Transcriptome Analysis Provides Insight into Venom Evolution in a Seed-Parasitic Wasp, *Megastigmus spermotrophus*

Amber R. Paulson, Cuong H. Le, Jamie C. Dickson, Jürgen Ehlting, Patrick von Aderkas and Steve J. Perlman

October 2016

The final publication is available at Wiley via:

https://doi.org/10.1111/imb.12247

Transcriptome Analysis Provides Insight into Venom Evolution in a Seed-Parasitic Wasp, *Megastigmus spermotrophus*

Amber R. Paulson[1], Cuong H. Le, Jamie C. Dickson, Jürgen Ehlting, Patrick von Aderkas[2] and Steve J. Perlman[3*]

Department of Biology, University of Victoria, Victoria, British Columbia, Canada.

[*] Integrated Microbial Biodiversity Program, Canadian Institute for Advanced Research, Toronto, Ontario, Canada.

[1] amber.rose.paulson@gmail.com and corresponding author, [2] pvonader@uvic.ca, [3] stevep@uvic.ca

Fax: Attention Dr. Steve Perlman, +1 250-721-7120

Submitting author postal address:

Amber Paulson (c/o Dr. Steve Perlman)
Department of Biology
University of Victoria
PO Box 3020,
Station CSC
Victoria, BC V8W 3N5

Key words: venom, phytophagy, seed-parasitism, Chalcidoidea, parasitoid, Hymenoptera, aspartylglucosaminidase

Running title: Putative venoms of *Megastigmus spermotrophus*

Abstract

One of the most striking host range transitions is the evolution of plant parasitism from animal parasitism. Parasitoid wasps that have secondarily evolved to attack plants (i.e., gall wasps and seed-feeders) demonstrate intimate associations with their hosts, yet the mechanism of plant-host manipulation is currently not known. There is, however, emerging evidence suggesting that ovipositional secretions play a role in plant manipulation. To investigate whether parasites have modified pre-existing adaptations to facilitate dramatic host shifts we aimed to characterize the expression of venom proteins in a plant parasite using a collection of parasitoid venom sequences as a guide. The transcriptome of a seed-feeding wasp, *Megastigmus spermotrophus*, was assembled *de novo* and three putative venoms were found to be highly expressed in adult females. One of these putative venoms, aspartylglucosaminidase, has been previously identified as a major venom component in two distantly-related parasitoid wasps (*Asobara tabida* and *Leptopilina heterotoma*) and may have originated via gene duplication within the Hymenoptera. Our study shows that *M. spermotrophus*, a specialized plant parasite, expresses putative venom transcripts that share homology to venoms identified in *Nasonia vitripennis* (both superfamily Chalcidoidea), which suggests that *M. spermotrophus* may have co-opted pre-existing machinery to develop as a plant parasite.

Introduction

Parasitism is perhaps the most successful and diverse strategy on the planet and parasites have evolved many incredibly sophisticated ways to subdue and manipulate their hosts. Within the class Insecta, an amazing diversity of parasitic lifestyles have evolved, with parasitoids notably being the most successful group of parasitic insects in terms of species diversity and host range. In parasitoids, the juvenile stage (i.e. larva) typically develops in or on an animal host, usually another insect, killing it, and developing into a free-living adult (Eggleton & Gaston, 1990; Eggleton & Belshaw, 1992). The parasitoid lifestyle has evolved independently in three major insect orders, beetles (Coleoptera), flies (Diptera), and perhaps most successfully, in wasps (Hymenoptera), where it evolved only once and yet has resulted in an explosive radiation of life history strategies and host range. Hosts include all insect orders and many other terrestrial invertebrates including snails, crabs and spiders (Godfray, 1994). In almost all parasitic

Hymenoptera, the adult female wasp lays an egg in or on the host, i.e. mothers locate and subdue hosts.

The most striking host range transition within higher parasitic Hymenoptera is the ability to parasitize plants. The goal of this paper is to begin addressing the question: How does a plant parasite evolve from an animal parasite? Plant parasitism, or endophytophagy, has evolved independently numerous times in this order, in the form of either gall-making or seed parasitism (Whitfield, 2003; Heraty *et al.*, 2011). The best known and most diverse plant-parasitic Hymenoptera are fig wasps (Agaonidae) and cynipid gall wasps (Cynipidae), although plant parasitism has been documented in many other groups, including the family Braconidae (Austin & Dangerfield, 1998) and several families of Chalcidoidea (Munro *et al.*, 2011). Work on the phylogenetic relationships among and within the major parasitoid lineages is still ongoing, making it difficult to understand the key transitions that have led to the evolution of plant parasitism from animal parasitism in Hymenoptera (Eggleton & Belshaw, 1992; Whitfield, 2003; Munro *et al.*, 2011). With the exception of one recent metatranscriptome investigation of gall induction by fig wasps (Martinson *et al.*, 2015) and a study on hormones produced by galling sawflies (Yamaguchi *et al.*, 2012), very little exploration about the mechanism of plant parasitism in Hymenoptera has been conducted.

We chose to examine the evolution of plant parasitism in *Megastigmus spermotrophus* Wachtl (Hymenoptera: Chalcidoidea: Torymidae), a well-studied, economically important pest of Douglas-fir, *Pseudotsuga menziesii* (Mirbel) Franco. *Megastigmus* is a diverse and speciose genus that includes both plant and animal parasites (Grissell, 1999), and therefore the evolution of obligate plant-parasitism in this group is relatively recent. *M. spermotrophus* is the best studied species in this genus and has been previously shown to exhibit a very sophisticated strategy of host manipulation. After the egg hatches the larva consumes the developing plant embryo, yet the host megagametophyte continues to accumulate storage products on which the larva feeds (von Aderkas et al. 2005a). Even when the eggs are laid earlier in developing ovules, the larva is able to redirect the development of unfertilized ovules that would normally abort (von Aderkas *et al.* 2005b). Thus *M. spermotrophus* is able to co-opt the conifer female reproductive tissue for its own reproductive success at the expense of the host, demonstrating a unique method of manipulating seed development (von Aderkas *et al.* 2005a; b). How *M.*

spermotrophus alters Douglas-fir seed development is not known, although data from hormone profiling suggested that the failure of the megagametophyte to abort in unpollinated infested treatments may be partially explained by changes in cytokinins (Chiwocha *et al.*, 2007). Cytokinins and other phytohormones have been shown to be involved the development of insect galls and green islands caused by leaf-mining insects (Mapes & Davies, 2001a; b; Giron *et al.*, 2007; Yamaguchi *et al.*, 2012).

We chose to focus our study on venoms, as these have been shown to be a crucial component of successful parasitism in Hymenoptera. In addition to laying an egg into their hosts, females also inject a diverse cocktail of compounds. Parasitoid venoms are known to disrupt host cells or tissues, enhance other virulence factors, induce paralysis, modify host metabolism and physiology, interfere with host development and/ or suppress the host immune response (Danneels *et al.*, 2010; Moreau, 2013; Moreau & Asgari, 2015; Mrinalini *et al.*, 2015). Several large-scale transcriptomic and/ or proteomic surveys have been recently performed (Danneels *et al.*, 2010; Vincent *et al.*, 2010; Zhu *et al.*, 2010; Colinet *et al.*, 2013; Dorémus *et al.*, 2013; Heavner *et al.*, 2013; Burke & Strand, 2014); however little is known about the composition of parasitoid venom from most parasitoid species. These studies have shown that parasitoid venoms are complex and diverse, consisting of many components, including small peptides, neurotoxins, amines and larger enzymes (Asgari & Rivers, 2011). In the last two decades there has been a surge in venom-based drug discovery programs (King, 2011). With rapid advances in next generation sequencing platforms we will likely see continued drug-bioprospecting of unstudied venomous lineages for novel drug compounds (Casewell *et al.*, 2013).

Given the importance and diverse functions of venoms within the Hymenoptera, it would be surprising if venoms were not involved in the manipulation of host plant tissues by endophytophagous wasps. In the case of *M. spermotrophus*, we hypothesize that venomous secretions may play a role in early host manipulation (i.e., the redirection of unfertilized ovules), potentially through interference of normal phytohormone pathways. At least some evidence exists to support the notion that gall-inducing wasps produce ovipositional secretions and that these secretions are associated with the induction of galls in sawflies (Tenthredinidae), fig-wasps (Agaonidae) and cynipid wasps (Cynipidae) (McCalla *et al.*, 1962; Price, 1992; Kjellberg *et al.*, 2005; Leggo & Shorthouse, 2006; Cox-Foster *et al.*, 2007; Martinson *et al.*, 2015). Furthermore,

a recent study on the morphological evolution of the venom apparatus from cynipoid wasps found that most phytophagous species have a larger venom apparatus than inquilines and parasitoids (Vårdal 2004, 2006); fig wasps also have large venom glands (Martinson *et al.*, 2014). However, the association of ovipositional secretions and gall induction by chalcid wasps has not been very well studied. An early investigation of the internal anatomy of a phytophagous chalcid from the genus *Harmolita* sp. revealed the presence of a well-developed poison apparatus, leading to speculation that secretions from the poison apparatus were injected during oviposition and that the fluid initiated and/or caused the gall to form (James 1926).

The focus of this study is to identify putative proteinaceous venom components that are highly expressed in female *M. spermotrophus*, which may play a role in early host manipulation of Douglas-fir ovules. To identify putative venoms of *M. spermotrophus*, we first used a comparative transcriptome approach. To this end, we identified potential candidate venom constituents based on sequence similarity to previously characterized Hymenoptera venoms in the *de novo* transcriptome of *M. spermotrophus*. Recently, *Nasonia vitripennis*, an ectoparasitoid of flesh fly pupae, became the first parasitoid and chalcid wasp to have its genome sequenced (Werren *et al.*, 2010). Supplemental to the genome, a recent study identified 79 constituents of *Nasonia* venom, obtained by a combination of bioinformatics and proteomics (de Graaf *et al.* 2010). Both *N. vitripennis* and *M. spermotrophus* belong to the superfamily Chalcidoidea. The availability of a sequenced genome combined with a diverse set of *N. vitripennis* venom protein sequences provided an excellent tool to investigate the possibility that *M. spermotrophus* may share homologous venom components. We used differential expression analysis, subsequently validated with qRT-PCR, to identify putative venom transcripts that were highly expressed in adult females, compared to adult males and larvae. Our work demonstrates that endophytophagous wasps express a number of transcripts with significant homology to *N. vitripennis* venoms, of which three putative venom transcripts were highly expressed in female *M. spermotrophus,* suggesting a potential role in early host manipulation of Douglas-fir ovules. These findings support the hypothesis that plant parasites may adapt mechanisms of host manipulation employed by their animal-parasite ancestors.

