2105-12-323.
- [50] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, in: *Proteomics Protoc. Handb.*, Humana Press, 2005: pp. 571–607. doi:10.1385/1-59259-890-0:571.
- [51] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *J. Mol. Biol.* 215 (1990) 403–410. doi:10.1016/S0022-2836(05)80360-2.

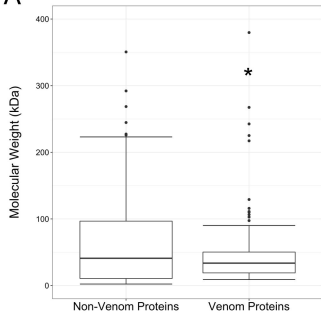
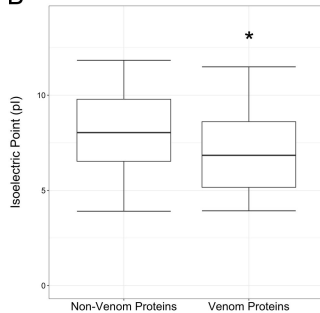
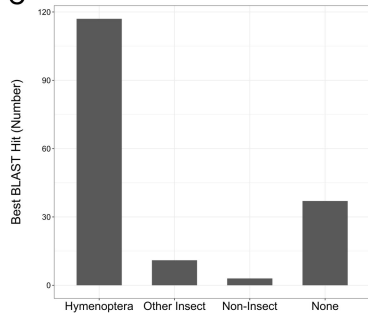
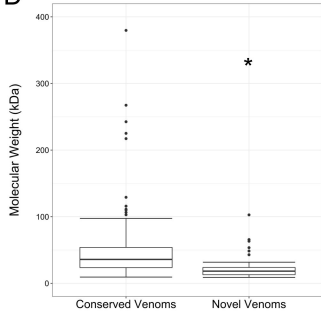
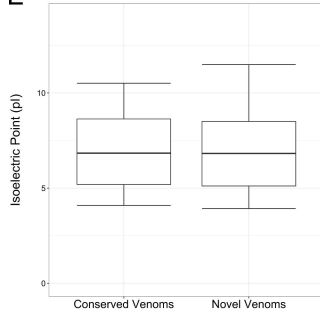
- [52] A. Marchler-Bauer, S. Lu, J.B. Anderson, F. Chitsaz, M.K. Derbyshire, C. DeWeese-Scott, J.H. Fong, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, J.D. Jackson, Z. Ke, C.J. Lanczycki, F. Lu, G.H. Marchler, M. Mullokandov, M.V. Omelchenko, C.L. Robertson, J.S. Song, N. Thanki, R.A. Yamashita, D. Zhang, N. Zhang, C. Zheng, S.H. Bryant, CDD: a Conserved Domain Database for the functional annotation of proteins, *Nucleic Acids Res.* 39 (2011) D225–D229. doi:10.1093/nar/gkq1189.
- [53] A. Marchler-Bauer, M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, R.C. Geer, J. He, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, S.H. Bryant, CDD: NCBI's conserved domain database, *Nucleic Acids Res.* 43 (2015) D222–D226. doi:10.1093/nar/gku1221.
- [54] K. Tamura, G. Stecher, D. Peterson, A. Filipinski, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis version 6.0, *Mol. Biol. Evol.* 30 (2013) 2725–2729. doi:10.1093/molbev/mst197.
- [55] J. Matysiak, J. Hajduk, Ł. Pietrzak, C.E.H. Schmelzer, Z.J. Kokot, Shotgun proteome analysis of honeybee venom using targeted enrichment strategies, *Toxicon.* 90 (2014) 255–264. doi:10.1016/j.toxicon.2014.08.069.
- [56] E. Habermann, Bee and Wasp Venoms, *Science.* 177 (1972) 314–322. doi:10.1126/science.177.4046.314.
- [57] N.B. Dias, B.M. de Souza, P.C. Gomes, M.S. Palma, Peptide diversity in the venom of the social wasp *Polybia paulista* (Hymenoptera): A comparison of the intra- and inter-colony compositions, *Peptides.* 51 (2014) 122–130. doi:10.1016/j.peptides.2013.10.029.