Results

Short read filtering and de novo assembly

Illumina sequencing of four *M. spermotrophus* whole insect cDNA libraries (larva, adult male, lab-reared adult female and wild adult female) generated 236,985,595 paired-end reads of 100 bp in length, equating to 47.40 giga-bases of total sequence. Fewer reads (9.2 Gbp) were sequenced from the wild adult female compared to the other samples (12.5 to 13.0 Gbp) . Quality filtering removed approximately 15.4 % of the reads prior to assembly, resulting in a mixed population of paired- and single-end reads (Table 1).

The transcriptome of *M. spermotrophus* was assembled *de novo* using multiple *k*mer values (Figure S1). Each of the individual *k*mer assemblies was combined using trans-ABySS, resulting in 1,361,656 assembled contigs. These assembled contigs were first clustered using the program CD-HIT-EST, which generated 296,711 clusters. A second clustering program, TIGR-TGICL, was used, resulting in 44,176 clusters and 149,236 singletons. Removal of all contigs less than or equal to 100 bp resulted in a final contig set of 143,306 transcripts (Table 2). The transcripts ranged in length from 101 (minimum contig length) up to 32,049 bp, with a N50 of 2,420 bp. The entire length of the transcriptome totalled 118,105,899 bp with an average contig length of 824 bp.

Annotation

From the transcriptome 1,639 contigs had significant similarity (E-value cut-off = 10^{-7}) to 41 of the 64 proteins in the *N. vitripennis* venom query dataset (Table 3, left column). In some cases, annotations representing proteins with other physiological functions from the NCBI non-redundant protein and/ or nucleotide databases had a smaller E-value than venom protein annotations and were not further considered. Consequently, there were 21 putative venom proteins, corresponding to 42 contigs, in the final *M. spermotrophus* annotation set (Table 3, right column; Supporting information Table S1).

Beyond the 42 contigs annotated as putative venoms, a total of 42,634 contigs (30 % of all transcripts) were assigned an annotation based on sequence similarity to entries present in the NCBI non-redundant protein database and nucleotide collection (including physiological

paralogs of putative *M. spermotrophus* venoms). Annotation of the *M. spermotrophus* transcriptome demonstrated redundancy, with many annotations (~ 72 %) being assigned to multiple contigs (average annotation assignment = 2.1, standard deviation = 3.5), resulting in 20,284 total non-redundant annotations (data not shown).

The majority of non-redundant annotations assigned from the nr protein database were from insects (77.3 %) (data not shown). The model parasitoid *N. vitripennis* had the greatest overall representation among annotations (47.3 %). Almost 15 % of annotations were based on closest matches to prokaryotes, and most of the bacterial annotations (71.7 %) represented sequences from a single genus of Betaproteobacteria, *Ralstonia*.

Transcript expression and differential analysis

The program RSEM was used to generate expression values by mapping the forward reads from all of the libraries onto the assembled contigs and calculating expected counts. In total 202,977,319 forward reads were processed, with 5,980,555 (3.0 %) read mapping failures. Expected count data were transformed using Conditional Quantile Normalization, which resulted in the standardization of the distribution of counts for all four libraries. The NOISeq-sim algorithm identified 404 transcripts that are likely differentially expressed in *M. spermotrophus* females compared to larvae and males (Supplemental data, Figure S2). Analysis focused on the 243 transcripts that are more highly expressed in females compared to males and larvae, since putative venom transcripts are more likely present in this specific set of transcripts.

Three of the proteins that were annotated as putative venoms were identified as likely being highly expressed in female *M. spermotrophus*: Gram-negative bacteria binding 1-2 precursor (GNB), venom protein R precursor (VPR) and aspartylglucosaminidase precursor (AGA-V) (Table 3). Three contigs were annotated as AGA-V, of which two were almost exclusively expressed in females compared to larvae and males (Figure 1A). These three contigs are identical in a large section in the C-terminus of the protein (224 amino acids). In contrast to this female-biased expression, two transcripts that were annotated as lysosomal aspartylglucosaminidase (N-(4)-(beta-N-acetylglucosaminyl)-L-asparaginase , AGA-L), a putative paralog of AGA-V, were not found to be more highly expressed in females than in males or larvae (Figure 1B). There were six contigs annotated as GNB, of which four were significantly differentially expressed,

196 with higher expression in females compared to males and larvae (Figure 1C). Three additional
197 transcripts were annotated as beta-1,3-glucan-binding protein (beta-GBP), another pattern
198 recognition protein and a putative paralog of GNB; all three putative beta-GBP transcripts were
199 also highly differentially expressed in females compared to larvae and males (Figure 1D). Two
200 contigs were identified as VPR, of which, one was highly expressed in females (Figure 1E).
201 There were no putative paralogs of VPR.

202 *Validation with Quantitative Real-Time Polymerase Chain Reaction*

203 We conducted qRT-PCR to validate differential expression of candidate venom genes AGA and
204 VPR in females, using primers that were designed to target all redundant putative venom contigs.
205 Normalized $\log_2$ transformed expression of AGA-V was significantly higher in both lab-reared
206 and wild female samples compared to larvae (p-value = 0.02) (Figure 2). In contrast, the non-
207 venomous paralog AGA-L was expressed in much lower levels in female samples compared to
208 larvae, although lab-reared females and larvae were significantly different (p-value = 0.02)
209 (Figure 2). We did not find evidence of differential expression of VPR using qRT-PCR (data not
210 shown).

211 *Phylogenetic Analysis of AGA*

212 A protein phylogeny of aspartylglucosaminidase (AGA) was re-constructed using Bayesian
213 methods from a wide sample of sequences from insect genomes, with a focus on Hymenoptera
214 (Figure 3). Additional AGA sequences that had been confirmed to be venoms in the parasitoids
215 *Leptopilina heterotoma* and *Asobara tabida* were also included, as well as sequences from
216 transcriptomes from a range of hymenopteran lineages whose genomes have not yet been
217 sequenced, such as sawflies. AGA appears to have been duplicated in the Hymenoptera, with a
218 number of independent duplications in ants and in chalcidoid wasps and their relatives (i.e. the
219 lineage that includes *Nasonia*, *Megastigmus*, *Ceratosolen* and *Pegoscapus* fig wasps, and
220 *Leptopilina*, which is a cynipoid wasp). The putative non-venomous *Megastigmus* paralog AGA-
221 L is found in a tight cluster with non-venomous AGA from *Nasonia* and an AGA from
222 *Ceratosolen* fig wasps, which is presumably non-venomous as well. On the other hand,
223 *Megastigmus* AGA-V lies on a longer branch, and does not form a tight cluster with *Nasonia* and

fig wasp AGA-V. Other AGA sequences that have been confirmed as venoms, such as in *Leptopilina heterotoma* and *Asobara tabida,* also lie on long branches.

Discussion:

Sequencing the transcriptome of *M. spermotrophus* recovered a sizeable fraction (21 out of 64) of the venoms used by the closely related parasitoid *N. vitripennis*. Differential expression analysis of the *M. spermotrophus* transcriptome revealed three interesting candidate genes that are highly expressed in females, with AGA-V being an especially promising candidate venom protein because it has also been identified as a venom protein in several divergent parasitic Hymenoptera species, including three parasitoids and a fig wasp (Moreau *et al.* 2004; Colinet *et al.* 2013; Martinson *et al.* 2015).

De novo assembly and post-assembly clustering of the *M. spermotrophus* transcriptome generated 143,306 contigs. The number of contigs is comparable to other insect transcriptomes constructed exclusively from Illumina generated sequences, such as that of soybean aphid (253,603 contigs) (Liu *et al.*, 2012), *Anopheles funestus* (46,987 contigs) (Crawford *et al.*, 2010), oriental fruit fly (484,628 contigs) (Shen *et al.*, 2011) and salt marsh beetle (65,766 contigs) (van Belleghem *et al.* 2012). Post-sequencing clustering and removal of singletons resulted in a relatively large average contig length of over 800 bp. Nearly half (47.3 %) of non-redundant annotations assigned to the *M. spermotrophus* transcriptome were from the model parasitoid and close relative *N. vitripennis* and 77.3% of all non-redundant annotations were of insect origin. Interestingly, many transcripts (10.6 % of non-redundant annotations) were assigned from the bacterial genus *Ralstonia* (Betaproteobacteria), which corroborated the recent findings that this bacterium is pervasively associated with different life stages of *M. spermotrophus* (Paulson *et al.*, 2014). Interestingly, the four most highly expressed *Ralstonia*-attributed contigs are related to mobile genetic elements, including: transposase IS66 (YP_001899056.1), ISPsy11 transposase OrfB (ZP_10987507.1), resolvase domain protein (ZP_10982688.1) and putative cointegrate resolution protein T (ZP_07678195.1).

Twenty-one *N. vitripennis* venom protein annotations were assigned to 42 contigs of the *M. spermotrophus* transcriptome, including venom proteins from all categories listed by de Graaf *et al.* (2010). This is maybe not so surprising since *M. spermotrophus* and *N. vitripennis* belong to

the same superfamily, suggesting that at least a portion of the *N. vitripennis* venom protein repertoire may have been retained by *M. spermotrophus* and perhaps modified to enable an endophytophagous lifestyle over evolutionary timescales. In some cases, multiple contigs were annotated as the same putative venom, which may be attributed to either assembly errors, splice variants, genomic DNA contamination, gene duplications or allelic variation among individuals within the population sampled.

Among the transcripts that were highly expressed in female libraries were a number of expected genes that are known to be associated with oogenesis, such as vitellogenin and vitellogenin-like protein (Guidugli *et al.*, 2005), vitellogenin receptor (Schonbaum *et al.*, 2000), nanos (Forbes & Lehmann, 1998) and the maternal effect protein oskar (Lehmann & Nüsslein-Volhard, 1986) (data not shown). Also, several transcripts associated with odor perception, such as chemosensory protein CSP-1 (Pelosi *et al.*, 2006) and putative odorant binding protein 70 (Vogt *et al.*, 1999) were highly expressed in females, which suggests that female *M. spermotrophus* may utilize chemical cues to locate susceptible host trees. Next we focused on candidate venom transcripts that were differentially expressed in females as a means to identify potential mechanisms of early host manipulation as we hypothesize venoms are likely injected into the host by the adult female during oviposition. Three candidate venom transcripts were identified to be highly expressed in females compared to larvae and males using differential expression analysis: gram-negative bacteria binding 1-2 precursor (GNB), venom protein R precursor (VPR) and aspartylglucosaminidase precursor (AGA-V).

The putative venom AGA-V is the most promising candidate for host manipulation identified from the *M. spermotrophus* transcriptome. Two contigs annotated as AGA-V were identified as highly differentially expressed in adult females and a third contig followed a similar pattern. Using qRT-PCR, we validated the strong and significant increase in AGA-V expression in adult female *M. spermotrophus* compared to larvae. Adding to the possibility of AGA-V being an important venom in *M. spermotrophus*, aspartylglucosaminidase has also been identified in a recently published fig wasp transcriptome study and in the venom of at least three other parasitoid species, in addition to *N. vitripennis.* An AGA-V homolog from the fig wasp *Pegoscapus hoffmeyeri* was identified in pollinated fig flowers (Martinson *et al.*, 2015). In fact, eight other Hymenoptera transcripts sharing homology with *N. vitripennis* venoms, including the

lysosomal enzyme acid phosphatase, were identified in fig flowers. AGA-V is also a major venom constituent of the *Drosophila* parasitoids *Asobara tabida* (Braconidae) (Moreau *et al.*, 2004; Vinchon *et al.*, 2010) and *Leptopilina heterotoma* (Figitidae) (Colinet *et al.*, 2013). In contrast to *L. heterotoma*, AGA-V was not found to be a major component in the venom of the congener *L. boulardi* (Colinet *et al.*, 2013), demonstrating the highly dynamic nature of Hymenoptera venom. Finally, we also found a significant match to *M. spermotrophus* AGA-V in an unpublished transcriptome study of teratocytes from *Cotesia plutellae* (Braconidae) (accession #GAKG01023507.1).

From bacteria to humans, aspartylglucosaminidase, otherwise known as N-(4)-(beta-N-acetylglucosaminyl)-L-asparaginase (AGA) is an essential lysosomal enzyme, involved in the digestion of glycoproteins (Tarentino *et al.*, 1995; Tenhunen *et al.*, 1995; Liu *et al.*, 1996), which acts on glycosylated asparagines by hydrolyzing the ß-N-glycosidic linkage between an asparagine residue and an N-acetylglucosamine moiety (Makino *et al.*, 1966). Here we provide evidence suggesting that AGA-V evolved through the duplication of its non-venomous lysosomal paralog, AGA-L. Many venom toxins evolve via gene duplication whereby a gene encoding a normal 'physiological' protein with an important bioactivity or regulatory function is duplicated and the duplicate copy becomes selectively expressed in the venom gland (Kordiš & Gubenšek, 2000; Fry *et al.*, 2003). Indeed, lysosomal enzymes are thought to be commonly recruited into hymenopteran venoms, with examples such as diverse hydrolases (Vinchon *et al.*, 2010) and acid phosphatase (Dani *et al.*, 2005; Zhu *et al.*, 2008, 2010). Such enzymes might play a role in catalyzing the release of nutrients from host hemolymph (Dani *et al.*, 2005) or serve a specific purpose in affecting the host's physiology (Zhu *et al.*, 2008). A phylogenetic analysis of a diverse set of insect AGA sequences revealed that AGA has been duplicated a number of times within Hymenoptera, including at least once in chalcidoid wasps (i.e. *Megastigmus*, *Nasonia*, and fig wasps) and their relatives (e.g. *Leptopilina*); AGA-L and AGA-V are directly adjacent to each other in the *N. vitripennis* genome (Munoz-Torres *et al.*, 2011). At this point, however, it is difficult to accurately reconstruct the number of times that AGA has been duplicated in chalcidoids and their relatives, as there are few available sequences, and often only one of the duplicates. Reconstruction is also made challenging by the fact that many of the AGA sequences that are suspected to have a venomous function (e.g. *Leptopilina*, *Megastigmus*) appear to lie on

long branches, suggesting that they are evolving rapidly, although more sequences, including more non-venomous paralogs, are required to examine this in more detail. It would also be interesting to examine AGA evolution and function in ants, as this gene appears to have been duplicated there as well.

Two other candidate venom proteins were identified as differentially expressed in *M. spermotrophus*, GNB and VPR. Both were identified in a proteomic study of *N. vitripennis* venom extract (de Graaf *et al.* 2010). VPR has no similarity to any known protein, so it is difficult to make predictions with respect to its function. GNB belongs to a family of recognition proteins called the gram-negative bacteria-binding proteins (GNBPs), some of which have a strong affinity for lipopolysaccharides (Kim et al. 2000, Ochiai 2000). Prior to the de Graaf et al. (2010) study, GNBPs were not known to be associated with insect venom. It is possible that GNB has an intrinsic immunological function within *M. spermotrophus*, rather than being secreted as a venom. Alternatively, GNB may have a role in reducing bacterial or fungal invasion of the ovule following oviposition, protecting the egg and/ or developing larva from microbial pathogens (Dani *et al.*, 2003; Moreau, 2013).

While our approach was successful in targeting potential venom constituents from adult females, it is probably a major underestimate of the entire arsenal of *M. spermotrophus* genes that contribute to host manipulation. For example, early larval instars are also likely very important in maintaining continued manipulation of ovule development and redirection of nutrients during the feeding stage. Larval secretions during feeding are known to be critical in gall formation in cynipid gall wasps (Leggo & Shorthouse, 2006). As our *M. spermotrophus* transcriptome is missing this key development stage, we were not able to identify candidate proteins that may be secreted by the larvae. Also, an unbiased screen of the venom gland itself, using proteomic or transcriptomic approaches, could provide more detailed insight into the venom repertoire of *M. spermotrophus.* It is interesting to note that dissections of *M. spermotrophus* revealed the presence of a noticeable venom gland (A. Paulson, personal observation) and that gall-inducing cynipid, agaonid and chalcid wasps are also known to have well developed venom glands (James, 1926; Vårdal, 2004, 2006; Martinson *et al.*, 2014). Furthermore, in focusing on only those putative venom proteins with very high expression in females compared to males and larvae, we may have underestimated potential venom proteins that have similar expression

profiles in all libraries in *M. spermotrophus*. A tiling expression microarray of *N. vitripennis* comparing female and male reproductive tissue found that expression levels for some venom proteins was only subtly higher in the reproductive tract of females compared to male testes (de Graaf *et al.*, 2010; Werren *et al.*, 2010).

Through the application of transcriptomic approaches we were able to determine that endophytophagous wasps share many homologous venoms with parasitoids, which suggests that the evolution of plant endoparasitism in *Megastigmus* may not have completely relied on wholesale innovations; sequencing the *Megastigmus* genome would help resolve this more clearly. On this note, a recent analysis of the genome of the fig wasp *Ceratosolen solmsi* did not identify any unique genes or gene family expansions related to host manipulation compared to *N. vitripennis* (Xiao *et al.*, 2013), which also supports the idea that endophytophagous hymenopteran lineages have likely adapted the parasitoid venom machinery for manipulating plants.

Experimental Procedures:

RNA extraction

Adult *M. spermotrophus* males and females were collected upon emergence and larvae were extracted from heavily infested seed from the Mt. Newton Seed Orchard, located in Saanichton, BC (48°35'54.00"N, 123°25'56.87"W). Wild females were collected from trees located on the University of Victoria campus in Victoria, BC (48°27'42.90"N, 123°18'37.50"W). All insect samples were flash-frozen in liquid nitrogen and then stored at -80 °C. Approximately 10-20 individuals were placed into 2 ml Micro tubes (Sarstedt) with one volume buffer RLT (Qiagen), 1/100 volume beta-mercaptoethanol and three 3.5mm dia. glass beads (BioSpec Products). Samples were homogenized using the Mini-Beadbeater (BioSpec Products) at half-speed for 90 seconds. The homogenate was centrifuged at 1,300 x g for 3 minutes. Total RNA was extracted using RNeasy (Qiagen), followed by on-column DNase digestion and RNA cleanup, using the manufacture's guidelines. Next, the RNA extract was purified using an isopropanol precipitation followed by a 100 % ethanol wash and then re-suspended in RNase-free water. The RNA extract was separated on a 1 % agarose gel stained with SYBR Safe (Invitrogen) and visualized under UV light. The RNA quality and quantity was determined using a Nanodrop 2000 instrument

(Thermo Scientific) and RNA quality was further analyzed using an Experion Electrophoresis Station (Bio-Rad).

Complementary DNA library construction with oligo(dT) primers, library fragmentation, size exclusion purifications (target average sequence length of 300 bp) and sequencing on the Illumina sequencing platform (HISeq 2000) were conducted by the BC Cancer Agency Genome Sciences Centre, Vancouver, Canada.

Short read filtering and de novo assembly

Short reads were first quality filtered with Trimmomatic (v0.22) (Bolger *et al.*, 2014) with the following parameters: minimum leading quality of three, minimum trailing quality of 20, minimum read length of 36 and sliding window of four bases with a minimum quality of 20. Filtered reads were then assessed using FASTQC (v0.10.1) to verify quality improvements (Andrews, 2010). The short reads were assembled *de novo* using the trans-ABySS (v1.3.2) pipeline using k-mer values of 30, 35, and even values from 52-96 (Robertson *et al.*, 2010). The assembly was further clustered using CD-HIT-EST with the default sequence identity threshold of 0.95 (v4.6) (Li & Godzik, 2006). Additional clustering was performed by using TIGR-TGICL (Pertea *et al.*, 2003) with the Cap3 specific overlap percent identity cut-off set to 98. Only contigs larger than 100 bases were used in subsequent analysis.

Annotation

A query of sixty-four *N. vitripennis* protein sequences including proteases/peptidases, protease inhibitors, carbohydrate metabolism, DNA metabolism, glutathione metabolism, esterases, recognition/binding, others and unknowns was obtained from GenBank (Supporting information, Table S2). The venom proteins included in the query were originally generated by de Graaf et al. (2010) using both bioinformatic and proteomic approaches. In the bioinformatic approach a query of 383 protein sequences from previously known adult hymenopteran venom proteins was used to identify putative venom protein homologs from the *N. vitripennis* genome using BLASTp. In the proteomic approach crude *N. vitripennis* venom was analyzed using two methods of two-dimensional liquid chromatography-mass spectrometry. In order to identify putative venom transcript homologs, the *N. vitripennis* venom protein query was compared to the

M. spermotrophus transcriptome using tBLASTn, with an E-value cut-off of 10^{-7}.

Additionally, the *M. spermotrophus* transcriptome contig set was annotated with BLASTX (v2.2.27+) against the NCBI non-redundant (nr) database with an E-value cut-off of 10^{-5}. Any contigs without BLASTx hits were then annotated with BLASTn using the NCBI nucleotide database, with the same E-value cut-off.