- [58] A. Touchard, J.M.S. Koh, S.R. Aili, A. Dejean, G.M. Nicholson, J. Orivel, P. Escoubas, The complexity and structural diversity of ant venom peptidomes is revealed by mass spectrometry profiling., (2015). doi:10.1002/rcm.7116.
- [59] K. Konno, K. Kazuma, K. Nihei, Peptide Toxins in Solitary Wasp Venoms, *Toxins.* 8 (2016). doi:10.3390/toxins8040114.
- [60] L. Zhao, P. Saelao, C.D. Jones, D.J. Begun, Origin and spread of de novo genes in *Drosophila melanogaster* populations, *Science.* 343 (2014) 769–772. doi:10.1126/science.1248286.
- [61] T. Domazet-Loso, D. Tautz, An Evolutionary Analysis of Orphan Genes in *Drosophila*, *Genome Res.* 13 (2003) 2213–2219. doi:10.1101/gr.1311003.
- [62] J. Ruiz-Orera, J. Hernandez-Rodriguez, C. Chiva, E. Sabidó, I. Kondova, R. Bontrop, T. Marqués-Bonet, M.M. Albà, Origins of De Novo Genes in Human and Chimpanzee, *PLOS Genet.* 11 (2015) e1005721. doi:10.1371/journal.pgen.1005721.
- [63] D.G. Knowles, A. McLysaght, Recent de novo origin of human protein-coding genes, *Genome Res.* 19 (2009) 1752–1759. doi:10.1101/gr.095026.109.
- [64] Z.W. Arendsee, L. Li, E.S. Wurtele, Coming of age: orphan genes in plants, *Trends Plant Sci.* 19 (2014) 698–708. doi:10.1016/j.tplants.2014.07.003.
- [65] M.A. Campbell, W. Zhu, N. Jiang, H. Lin, S. Ouyang, K.L. Childs, B.J. Haas, J.P. Hamilton, C.R. Buell, Identification and Characterization of Lineage-Specific Genes within the Poaceae, *Plant Physiol.* 145 (2007) 1311–1322. doi:10.1104/pp.107.104513.
- [66] P.J. Keeling, J.D. Palmer, Horizontal gene transfer in eukaryotic evolution, *Nat. Rev. Genet.* 9 (2008) 605–618. doi:10.1038/nrg2386.
- [67] H. Ochman, J.G. Lawrence, E.A. Groisman, Lateral gene transfer and the nature of bacterial innovation, *Nature.* 405 (2000) 299–304. doi:10.1038/35012500.
- [68] P.F.L. Boreham, D.J. Stenzel, Blastocystis in Humans and Animals: Morphology, Biology, and Epizootiology, in: J.R.B. and R. Muller (Ed.), *Adv. Parasitol.*, Academic Press, 1993: pp. 1–70. <http://www.sciencedirect.com/science/article/pii/S0065308X08602067> (accessed December 13, 2016).