Differential expression

Transcript expression was quantified using the RSEM software package (v1.2.0) (Li & Dewey, 2011), aligning forward reads only and providing a mean fragment length of 300 bp. As the mean fragment length was set to 300 bp, expression values were only calculated for contigs of length 300 bp or longer. Expected count values were normalized using the conditional quantile normalization (CQN) R package (v1.7.0) (Hansen *et al.*, 2012). Differential expression analysis was implemented using the non-parametric statistical analysis package NOISeq (v2.0.0) (Tarazona *et al.*, 2011). The NOISeq-sim feature was utilized to simulate technical replicates with the following parameters: size of simulated samples equal to twenty percent of sequencing depth, five simulation replicates and allowance of two percent variability. Differential expression probability was increased from 0.8 to 0.9 to account for the lack of technical replicates. Bioinformatics packages were implemented using R (v3.0.1) in RStudio (v0.97.551).

Validation with Quantitative Real-Time Polymerase Chain Reaction

Quantitative real-time polymerase chain reaction (qRT-PCR) was used to validate the expression of putative venom transcripts in adult female *M. spermotrophus*. RNA was extracted from whole body females (lab-reared and wild) and final-instar larvae (six biological replicates each) using 450 µl of TRIzol per sample according to the manufacturer's guidelines (Invitrogen) in a Mini-Beadbeater (Biospec Products). Reverse transcription was completed using Superscript III according to the manufacturer's protocol (Invitrogen) and with oligo(dT) primers (Integrated DNA technologies). The qRT-PCR reactions were performed in Bio-Rad CFX96 on a C1000 thermocycling platform with EvaGreen (Biotium Inc.) and HotStart DNA polymerase (Applied Biological Materials). Primers were designed using Primer-BLAST (Ye *et al.*, 2012) (Supporting information, Table S3). Amplicon sizes of 70 to 150 bp were selected with an

optimal primer annealing temperature of 60 ºC. For each primer our transcriptome data (with our targets removed) was used for specificity testing and only primers that would not generate products on any target in the database were used. Each reaction was completed in triplicate, with customized conditions to optimize PCR efficiency for each of the products.

Target expression was normalized against the expression of the predicted 60S ribosomal gene L13a (contig 178886). PCR efficiency-corrected relative normalized expression for each target were calculated (Pfaffl, 2001) and data were $\log_2$ transformed. Statistical analysis was performed using the Mann-Whitney U-test with the Bonferroni multiple-test correction using R (v3.1.1) in RStudio (v0.98.1049).

Aspartylglucosaminidase phylogeny

The phylogenetic relationships of AGA-V protein and its paralog AGA-L were reconstructed using a wide range of sequences that were collected from GenBank and the Hymenoptera Genome Database (Munoz-Torres *et al.*, 2011) using BLAST. AGA sequence from *Ceratosolen solmsi marchali* was provided y E.O. Martinson (personal communication). Contigs 4602 and 8722 were chosen from the *M. spermotrophus* transcriptome, as these were the longest and best aligning venomous and physiological AGA contigs, with 315 and 290 amino acids in the final alignment, respectively. Sequences were aligned in Geneious using MAFFT (Katoh & Standley, 2013), with the E-INS-I alignment algorithm, a BLOSUM 62 matrix, and a gap opening penalty of 1.53. Phylogenetic analysis was performed using Bayesian methods, in MrBayes v 3.2.5 (Ronquist *et al.*, 2012), with a WAG model of amino acid substitution and default settings.

Acknowledgements

We would like to thank Dave Koletelo at the Surrey Seed Centre, Don Piggott at Yellow Point Propagation Ltd, and orchard managers Tim Crowder (Mt. Newton Seed Orchard) and the late Tim Lee (Vernon Seed Orchard) for providing infested seed. We would especially like to thank Belaid Moa for assistance with transcriptome assembly and using the HPC on Westgrid / ComputeCanada. Finn Hamilton helped with the phylogenetic analysis. Jean-Noël Candau provided guidance on rearing and husbandry of *Megastigmus*. This work was funded by the Strategic Project Partnership Grants Program of NSERC and the Agence Nationale de Recherche

- Programme Blanc International.

References:

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Asgari, S. & Rivers, D.B. 2011. Venom proteins from endoparasitoid wasps and their role in host-parasite interactions. *Annu. Rev. Entomol.* **56**: 313–335.

Austin, A.D. & Dangerfield, P.C. 1998. Biology of *Mesostoa kerri* (Insecta: Hymenoptera: Braconidae: Mesostoinae), an endemic Australian wasp that causes stem galls on *Banksia marginata*. *Aust. J. Bot.* **46**: 559–569.

Bolger, A.M., Lohse, M. & Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Burke, G.R. & Strand, M.R. 2014. Systematic analysis of a wasp parasitism arsenal. *Mol. Ecol.* **23**: 890–901.

Casewell, N.R., Wüster, W., Vonk, F.J., Harrison, R.A. & Fry, B.G. 2013. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol. Evol.* **28**: 1–11.

Chiwocha, S., Rouault, G., Abrams, S. & von Aderkas, P. 2007. Parasitism of seed of Douglas fir (*Pseudotsuga menziesii*) by the seed chalcid, *Megastigmus spermotrophus*, and its influence on seed hormone physiology. *Sex. Plant Reprod.* **20**: 19–25.

Colinet, D., Deleury, E., Anselme, C., Cazes, D., Poulain, J., Azema-Dossat, C., *et al.* 2013. 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.

Cox-Foster, A.D.L., Conlan, S., Holmes, E.C., Palacios, G., Jay, D., Moran, N.A., *et al.* 2007. A metagenomic survey of in honey bee colony microbes disorder collapse. *Science* **318**: 283–286.

Crawford, J.E., Guelbeogo, W.M., Sanou, A., Traoré, A., Vernick, K.D., Sagnon, N., *et al.* 2010. De novo transcriptome sequencing in Anopheles funestus using Illumina RNA-Seq technology. *PLoS One* **5**: e14202.

Dani, M.P., Edwards, J.P. & Richards, E.H. 2005. Hydrolase activity in the venom of the pupal endoparasitic wasp, *Pimpla hypochondriaca*. *Comp. Biochem. Physiol.* **141B**: 373–381.

Dani, M.P., Richards, E.H., Isaac, R.E. & Edwards, J.P. 2003. Antibacterial and proteolytic activity in venom from the endoparasitic wasp *Pimpla hypochondriaca* (Hymenoptera: Ichneumonidae). *J. Insect Physiol.* **49**: 945–954.

Danneels, E.L., Rivers, D.B. & de Graaf, D.C. 2010. Venom proteins of the parasitoid wasp *Nasonia vitripennis*: recent discovery of an untapped pharmacopee. *Toxins* **2**: 494–516.

de Graaf, D.C., Aerts, M., Brunain, M., Desjardins, C.A., Jacobs, F.J., Werren, J.H., *et al.* 2010. Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from

493 bioinformatic and proteomic studies. *Insect Mol. Biol.* **19**: 11–26.

494 Dorémus, T., Urbach, S., Jouan, V., Cousserans, F., Ravallec, M., Demettre, E., *et al.* 2013.
495 Venom gland extract is not required for successful parasitism in the polydnavirus-associated
496 endoparasitoid *Hyposoter didymator* (Hym. Ichneumonidae) despite the presence of
497 numerous novel and conserved venom proteins. *Insect Biochem. Mol. Biol.* **43**: 292–307.

498 Eggleton, P. & Belshaw, R. 1992. Insect parasitoids: An evolutionary overview. *Philos. Trans.*
499 *Biol. Sci.* **337**: 1–20.

500 Eggleton, P. & Gaston, K.J. 1990. "Parasitoid" species and assemblages: Convenient definitions
501 or misleading compromises? *Oikos* **59**: 417–421.

502 Forbes, A. & Lehmann, R. 1998. Nanos and Pumilio have critical roles in the development and
503 function of *Drosophila* germline stem cells. *Development* **125**: 679–690.

504 Fry, B.G., Wüster, W., Kini, R.M., Brusic, V., Khan, A., Venkataraman, D., *et al.* 2003.
505 Molecular evolution and phylogeny of elapid snake venom three-finger toxins. *J. Mol. Evol.*
506 **57**: 110–129.

507 Giron, D., Kaiser, W., Imbault, N. & Casas, J. 2007. Cytokinin-mediated leaf manipulation by a
508 leafminer caterpillar. *Biol. Lett.* **3**: 340–343.

509 Godfray, H.C.J. 1994. *Parasitoids: behavioural and evolutionary ecology*. Princeton University
510 Press, Princeton, N.J.

511 Grissell, E.E. 1999. An annotated catalog of world Megastigminae (Hymenoptera: Chalcidoidea:
512 Torymidae). *Contrib. Am. Entomol. Inst.* **31**: 1–92.

513 Guidugli, K.R., Piulachs, M.-D., Bellés, X., Lourenço, A.P. & Simões, Z.L.P. 2005. Vitellogenin
514 expression in queen ovaries and in larvae of both sexes of *Apis mellifera*. *Arch. Insect*
515 *Biochem. Physiol.* **59**: 211–218.

516 Hansen, K.D., Irizarry, R.A. & Wu, Z. 2012. Removing technical variability in RNA-seq data
517 using conditional quantile normalization. *Biostatistics* **13**: 204–216.

518 Heavner, M.E., Gueguen, G., Rajwani, R., Pagan, P.E., Small, C. & Govind, S. 2013. Partial
519 venom gland transcriptome of a *Drosophila* parasitoid wasp, *Leptopilina heterotoma*,
520 reveals novel and shared bioactive profiles with stinging Hymenoptera. *Gene* **526**: 195–204.

521 Heraty, J., Ronquist, F., Carpenter, J.M., Hawks, D., Schulmeister, S., Dowling, A.P., *et al.*
522 2011. Evolution of the hymenopteran megaradiation. *Mol. Phylogenet. Evol.* **60**: 73–88.

523 James, H.C. 1926. The anatomy of a British phytophagous Chalcidoid of the genus *Harmolita*
524 (Isosoma). *Proc. Zool. Soc. London* **96**: 75–182.

525 Katoh, K. & Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7:
526 Improvements in performance and usability. *Mol. Biol. Evol.* **30**: 772–780.

527 Kim, Y.S., Ryu, J.H., Han, S.J., Choi, K.H., Nam, K.B., Jang, I.H., *et al.* 2000. Gram-negative
528 bacteria-binding protein, a pattern recognition receptor for lipopolysaccharide and beta-1,3-
529 glucan that mediates the signaling for the induction of innate immune genes in *Drosophila*
530 *melanogaster* cells. *J. Biol. Chem.* **275**: 32721–32727.

King, G.F. 2011. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert Opin. Biol. Ther.* **11**: 1469–1484.

Kjellberg, F., Jousselin, E., Hossaert-McKey, M. & Rasplus, J.-Y. 2005. Biology, ecology, and evolution of fig-pollinating wasps (Chalcidoidea, Agaonidae). In: *Biology, Ecology, and Evolution of Gall-inducing Arthropods, Volume 2* (A. Raman et al., eds), pp. 539–572. Science Publishers, Inc., Enfield, USA.

Kordiš, D. & Gubenšek, F. 2000. Adaptive evolution of animal toxin multigene families. *Gene* **261**: 43–52.

Leggo, J.J. & Shorthouse, J.D. 2006. Development of stem galls induced by *Diplolepis triforma* (Hymenoptera: Cynipidae) on *Rosa acicularis* (Rosaceae). *Can. Entomol.* **138**: 661–680.

Lehmann, R. & Nüsslein-Volhard, C. 1986. Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in *Drosophila*. *Cell* **47**: 141–152.

Li, B. & Dewey, C.N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.

Li, W. & Godzik, A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.

Liu, S., Chougule, N.P., Vijayendran, D. & Bonning, B.C. 2012. Deep sequencing of the transcriptomes of soybean aphid and associated endosymbionts. *PLoS One* **7**: e45161.

Liu, Y., Dunn, G.S. & Aronson, N.N. 1996. Purification, biochemistry and molecular cloning of an insect glycosylasparaginase from *Spodoptera frugiperda*. *Glycobiology* **6**: 527–536.

Makino, M., Toshimasa, K. & Ikuo, Y. 1966. Enzymatic cleavage of glycopeptides. *Biochem. Biophys. Res. Commun.* **24**: 961–966.

Mapes, C.C. & Davies, P.J. 2001a. Cytokinins in the ball gall of *Solidago altissima* and in the gall forming larvae of *Eurosta solidaginis*. *New Phytol.* **151**: 203–212.

Mapes, C.C. & Davies, P.J. 2001b. Indole-3-acetic acid and ball gall development on *Solidago altissima*. *New Phytol.* **151**: 195–202.

Martinson, E.O., Hackett, J.D., Machado, C.A. & Arnold, A.E. 2015. Metatranscriptome analysis of fig flowers provides insights into potential mechanisms for mutualism stability and gall induction. *PLoS One* **10**: e0130745.

Martinson, E.O., Jandér, K.C., Peng, Y.Q., Chen, H.H., Machado, C.A., Arnold, A.E., *et al.* 2014. Relative investment in egg load and poison sac in fig wasps: Implications for physiological mechanisms underlying seed and wasp production in figs. *Acta Oecologica* **57**: 58–66.

McCalla, D.R., Genthe, M.K. & Hovanitz, W. 1962. Chemical nature of an insect gall growth-factor. *Plant Physiol.* **37**: 98–103.

Moreau, S.J.M. 2013. "It stings a bit but it cleans well": Venoms of Hymenoptera and their antimicrobial potential. *J. Insect Physiol.* **59**: 186–204.

569 Moreau, S.J.M. & Asgari, S. 2015. Venom proteins from parasitoid wasps and their biological
570 functions. *Toxins* **7**: 2385–2412.

571 Moreau, S.J.M., Cherqui, A., Doury, G., Dubois, F., Fourdrain, Y., Sabatier, L., *et al.* 2004.
572 Identification of an aspartylglucosaminidase-like protein in the venom of the parasitic wasp
573 *Asobara tabida* (Hymenoptera: Braconidae). *Insect Biochem. Mol. Biol.* **34**: 485–492.

574 Mrinalini, Siebert, A.L., Wright, J., Martinson, E., Wheeler, D. & Werren, J.H. 2015. Parasitoid
575 venom induces metabolic cascades in fly hosts. *Metabolomics* **11**: 350–366.

576 Munoz-Torres, M.C., Reese, J.T., Childers, C.P., Bennett, A.K., Sundaram, J.P., Childs, K.L., *et*
577 *al.* 2011. Hymenoptera genome database: Integrated community resources for insect species
578 of the order Hymenoptera. *Nucleic Acids Res.* **39**: D658–D662.

579 Munro, J.B., Heraty, J.M., Burks, R.A., Hawks, D., Mottern, J., Cruaud, A., *et al.* 2011. A
580 molecular phylogeny of the Chalcidoidea (Hymenoptera). *PLoS One* **6**: e27023.

581 Ochiai, M. 2000. A pattern-recognition protein for ß-1,3-glucan: Rhe binding domain and the
582 cDNA cloning of ß-1,3-glucan recognition protein from the silkworm, *Bombyx mori*. *J.*
583 *Biol. Chem.* **275**: 4995–5002.

584 Paulson, A.R., von Aderkas, P. & Perlman, S.J. 2014. Bacterial associates of seed-parasitic
585 wasps (Hymenoptera: *Megastigmus*). *BMC Microbiol.* **14**: 224.

586 Pelosi, P., Zhou, J.-J., Ban, L.P. & Calvello, M. 2006. Soluble proteins in insect chemical
587 communication. *Cell. Mol. Life Sci.* **63**: 1658–1676.

588 Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., *et al.* 2003. TIGR
589 Gene Indices clustering tools (TGICL): A software system for fast clustering of large EST
590 datasets. *Bioinformatics* **19**: 651–652.

591 Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR.
592 *Nucleic Acids Res.* **29**: 45e–45.

593 Price, P.W. 1992. Evolution and Ecology of Gall-inducing Sawflies. In: *Biology of Insect-*
594 *induced Galls* (J. D. Shorthouse & O. Rohfritsch, eds), pp. 208–224. Oxford University
595 Press, New York, USA.

596 Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., *et al.* 2010. *De novo*
597 assembly and analysis of RNA-seq data. *Nat. Methods* **7**: 909–912.

598 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., *et al.* 2012.
599 Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large
600 model space. *Syst. Biol.* **61**: 539–542.

601 Schonbaum, C.P., Perrino, J.J. & Mahowald, A.P. 2000. Regulation of the vitellogenin receptor
602 during *Drosophila melanogaster* oogenesis. *Mol. Biol. Cell* **11**: 511–521.

603 Shen, G.-M., Dou, W., Niu, J.-Z., Jiang, H.-B., Yang, W.-J., Jia, F.-X., *et al.* 2011.
604 Transcriptome analysis of the oriental fruit fly (*Bactrocera dorsalis*). *PLoS One* **6**: e29127.

605 Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. 2011. Differential
606 expression in RNA-seq: A matter of depth. *Genome Res.* **21**: 2213–2223.

Tarentino, A.L., Quinones, G., Hauer, C.R., Changchien, L.M. & Plummer, T.H. 1995. Molecular cloning and sequence analysis of *Flavobacterium meningosepticum* glycosylasparaginase: a single gene encodes the alpha and beta subunits. *Arch. Biochem. Biophys.* **316**: 399–406.

Tenhunen, K., Laan, M., Manninen, T., Palotie, A., Peltonen, L. & Jalanko, A. 1995. Molecular cloning, chromosomal assignment, and expression of the mouse aspartylglucosaminidase gene. *Genomics* **30**: 244–250.

van Belleghem, S.M., Roelofs, D., van Houdt, J. & Hendrickx, F. 2012. *De novo* transcriptome assembly and SNP discovery in the wing polymorphic salt marsh beetle *Pogonus chalceus* (Coleoptera, Carabidae). *PLoS One* **7**: e42605.

Vårdal, H. 2004. From Parasitoids to Gall Inducers and Inquilines - Morphological Evolution in Cynipoid Wasps. Ph.D. Thesis. Uppsala University, Sweden.

Vårdal, H. 2006. Venom gland and reservoir morphology in cynipoid wasps. *Arthropod Struct. Dev.* **35**: 127–136.

Vincent, B., Kaeslin, M., Roth, T., Heller, M., Poulain, J., Cousserans, F., *et al.* 2010. The venom composition of the parasitic wasp *Chelonus inanitus* resolved by combined expressed sequence tags analysis and proteomic approach. *BMC Genomics* **11**: 693.

Vinchon, S., Moreau, S.J.M., Drezen, J.M., Prévost, G. & Cherqui, A. 2010. Molecular and biochemical analysis of an aspartylglucosaminidase from the venom of the parasitoid wasp *Asobara tabida* (Hymenoptera: Braconidae). *Insect Biochem. Mol. Biol.* **40**: 38–48.

Vogt, R.G., Callahan, F.E., Rogers, M.E. & Dickens, J.C. 1999. Odorant binding protein diversity and distribution among the insect orders, as indicated by LAP, an OBP-related protein of the true bug *Lygus lineolaris* (Hemiptera, Heteroptera). *Chem. Senses* **24**: 481–495.

von Aderkas, P., Rouault, G., Wagner, R., Chiwocha, S. & Roques, A. 2005a. Multinucleate storage cells in Douglas-fir (*Pseudotsuga menziesii* (Mirbel) Franco) and the effect of seed parasitism by the chalcid Megastigmus spermotrophus Wachtl. *Heredity* **94**: 616–622.

von Aderkas, P., Rouault, G., Wagner, R., Rohr, R. & Roques, A. 2005b. Seed parasitism redirects ovule development in Douglas fir. *Proc. R. Soc. B Biol. Sci.* **272**: 1491–1496.

Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J. & Colbourne, J.K. 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**: 343–348.

Whitfield, J.B. 2003. Phylogenetic insights into the evolution of parasitism in Hymenoptera. *Adv. Parasitol.* **54**: 69–100.

Xiao, J.-H., Yue, Z., Jia, L.-Y., Yang, X.-H., Niu, L.-H., Wang, Z., *et al.* 2013. Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. *Genome Biol.* **14**: R141.

Yamaguchi, H., Tanaka, H., Hasegawa, M., Tokuda, M., Asami, T. & Suzuki, Y. 2012. Phytohormones and willow gall induction by a gall-inducing sawfly. *New Phytol.* **196**: 586–595.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T.L. 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**: 134.

Zhu, J., Ye, G. & Hu, C. 2008. Molecular cloning and characterization of acid phosphatase in venom of the endoparasitoid wasp *Pteromalus puparum* (Hymenoptera: Pteromalidae). *Toxicon* **51**: 1391–9.

Zhu, J.-Y., Fang, Q., Wang, L., Hu, C. & Ye, G.-Y. 2010. Proteomic analysis of the venom from the endoparasitoid wasp *Pteromalus puparum* (Hymenoptera: Pteromalidae). *Arch. Insect Biochem. Physiol.* **75**: 28–44.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

AP designed experiments, collected and analyzed data and wrote the paper; PvA and JE conceived the project, designed experiments and commented on the manuscript; SP conceived the project, designed experiments and wrote the paper. CL and JD conducted RT-PCR validation.