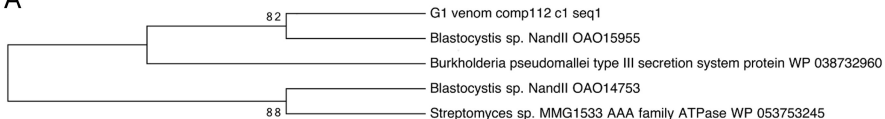
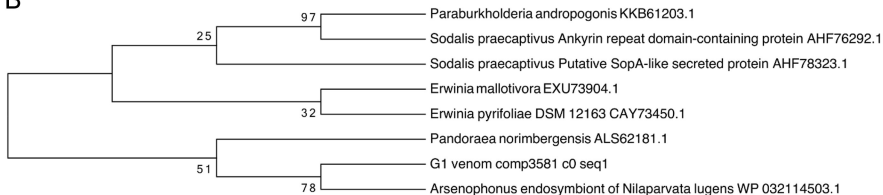
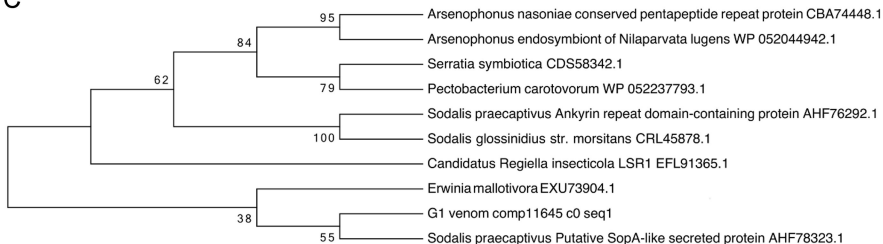
- [69] C. Noël, F. Dufernez, D. Gerbod, V.P. Edgcomb, P. Delgado-Viscogliosi, L.-C. Ho, M. Singh, R. Wintjens, M.L. Sogin, M. Capron, R. Pierce, L. Zenner, E. Viscogliosi, Molecular Phylogenies of Blastocystis Isolates from Different Hosts: Implications for Genetic Diversity, Identification of Species, and Zoonosis, *J. Clin. Microbiol.* 43 (2005) 348–355. doi:10.1128/JCM.43.1.348-355.2005.
- [70] C. Dale, S.C. Welburn, The endosymbionts of tsetse flies: manipulating host–parasite interactions, *Int. J. Parasitol.* 31 (2001) 628–631. doi:10.1016/S0020-7519(01)00151-5.
- [71] E. Nováková, V. Hypša, N.A. Moran, Arsenophonus, an emerging clade of intracellular symbionts with a broad host distribution, *BMC Microbiol.* 9 (2009) 143. doi:10.1186/1471-2180-9-143.
- [72] G. Koutsovoulos, S. Kumar, D.R. Laetsch, L. Stevens, J. Daub, C. Conlon, H. Maroon, F. Thomas, A.A. Aboobaker, M. Blaxter, No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*, *Proc. Natl. Acad. Sci.* 113 (2016) 5053–5058. doi:10.1073/pnas.1600338113.
- [73] P.-Y. Dupont, M.P. Cox, Genomic Data Quality Impacts Automated Detection of Lateral Gene Transfer in Fungi, *G3 Genes Genomes Genet.* (2017) g3.116.038448. doi:10.1534/g3.116.038448.
- [74] A. Crisp, C. Boschetti, M. Perry, A. Tunnacliffe, G. Micklem, Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes, *Genome Biol.* 16 (2015) 50. doi:10.1186/s13059-015-0607-3.
- [75] J.C.D. Hotopp, M.E. Clark, D.C.S.G. Oliveira, J.M. Foster, P. Fischer, M.C.M. Torres, J.D. Giebel, N. Kumar, N. Ishmael, S. Wang, J. Ingram, R.V. Nene, J. Shepard, J. Tomkins, S. Richards, D.J. Spiro, E. Ghedin, B.E. Slatko, H. Tettelin, J.H. Werren, Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes, *Science.* 317 (2007) 1753–1756. doi:10.1126/science.1142490.
- [76] J.C. Dunning Hotopp, Horizontal gene transfer between bacteria and animals, *Trends Genet. TIG.* 27 (2011) 157–163. doi:10.1016/j.tig.2011.01.005.
- [77] N. Kondo, N. Nikoh, N. Ijichi, M. Shimada, T. Fukatsu, Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 14280–14285. doi:10.1073/pnas.222228199.
- [78] N. Nikoh, K. Tanaka, F. Shibata, N. Kondo, M. Hizume, M. Shimada, T. Fukatsu, *Wolbachia* genome integrated in an insect chromosome: Evolution and fate of laterally transferred endosymbiont genes, *Genome Res.* 18 (2008) 272–280. doi:10.1101/gr.7144908.
- [79] J.H. Werren, S. Richards, C.A. Desjardins, O. Niehuis, J. Gadau, J.K. Colbourne, T.N.G.W. Group, Functional and Evolutionary Insights from the Genomes of Three Parasitoid *Nasonia* Species, *Science.* 327 (2010) 343–348. doi:10.1126/science.1178028.
- [80] J.-M. Drezen, J. Gauthier, T. Josse, A. Bézier, E. Herniou, E. Huguet, Foreign DNA acquisition by invertebrate genomes, *J. Invertebr. Pathol.* (n.d.). doi:10.1016/j.jip.2016.09.004.
- [81] R. Acuña, B.E. Padilla, C.P. Flórez-Ramos, J.D. Rubio, J.C. Herrera, P. Benavides, S.-J. Lee, T.H. Yeats, A.N. Egan, J.J. Doyle, J.K.C. Rose, Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee, *Proc. Natl. Acad. Sci.* 109 (2012) 4197–4202. doi:10.1073/pnas.1121190109.
- [82] B.F. Sun, J.H. Xiao, S.M. He, L. Liu, R.W. Murphy, D.W. Huang, Multiple ancient horizontal gene transfers and duplications in lepidopteran species, *Insect Mol. Biol.* 22 (2013) 72–87. doi:10.1111/imb.12004.
- [83] C. Zhao, D. Doucet, O. Mittapalli, Characterization of horizontally transferred β -fructofuranosidase (*ScrB*) genes in *Agrilus planipennis*, *Insect Mol. Biol.* 23 (2014) 821–832. doi:10.1111/imb.12127.
- [84] S. Chou, M.D. Daugherty, S.B. Peterson, J. Biboy, Y. Yang, B.L. Jutras, L.K. Fritz-Laylin, M.A. Ferrin, B.N. Harding, C. Jacobs-Wagner, X.F. Yang, W. Vollmer, H.S. Malik, J.D.