Data accessibility

The *M. spermotrophus* final transcriptome has been deposited at DDB/EMBL/GenBank under the accession GCPB00000000. The version described in this paper is the first version, GCPB01000000.

667 Table 1. Illumina sequencing output for the *Megastigmus spermotrophus* whole insect cDNA

Library:	Larva	Male	Lab-Reared Female	Wild Female	**Total**
Total number raw paired-end reads:	62,396,989	64,991,169	63,560,118	46,037,319	**236,985,595**
Raw paired-end read total length (Gbp):	12.48	13.00	12.71	9.21	**47.40**
Total number filtered paired-end reads:	49,883,162	52,200,155	50,592,674	36,311,907	**188,987,898**
Total number filtered single-end reads:	6,023,891	6,299,194	6,135,548	4,478,658	**22,937,291**

668

669 Table 2. *Megastigmus spermotrophus* transcriptome clustering results

Clustering Method	Number of Contigs	N50 (min:200 bp)
Trans-ABySS	1,361,656	1,690
CD-HIT-EST	296,711	1,570
TIGR-TGICL	193,412	2,420
Singleton Removal	143,306	2,420

Table 3. List of *Nasonia vitripennis* venom proteins with significant similarities to *de novo* assembled sequence in the *Megastigmus spermotrophus* transcriptome, organized by venom type.

N. vitripennis venom with significant sequence similarity (E-value = 10^{-7}) to *M. spermotrophus* assembled transcripts	Final annotation assigned in the *M. spermotrophus* transcriptome[a]
Proteases and peptidases	
Metalloprotease-like precursor	
Serine protease precursor	
Serine protease 16 precursor	●
Serine protease homolog 21 precursor	●
Serine protease 22 precursor	●
Serine protease homolog 29 precursor	
Serine protease 33 precursor	
Serine protease homolog 42 isoform 2 precursor	
Serine protease homolog 42 isoform 1 precursor	●
Serine protease 50 precursor	
Serine protease 96 precursor	
Serine protease 97 precursor	
Protease inhibitors	
Cysteine-rich/KU venom protein precursor	●
Cysteine-rich/pacifastin venom protein 1 precursor	
Cysteine-rich/pacifastin venom protein 2 precursor	●
Kazal type serine protease inhibitor-like venom protein 1 precursor	●
Carbohydrate metabolism	
Chitinase 5 precursor	●
Glucose dehydrogenase-like venom protein	
Glucose dehydrogenase-like venom protein	
DNA metabolism	
Endonuclease-like venom protein precursor	●
Inosine-uridine preferring nucleoside hydrolase-like precursor	●
Glutathione metabolism	
Gamma-glutamyl cyclotransferase-like venom protein isoform 1 precursor	●
Gamma-glutamyl cyclotransferase-like venom protein isoform 2	
Esterases	
Venom acid phosphatase-like precursor	●
Venom acid phosphatase-like precursor	●
Multiple inositol polyphosphate phosphatase-like venom protein precursor	●
Carboxylesterase clade B, member 2 precursor	
Lipase A-like precursor	

Table 3 (Continued)

N. vitripennis venom with significant sequence similarity (E-value = 10^{-7}) to *M. spermotrophus* assembled transcripts	Final annotation assigned in the *M. spermotrophus* transcriptome[a]
Recognition/binding proteins	
Gram-negative bacteria binding protein 1-2 precursor	●[b]
Low-density lipoprotein receptor-like venom protein precursor	
Immunity related proteins	
C1q-like venom protein precursor	●
Others	
Aminotransferase-like venom protein 1 precursor	
Aminotransferase-like venom protein 2 precursor	
Antigen 5-like protein 1 precursor	
Aspartylglucosaminidase precursor	●[b]
Laccase-like precursor	
Venom laccase precursor	●
Unknown	
Venom protein D precursor	
Venom protein F precursor	●
Venom protein M precursor	●
Venom protein R precursor	●[b]

[a] BLAST hit to *N. vitripennis* venom with lower E-value than to non-venomous homologs

[b] Differentially greater expression in lab-reared and wild female *Megastigmus spermotrophus* transcriptome libraries compared to larvae and males

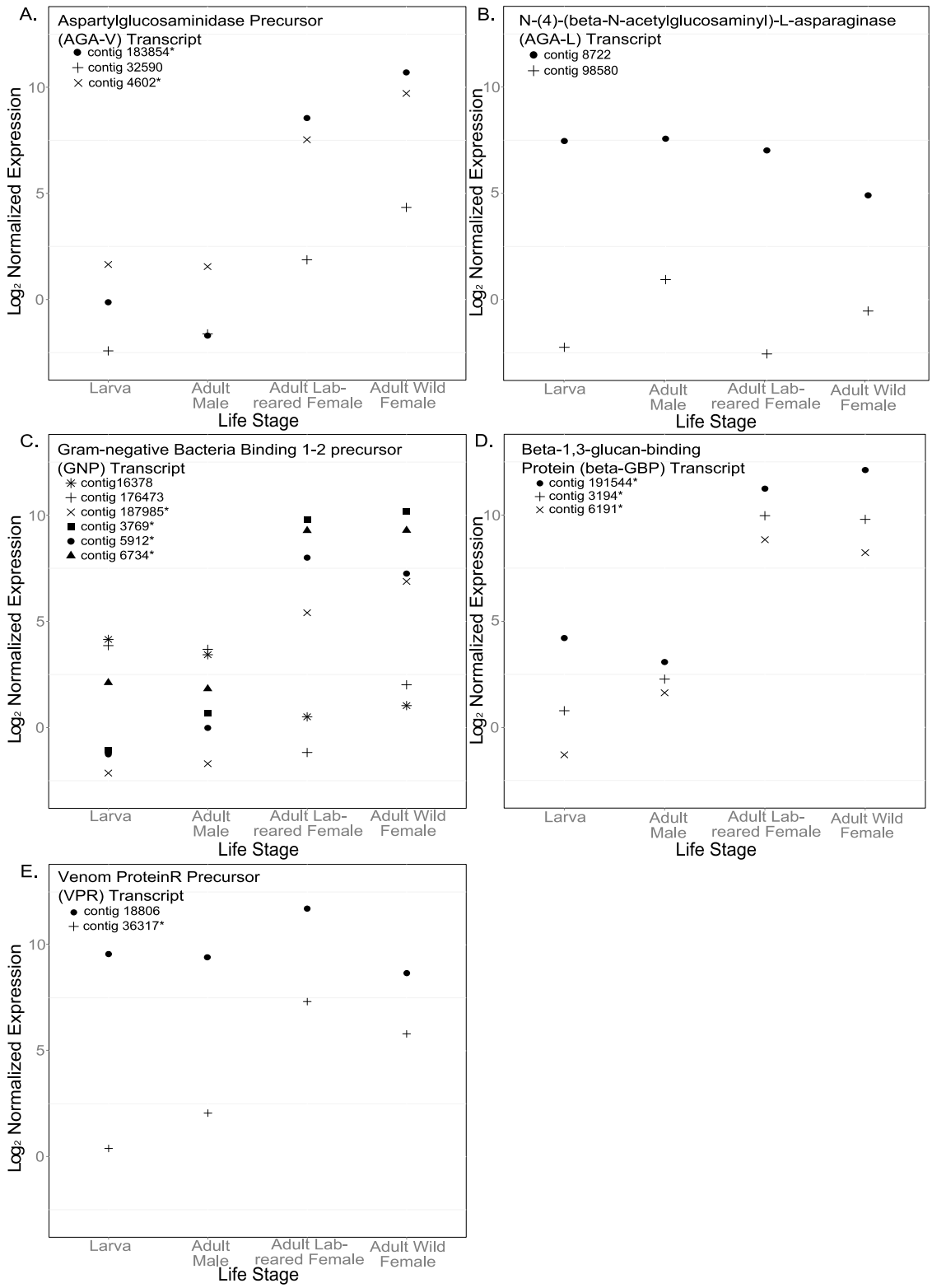
A.
Aspartylglucosaminidase Precursor
(AGA-V) Transcript
contig 183854*
contig 32590
contig 4602*
B.
N-(4)-(beta-N-acetylglucosaminyl)-L-asparaginase
(AGA-L) Transcript
contig 8722
contig 98580
C.
Gram-negative Bacteria Binding 1-2 precursor
(GNP) Transcript
contig16378
contig 176473
contig 187985*
contig 3769*
contig 5912*
contig 6734*
D.
Beta-1,3-glucan-binding
Protein (beta-GBP) Transcript
contig 191544*
contig 3194*
contig 6191*
E.
Venom ProteinR Precursor
(VPR) Transcript
contig 18806
contig 36317*
Log2 Normalized Expression
0
5
10
Larva
Adult Male
Adult Lab-reared Female
Adult Wild Female
Life Stage

Figure 1. Normalized expression of putative venom transcripts and their physiological paralogs in larva, male, lab-reared female and wild female *Megastigmus spermotrophus* transcriptome libraries. **A.** aspartylglucosaminidase precursor (AGA-V), **B.** N-(4)-(beta-N-acetylglucosaminyl)-L-asparaginase (AGA-L), **C.** gram-negative bacteria binding 1-2 precursor (GNP), **D.** beta-1,3-glucan-binding protein (beta-GBP) and **E.** protein R precursor (VPR). *Denotes contigs that are highly differentially expressed in female*s* based on non-parametric statistical analysis.

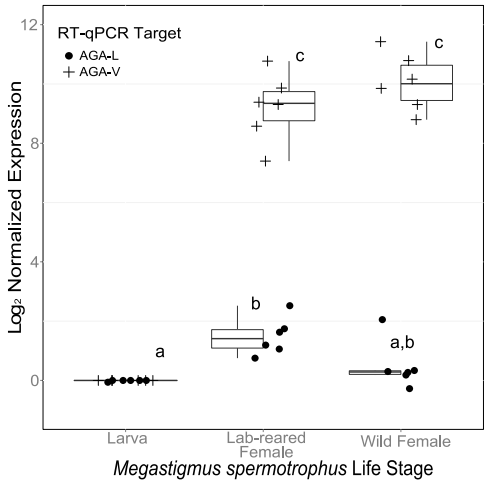

Figure 2. Normalized expression of putative venom AGA-V and its non-venomous paralog AGA-L in larva, lab-reared female and wild female *Megastigmus spermotrophus* based on quantitative real-time PCR.

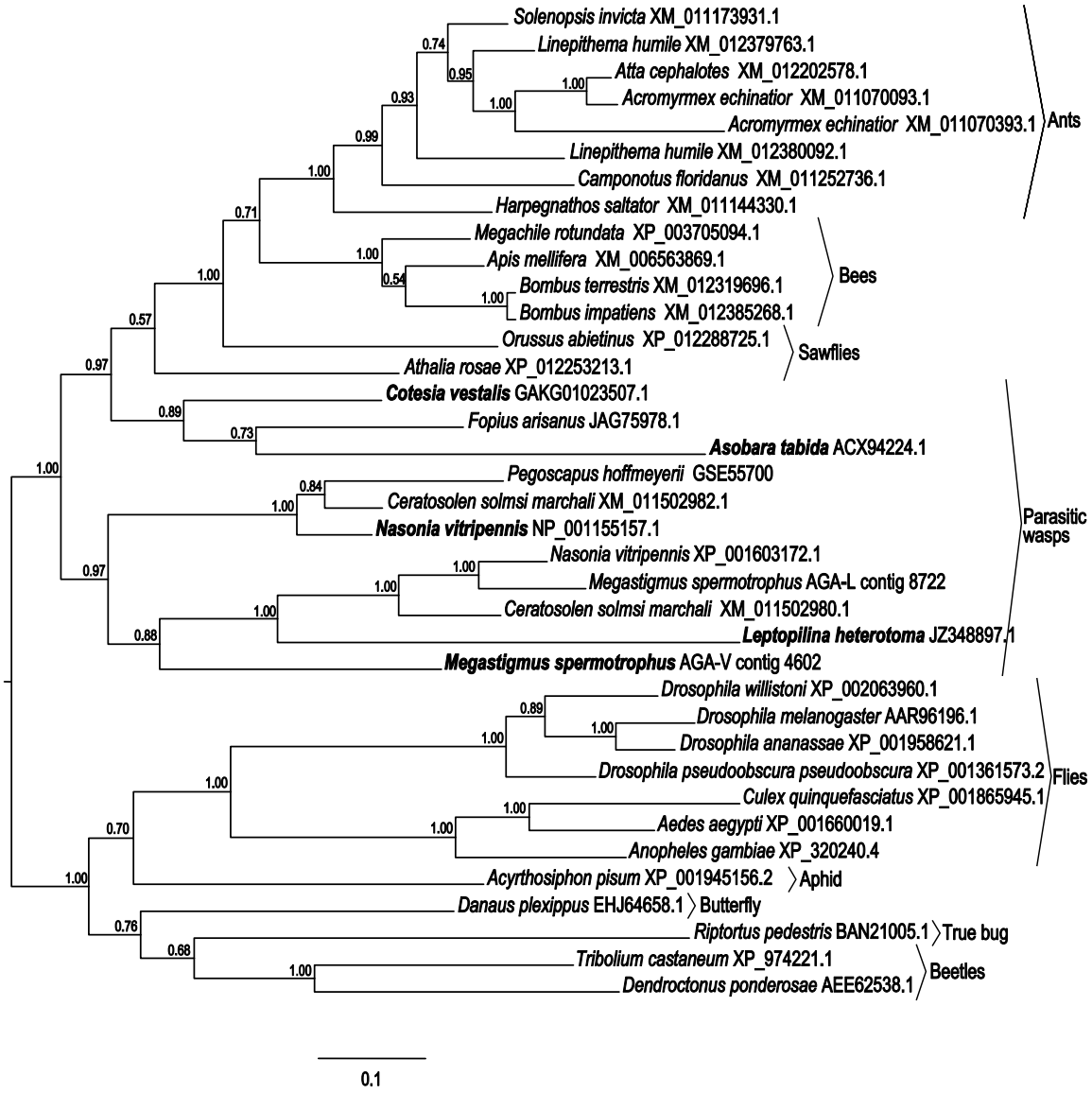

Figure 3: Molecular phylogenetic analysis for aspartylglucosaminidase protein sequence from insects using Bayesian methods and a WAG model of amino acid substitution, and with mid-point rooting. Numbers next to the nodes indicate posterior probabilities. Taxa in bold text represent putative or known venomous proteins.

Supporting Information: Transcriptome Analysis Provides Insight into Venom Evolution in a Seed-Parasitic Wasp, *Megastigmus spermotrophus*.

Amber R. Paulson[1], Cuong H. Le, Jamie C. Dickson, Jürgen Ehlting, Patrick von Aderkas[2] and Steve J. Perlman[3*]

Department of Biology, University of Victoria, Victoria, British Columbia, Canada.

[*] Integrated Microbial Biodiversity Program, Canadian Institute For Advanced Research, Toronto, Ontario, Canada.

[1] amber.rose.paulson@gmail.com and corresponding author, [2] pvonader@uvic.ca, [3] stevep@uvic.ca

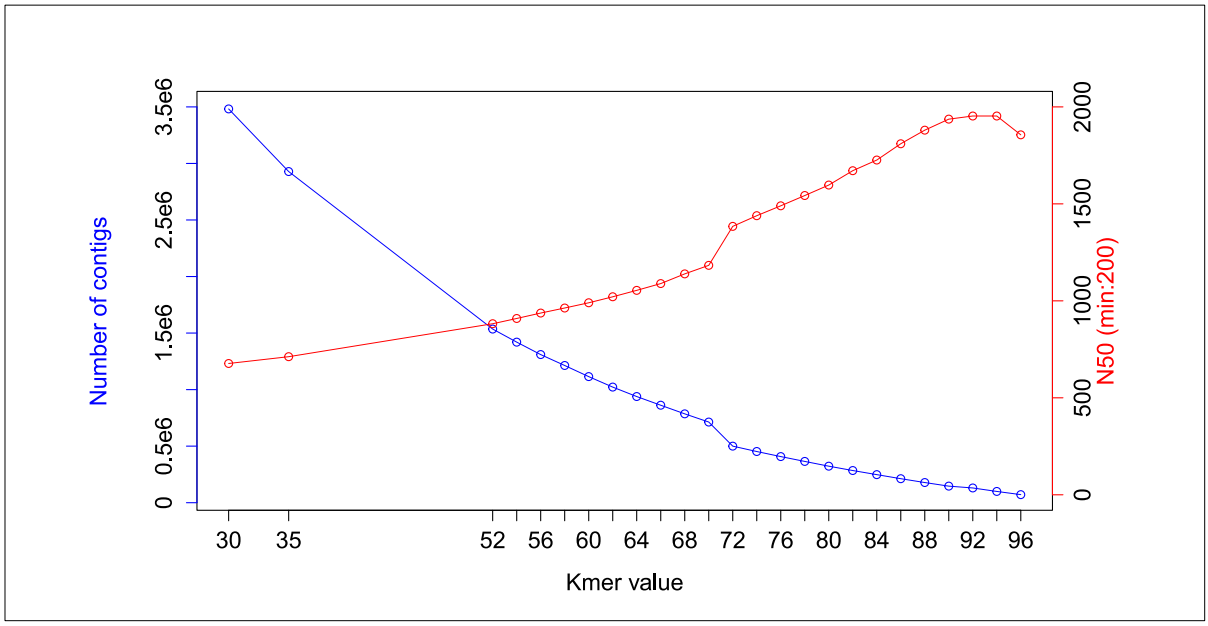

Figure S1. *Megastigmus spermotrophus* ABySS multi-k *de novo* transcriptome assembly results: Number of contigs (blue) and N50 value (red).

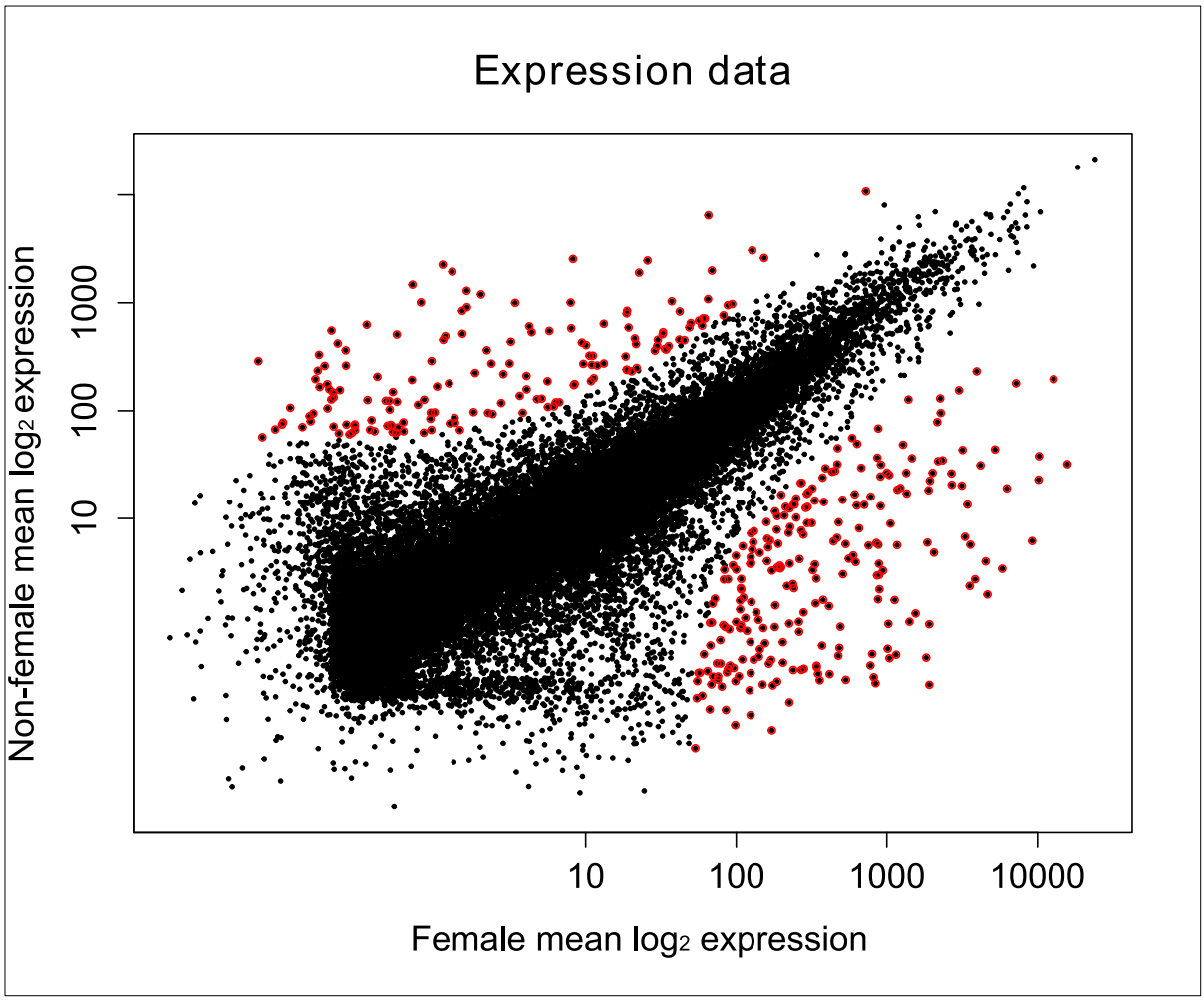

Figure S2. Log_2 mean normalized expression values from female (wild and lab-reared) and non-female (larva and adult male) transcriptome libraries of *Megastigmus spermotrophus*. Likely differentially expressed features are highlighted in red, as determined by the NOISeq-sim algorithm.