- Mougous, Transferred interbacterial antagonism genes augment eukaryotic innate immune function, *Nature*. 518 (2015) 98–101. doi:10.1038/nature13965.
- [85] M. Lynch, J.S. Conery, The Evolutionary Fate and Consequences of Duplicate Genes, *Science*. 290 (2000) 1151–1155. doi:10.1126/science.290.5494.1151.
- [86] F. Seifert, E. Ciszak, L. Korotchkina, R. Golbik, M. Spinka, P. Dominiak, S. Sidhu, J. Brauer, M.S. Patel, K. Tittmann, Phosphorylation of Serine 264 Impedes Active Site Accessibility in the E1 Component of the Human Pyruvate Dehydrogenase Multienzyme Complex, *Biochemistry (Mosc.)*. 46 (2007) 6277–6287. doi:10.1021/bi700083z.
- [87] R.G. McCartney, J.E. Rice, S.J. Sanderson, V. Bunik, H. Lindsay, J.G. Lindsay, Subunit Interactions in the Mammalian α -Ketoglutarate Dehydrogenase Complex: Evidence for Direct Association of the α -Ketoglutarate Dehydrogenase and Dihydrolipoamide Dehydrogenase Components, *J. Biol. Chem.* 273 (1998) 24158–24164. doi:10.1074/jbc.273.37.24158.
- [88] P. Arjunan, K. Chandrasekhar, M. Sax, A. Brunskill, N. Nemeria, F. Jordan, W. Furey, Structural determinants of enzyme binding affinity: the E1 component of pyruvate dehydrogenase from *Escherichia coli* in complex with the inhibitor thiamin thiazolone diphosphate, *Biochemistry (Mosc.)*. 43 (2004) 2405–2411. doi:10.1021/bi030200y.
- [89] F. Mastrogiacomo, C. Bergeron, S.J. Kish, Brain α -Ketoglutarate Dehydrogenase Complex Activity in Alzheimer's Disease, *J. Neurochem.* 61 (1993) 2007–2014. doi:10.1111/j.1471-4159.1993.tb07436.x.
- [90] D. Lonsdale, Thiamine and magnesium deficiencies: keys to disease, *Med. Hypotheses*. 84 (2015) 129–134. doi:10.1016/j.mehy.2014.12.004.
- [91] M.S. Patel, N.S. Nemeria, W. Furey, F. Jordan, The Pyruvate Dehydrogenase Complexes: Structure-based Function and Regulation, *J. Biol. Chem.* 289 (2014) 16615–16623. doi:10.1074/jbc.R114.563148.
- [92] C. Labrosse, K. Stasiak, J. Lesobre, A. Grangeia, E. Huguet, J.M. Drezen, M. Poirie, A RhoGAP protein as a main immune suppressive factor in the *Leptopilina boulardi* (Hymenoptera, Figitidae)–*Drosophila melanogaster* interaction, *Insect Biochem. Mol. Biol.* 35 (2005) 93–103. doi:10.1016/j.ibmb.2004.10.004.
- [93] 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 (2005) 161–170. doi:10.1016/j.jinsphys.2004.10.004.
- [94] J. Lytton, M. Westlin, M.R. Hanley, Thapsigargin inhibits the sarcoplasmic or endoplasmic reticulum Ca-ATPase family of calcium pumps., *J. Biol. Chem.* 266 (1991) 17067–17071.
- [95] W.J. Andrews, C.A. Bradley, E. Hamilton, C. Daly, T. Mallon, D.J. Timson, A calcium-dependent interaction between calmodulin and the calponin homology domain of human IQGAP1, *Mol. Cell. Biochem.* 371 (2012) 217–223. doi:10.1007/s11010-012-1438-0.
- [96] K. Takahashi, M. Inuzuka, T. Ingi, Cellular signaling mediated by calphoglin-induced activation of IPP and PGM, *Biochem. Biophys. Res. Commun.* 325 (2004) 203–214. doi:10.1016/j.bbrc.2004.10.021.
- [97] V. Gerke, S.E. Moss, Annexins: from structure to function, *Physiol. Rev.* 82 (2002) 331–371. doi:10.1152/physrev.00030.2001.
- [98] B. Lemaitre, J. Hoffmann, The Host Defense of *Drosophila melanogaster*, *Annu. Rev. Immunol.* 25 (2007) 697–743. doi:10.1146/annurev.immunol.25.022106.141615.
- [99] M.J. Gorman, S.M. Paskewitz, Serine proteases as mediators of mosquito immune responses, *Insect Biochem. Mol. Biol.* 31 (2001) 257–262. doi:10.1016/S0965-1748(00)00145-4.
- [100] M.R. Kanost, H. Jiang, Clip-domain serine proteases as immune factors in insect hemolymph, *Curr. Opin. Insect Sci.* 11 (2015) 47–55. doi:10.1016/j.cois.2015.09.003.

- [101] E. De Gregorio, S.-J. Han, W.-J. Lee, M.-J. Baek, T. Osaki, S.-I. Kawabata, B.-L. Lee, S. Iwanaga, B. Lemaitre, P.T. Brey, An Immune-Responsive Serpin Regulates the Melanization Cascade in *Drosophila*, *Dev. Cell.* 3 (2002) 581–592. doi:10.1016/S1534-5807(02)00267-8.
- [102] H. Tang, Z. Kambris, B. Lemaitre, C. Hashimoto, A Serpin that Regulates Immune Melanization in the Respiratory System of *Drosophila*, *Dev. Cell.* 15 (2008) 617–626. doi:10.1016/j.devcel.2008.08.017.
- [103] H. Mathé-Hubert, D. Colinet, E. Deleury, M. Belghazi, M. Ravallec, J. Poulain, C. Dossat, M. Poirié, J.-L. Gatti, Comparative venomomics of *Psytalia lounsburyi* and *P. concolor*, two olive fruit fly parasitoids: a hypothetical role for a GH1 β -glucosidase, *Sci. Rep.* 6 (2016) 35873. doi:10.1038/srep35873.
- [104] B. Kobe, A.V. Kajava, The leucine-rich repeat as a protein recognition motif, *Curr. Opin. Struct. Biol.* 11 (2001) 725–732. doi:10.1016/S0959-440X(01)00266-4.
- [105] E. Özkan, R.A. Carrillo, C.L. Eastman, R. Weiszmann, D. Waghray, K.G. Johnson, K. Zinn, S.E. Celniker, K.C. Garcia, An Extracellular Interactome of Immunoglobulin and LRR Proteins Reveals Receptor-Ligand Networks, *Cell.* 154 (2013) 228–239. doi:10.1016/j.cell.2013.06.006.
- [106] P. Fucini, C. Renner, C. Herberhold, A.A. Noegel, T.A. Holak, The repeating segments of the F-actin cross-linking gelation factor (ABP-120) have an immunoglobulin-like fold., *Nat. Struct. Biol.* 4 (1997) 223–230. doi:10.1038/nsb0397-223.
- [107] Y. Feng, C.A. Walsh, The many faces of filamin: A versatile molecular scaffold for cell motility and signalling, *Nat. Cell Biol.* 6 (2004) 1034–1038. doi:10.1038/ncb1104-1034.
- [108] S. Light, R. Sagit, S.S. Ithychanda, J. Qin, A. Elofsson, The evolution of filamin – A protein domain repeat perspective, *J. Struct. Biol.* 179 (2012) 289–298. doi:10.1016/j.jsb.2012.02.010.
- [109] M.A. Starovasnik, A.C. Braisted, J.A. Wells, Structural mimicry of a native protein by a minimized binding domain, *Proc. Natl. Acad. Sci.* 94 (1997) 10080–10085.
- [110] B.C. Cunningham, J.A. Wells, Minimized proteins, *Curr. Opin. Struct. Biol.* 7 (1997) 457–462. doi:10.1016/S0959-440X(97)80107-8.
- [111] P. Bork, G. Beckmann, The CUB Domain, *J. Mol. Biol.* 231 (1993) 539–545. doi:10.1006/jmbi.1993.1305.
- [112] E. Appella, I.T. Weber, F. Blasi, Structure and function of epidermal growth factor-like regions in proteins, *FEBS Lett.* 231 (1988) 1–4. doi:10.1016/0014-5793(88)80690-2.
- [113] R. Feuerstein, X. Wang, D. Song, N.E. Cooke, S.A. Liebhaber, The LIM/double zinc-finger motif functions as a protein dimerization domain., *Proc. Natl. Acad. Sci. U. S. A.* 91 (1994) 10655–10659.
- [114] K.L. Schmeichel, M.C. Beckerle, The LIM domain is a modular protein-binding interface, *Cell.* 79 (1994) 211–219. doi:10.1016/0092-8674(94)90191-0.
- [115] N.T. Mortimer, K.H. Moberg, The *Drosophila* F-box protein Archipelago controls levels of the Trachealess transcription factor in the embryonic tracheal system, *Dev. Biol.* 312 (2007) 560–571. doi:10.1016/j.ydbio.2007.10.002.
- [116] R.S. Harlander, M. Way, Q. Ren, D. Howe, S.S. Grieshaber, R.A. Heinzen, Effects of Ectopically Expressed Neuronal Wiskott-Aldrich Syndrome Protein Domains on *Rickettsia rickettsii* Actin-Based Motility, *Infect. Immun.* 71 (2003) 1551–1556. doi:10.1128/IAI.71.3.1551-1556.2003.
- [117] M.J. Renzi, L. Feiner, A.M. Koppel, J.A. Raper, A Dominant Negative Receptor for Specific Secreted Semaphorins Is Generated by Deleting an Extracellular Domain from Neuropilin-1, *J. Neurosci.* 19 (1999) 7870–7880.
- [118] R.A. Veitia, Exploring the Molecular Etiology of Dominant-Negative Mutations, *Plant Cell.* 19 (2007) 3843–3851. doi:10.1105/tpc.107.055053.

- [119] C.E. Stebbins, J.E. Galán, Structural mimicry in bacterial virulence, *Nature*. 412 (2001) 701–705. doi:10.1038/35089000.
- [120] C.R. King, M.J. Cohen, G.J. Fonseca, B.S. Dirk, J.D. Dikeakos, J.S. Mymryk, Functional and Structural Mimicry of Cellular Protein Kinase A Anchoring Proteins by a Viral Oncoprotein, *PLOS Pathog.* 12 (2016) e1005621. doi:10.1371/journal.ppat.1005621.
- [121] C. Delevoeye, M. Nilges, P. Dehoux, F. Paumet, S. Perrinet, A. Dautry-Varsat, A. Subtil, SNARE Protein Mimicry by an Intracellular Bacterium, *PLOS Pathog.* 4 (2008) e1000022. doi:10.1371/journal.ppat.1000022.
- [122] Z.A. Hamburger, M.S. Brown, R.R. Isberg, P.J. Bjorkman, Crystal Structure of Invasin: A Bacterial Integrin-Binding Protein, *Science*. 286 (1999) 291–295. doi:10.1126/science.286.5438.291.
- [123] P. Ligoxygakis, P. Bulet, J.-M. Reichhart, Critical evaluation of the role of the Toll-like receptor 18-Wheeler in the host defense of *Drosophila*, *EMBO Rep.* 3 (2002) 666–673. doi:10.1093/embo-reports/kvf130.
- [124] M.J. Williams, A. Rodriguez, D.A. Kimbrell, E.D. Eldon, The 18-wheeler mutation reveals complex antibacterial gene regulation in *Drosophila* host defense, *EMBO J.* 16 (1997) 6120–6130. doi:10.1093/emboj/16.20.6120.
- [125] B. Lemaitre, J.-M. Reichhart, J.A. Hoffmann, *Drosophila* host defense: Differential induction of antimicrobial peptide genes after infection by various classes of microorganisms, *Proc. Natl. Acad. Sci.* 94 (1997) 14614–14619.
- [126] B. Lemaitre, E. Nicolas, L. Michaut, J.-M. Reichhart, J.A. Hoffmann, The Dorsoventral Regulatory Gene Cassette *spätzle/Toll/cactus* Controls the Potent Antifungal Response in *Drosophila* Adults, *Cell*. 86 (1996) 973–983. doi:10.1016/S0092-8674(00)80172-5.
- [127] U. Theopold, D. Li, M. Fabbri, C. Scherfer, O. Schmidt, The coagulation of insect hemolymph, *Cell. Mol. Life Sci. CMLS.* 59 (2002) 363–372. doi:10.1007/s00018-002-8428-4.
- [128] R.H. Moy, B. Gold, J.M. Molleston, V. Schad, K. Yanger, M.-V. Salzano, Y. Yagi, K.A. Fitzgerald, B.Z. Stanger, S.S. Soldan, S. Cherry, Antiviral Autophagy Restricts Rift Valley Fever Virus Infection and Is Conserved from Flies to Mammals, *Immunity*. 40 (2014) 51–65. doi:10.1016/j.immuni.2013.10.020.
- [129] M. Nakamoto, R.H. Moy, J. Xu, S. Bambina, A. Yasunaga, S.S. Shelly, B. Gold, S. Cherry, Virus Recognition by Toll-7 Activates Antiviral Autophagy in *Drosophila*, *Immunity*. 36 (2012) 658–667. doi:10.1016/j.immuni.2012.03.003.
- [130] E. Foley, P.H. O’Farrell, Functional Dissection of an Innate Immune Response by a Genome-Wide RNAi Screen, *PLOS Biol.* 2 (2004) e203. doi:10.1371/journal.pbio.0020203.
- [131] H.-L. Lu, J.B. Wang, M.A. Brown, C. Euerle, R.J.S. Leger, Identification of *Drosophila* Mutants Affecting Defense to an Entomopathogenic Fungus, *Sci. Rep.* 5 (2015) 12350. doi:10.1038/srep12350.
- [132] M. Lewis, C.J. Arnot, H. Beeston, A. McCoy, A.E. Ashcroft, N.J. Gay, M. Gangloff, Cytokine *Spätzle* binds to the *Drosophila* immunoreceptor Toll with a neurotrophin-like specificity and couples receptor activation, *Proc. Natl. Acad. Sci.* 110 (2013) 20461–20466. doi:10.1073/pnas.1317002110.
- [133] C. Parthier, M. Stelter, C. Ursel, U. Fandrich, H. Lilie, C. Breithaupt, M.T. Stubbs, Structure of the Toll-*Spätzle* complex, a molecular hub in *Drosophila* development and innate immunity, *Proc. Natl. Acad. Sci.* 111 (2014) 6281–6286. doi:10.1073/pnas.1320678111.
- [134] J.L. Imler, J.A. Hoffmann, Toll receptors in innate immunity, *Trends Cell Biol.* 11 (2001) 304–311.
- [135] T. Horng, R. Medzhitov, *Drosophila* MyD88 is an adapter in the Toll signaling pathway, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 12654–12658. doi:10.1073/pnas.231471798.

A**B****C****D****E**

A**B****C**

A

Cellular RFEVIDSALSSEYAVMGFEYGYSLDSPYSLVLWEGQFGDFANGAQI
Venom PYTVCNSSLSSEYGVLGFEFGYSMTNPNALVIWEAQFGDFNNTAQC

Cellular MIDQFIASAETKWLRRSSGLVLLPHGYEGQGPEHSSARIERFLQL
Venom IIDQFISSGQAKWVRQSGLVMLQPHGLEGGPEHSSARMERFLQM

B

Cellular SHLEAVNPVLAGRVRKQN-----TRSVLGISIHGDAAFIFGQGV
Venom SHLEAVDPIVQKTRAEQFYRGDGEGKKVMSVLLHGDAAFVCGQGI

Cellular VAETLTLNIEGYKVGGIVHIVINNQVG
Venom VFETMHLSDLPDYTTHTGTHIVANNQIG

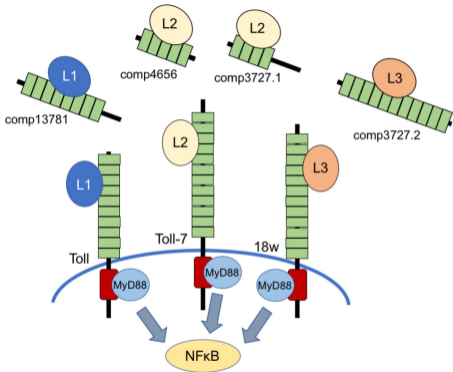


Table 1. Predicted venom paralogous genes are shown with their fly homolog and fly mutant phenotype.

Gene ID	Venom Paralog	Fly Homolog	Mutant Phenotype
Aldo-Keto Reductase	comp2501_c0_seq1	<i>CG10638</i>	Viable and Fertile
Aldolase	comp488_c0_seq2	<i>Ald</i>	Lethal
Annexin B9	comp829_c0_seq1	<i>AnxB9</i>	Partially Lethal
ATP Synthase Alpha	comp448_c0_seq1	<i>blw</i>	Lethal/Male Sterile
ATP Synthase Beta	comp393_c0_seq1	<i>ATPsynB</i>	Lethal/Female Sterile
Cyclophilin	comp148_c1_seq1	<i>Cyp1</i>	Lethal
Cytochrome c	comp346_c0_seq1	<i>Cyt-c-d</i>	Male Sterile
Myosin Light Chain	comp1347_c0_seq1	<i>Mlc2</i>	Lethal
Neprilysin	comp844_c0_seq1	<i>Nep2</i>	Partially Lethal
Oxoglutarate Dehydrogenase	comp1724_c0_seq2	<i>Nc73EF</i>	Partially Lethal

Table 2. Conserved domains showing enrichment in the venom protein sample.

Domain Name	Domain ID	Number	P Value
Leucine Rich Repeat (LRR8)	pfam13855	75	8.14E-15
Filamin Repeats	pfam00630	19	4.13E-05
Leucine Rich Repeat (LRR_RI)	cd00116	16	2.43E-03
DUF3584	pfam12128	10	1.26E-02
AAA Domain	pfam13514	9	1.85E-02
Mad Mitotic Checkpoint Protein	pfam05557	9	1.85E-02
Annexin	pfam00191	8	2.95E-02
Macoilin TM Protein	pfam09726	8	2.95E-02
Alpha Helical Coiled-Coil Rod Protein	pfam07111	7	4.17E-02
Autophagy Protein Apg6	pfam04111	7	4.17E-02
Calcium Binding Coiled-Coil Domain (CALCOCO1)	pfam07888	7	4.17E-02
Smc N Term	pfam02463	7	4.17E-02
Calponin Homology Domain	pfam00307	6	4.80E-02
Serpin	pfam00079	6	4.80E-02

Table 3. Minimized venom proteins and their *Drosophila* homologs listed by principal conserved domain, size and number of distinct and total conserved domain. *Drosophila* homologs with a known role in immunity are indicated.

Domain	Venom	Size (kDa)	Distinct CD	Total CD	Homolog	Size (kDa)	Distinct CD	Total CD	Immune Role
LRR	comp3727_c1_seq2	69.8	1	6	18 wheeler	154.8	3	13	[121,122]
LRR	comp6788_c0_seq1	40.9	1	5	artichoke	172.1	2	15	
EGF	comp248_c0_seq1	24.8	1	1	CG31999	102.1	2	12	
CUB	comp282_c0_seq1	15.3	1	1	CG42255	403.5	2	25	
LRR	comp1786_c0_seq1	50.3	1	4	CG7896	154.1	3	18	
LRR	comp2027_c0_seq2	35.5	1	1	kek6	93.4	3	4	[128,129]
LRR	comp2027_c0_seq22	9.4	1	1	kek6	93.4	3	4	[128,129]
LRR	comp4656_c0_seq11	21.2	1	2	tartan	81.9	2	7	
LRR	comp2027_c0_seq1	33.1	1	4	tartan	81.9	2	7	
LRR	comp13781_c0_seq1	90.1	1	8	Toll	124.6	3	11	[123,124]
LRR	comp3727_c1_seq1	89.0	1	4	Toll-7	160.9	3	14	[126,127]
LRR	comp4656_c0_seq2	38.7	1	5	Toll-7	160.9	3	14	[126,127]
LRR	comp6082_c0_seq1	32.2	1	2	CG1504	46.7	1	1	
LRR	comp2027_c0_seq10	22.4	1	1	CG32055	61.1	1	6	[125]
LRR	comp2027_c0_seq4	34.0	1	3	chaoptin	152.0	1	15	
LRR	comp2027_c0_seq6	18.0	1	2	Connectin	77.2	1	7	[129]
LIM	comp934_c0_seq1	8.3	1	1	Mlp84B	53.5	1	5	