Table S1. *Megastigmus spermotrophus* transcriptome putative venom proteins and library expression

Nasonia vitripennis Annotation	Accession	E-Value	ContigID	Length	GC content	Log_2 Normalized Expression			
						Larvae	Male	Reared female	Wild female
Proteases and peptidases									
Serine protease 16 precursor	NP_001155077.1	0	contig13012	2045	0.438	7.50	9.71	9.18	9.28
Serine protease 16 precursor	NP_001155077.1	2.00E-22	contig188559	410	0.502	3.58	5.77	4.77	0.53
Serine protease homolog 21 precursor	NP_001155060.1	0	contig17900	2237	0.45	10.54	9.94	10.84	10.01
Serine protease homolog 21 precursor	NP_001155060.1	0	contig17902	997	0.504	2.37	3.18	3.12	5.14
Serine protease 22 precursor	NP_001155043.1	2.00E-168	contig190198	2162	0.42	1.77	3.19	5.40	5.70
Serine protease 22 precursor	NP_001155043.1	3.00E-166	contig8097	2565	0.408	5.03	5.48	5.82	4.23
Serine protease 22 precursor	NP_001155043.1	1.00E-110	contig192437	1510	0.465	4.78	4.52	5.42	4.11
Serine protease homolog 42 isoform 1 precursor	NP_001155078.1	2.00E-171	contig16485	1114	0.411	4.83	7.02	6.55	5.37
Protease inhibitors									
Cysteine-rich/KU venom protein precursor	NP_001154998.1	0	contig24831	3485	0.453	8.37	8.65	8.15	9.44
Cysteine-rich/pacifastin venom protein 2 precursor	NP_001154996.1	5.00E-31	contig188721	907	0.411	8.36	5.14	7.73	2.96
Cysteine-rich/pacifastin venom protein 2 precursor	NP_001154996.1	3.00E-21	contig176319	576	0.382	3.42	-1.77	0.29	-1.84
Kazal type serine protease inhibitor-like venom protein 1 precursor	NP_001154995.1	6.00E-11	contig177945	869	0.328	-1.40	-1.74	-1.44	-1.75
Kazal type serine protease inhibitor-like venom protein 1 precursor	NP_001154995.1	2.00E-11	contig5130	1877	0.317	3.94	3.80	3.94	3.36
Carbohydrate metabolism									
Chitinase 5 precursor	NP_001155084.1	0	contig12716	2407	0.482	8.49	8.39	9.65	7.33
Chitinase 5 precursor	NP_001155084.1	3.00E-96	contig188098	663	0.529	2.26	1.34	-2.26	-1.51
Chitinase 5 precursor	NP_001155084.1	9.00E-94	contig12717	1155	0.508	1.28	-0.39	2.27	-0.67
Chitinase 5 precursor	NP_001155084.1	1.00E-89	contig163647	650	0.471	1.85	-1.34	2.70	2.12
Chitinase 5 precursor	NP_001155084.1	3.00E-11	contig150614	471	0.548	-1.55	-0.03	1.91	-0.54
DNA metabolism									
Endonuclease-like venom protein precursor	NP_001155087.1	9.00E-118	contig99625	1399	0.375	1.07	2.30	0.24	0.61
Inosine-uridine preferring nucleoside hydrolase-like precursor	NP_001155174.1	6.00E-112	contig43405	1290	0.346	-1.60	4.05	1.73	2.46
Glutathione metabolism									
Gamma-glutamyl cyclotransferase-like venom protein isoform 1 precursor	NP_001155144.1	1.00E-88	contig1982	1055	0.344	6.96	6.98	5.89	4.75
Gamma-glutamyl cyclotransferase-like venom protein isoform 1 precursor	NP_001155144.1	1.00E-49	contig14516	981	0.293	2.96	0.47	0.94	0.71

Table S1 (Continued)

Nasonia vitripennis Annotation	Accession	E-Value	ContigID	Length	GC content	Log_2 Normalized Expression			
						Larvae	Male	Reared female	Wild female
Esterases									
Venom acid phosphatase-like precursor	NP_001155146.1	4.00E-145	contig188037	1587	0.378	3.67	5.49	4.55	4.40
Venom acid phosphatase-like precursor	NP_001155147.1	2.00E-128	contig18791	2289	0.366	6.28	4.99	4.94	5.61
Multiple inositol polyphosphate phosphatase-like venom protein precursor	NP_001155183.1	3.00E-162	contig149688	2412	0.432	1.55	4.67	3.57	0.64
Recognition/binding proteins									
Gram-negative bacteria binding protein 1-2 precursor	NP_001155149.1	2.00E-60	contig176473	949	0.364	3.86	3.68	-1.17	2.01
Gram-negative bacteria binding protein 1-2 precursor	NP_001155149.1	3.00E-56	contig16378	988	0.347	4.15	3.43	0.51	1.04
Gram-negative bacteria binding protein 1-2 precursor	NP_001155149.1	2.00E-45	contig6734	816	0.349	2.11	1.83	9.27	9.28
Gram-negative bacteria binding protein 1-2 precursor	NP_001155149.1	5.00E-39	contig3769	476	0.389	-1.05	0.68	9.79	10.19
Gram-negative bacteria binding protein 1-2 precursor	NP_001155149.1	1.00E-36	contig187985	415	0.429	-2.14	-1.70	5.41	6.88
Gram-negative bacteria binding protein 1-2 precursor	NP_001155149.1	4.00E-33	contig5912	764	0.322	-1.26	-0.02	8.00	7.25
Immunity related proteins									
C1q-like venom protein precursor	NP_001155152.1	8.00E-68	contig188300	1255	0.474	4.03	6.61	5.08	6.26
C1q-like venom protein precursor	NP_001155152.1	6.00E-68	contig190886	1213	0.472	4.36	7.39	6.62	7.04
C1q-like venom protein precursor	NP_001155152.1	2.00E-66	contig188931	1660	0.502	-0.86	3.56	1.48	1.02
Others									
Aspartylglucosaminidase precursor	NP_001155157.1	8.00E-129	contig4602	4398	0.388	1.65	1.55	7.53	9.71
Aspartylglucosaminidase precursor	NP_001155157.1	8.00E-107	contig183854	1038	0.425	-0.13	-1.69	8.55	10.69
Aspartylglucosaminidase precursor	NP_001155157.1	3.00E-78	contig32590	912	0.419	-2.42	-1.62	1.87	4.34
Venom laccase precursor	NP_001155158.1	0	contig1432	3014	0.363	3.68	2.37	-2.32	-0.33
Unknown									
Venom protein F precursor	NP_001155160.1	4.00E-97	contig40539	1307	0.438	9.57	8.41	9.04	9.05
Venom protein M precursor	NP_001155030.1	1.00E-07	contig1276	551	0.423	6.44	3.30	2.78	2.01
Venom protein R precursor	NP_001155164.1	2.00E-38	contig18806	1389	0.341	9.59	9.42	11.73	8.68
Venom protein R precursor	NP_001155164.1	1.00E-09	contig36317	657	0.393	0.39	2.06	7.32	5.80

Table S2. *Nasonia vitripennis* venom query

Protein Name	Accession number
Proteases and peptidases	
metalloprotease-like precursor	NP_001155006.1
serine protease precursor	NP_001155015.1
serine protease precursor	NP_001164348.1
serine protease 16 precursor	NP_001155077.1
serine protease homolog 21 precursor	NP_001155060.1
serine protease 22 precursor	NP_001155043.1
serine protease homolog 29 precursor	NP_001155016.1
serine protease 33 precursor	NP_001155017.1
serine protease homolog 42 isoform 2 precursor	NP_001155079.1
serine protease homolog 42 isoform 1 precursor	NP_001155078.1
serine protease 50 precursor	NP_001155076.1
serine protease 96 precursor	NP_001155014.1
serine protease 97 precursor	NP_001155042.1
Protease inhibitors	
cysteine-rich/KU venom protein precursor	NP_001154998.1
cysteine-rich/pacifastin venom protein 1 precursor	NP_001154997.1
cysteine-rich/pacifastin venom protein 2 precursor	NP_001154996.1
cysteine-rich/TIL venom protein 2 precursor	NP_001155022.1
Kazal type serine protease inhibitor-like venom protein 1 precursor	NP_001154995.1
Kazal type serine protease inhibitor-like venom protein 2 precursor	NP_001164350.1
small serine proteinase inhibitor-like venom protein precursor	NP_001155083.1
Carbohydrate metabolism	
chitinase 5 precursor	NP_001155084.1
glucose dehydrogenase-like venom protein	NP_001155086.1
glucose dehydrogenase-like venom protein	NP_001155085.1
DNA metabolism	
endonuclease-like venom protein precursor	NP_001155087.1
inosine-uridine preferring nucleoside hydrolase-like precursor	NP_001155172.1
Glutathione metabolism	
gamma-glutamyl cyclotransferase-like venom protein isoform 1 precursor	NP_001155144.1
gamma-glutamyl cyclotransferase-like venom protein isoform 2	NP_001155145.1
Esterases	
venom acid phosphatase-like precursor	NP_001155147.1
venom acid phosphatase-like precursor	NP_001155146.1
multiple inositol polyphosphate phosphatase-like venom protein precursor	NP_001155183.1
carboxylesterase clade B, member 2 precursor	NP_001155148.1
lipase-like venom protein precursor	NP_001155039.1
lipase A-like precursor	NP_001154991.1

Table S2 (Continued)

Protein Name	Accession number
Recognition/binding proteins	
gram-negative bacteria binding protein 1-2 precursor	NP_001155149.1
chitin binding protein-like venom protein precursor	NP_001164343.1
GOBP-like venom protein precursor	NP_001155150.1
low-density lipoprotein receptor-like venom protein precursor	NP_001155040.1
Immunity related proteins	
calreticulin precursor	NP_001155151.1
C1q-like venom protein precursor	NP_001155152.1
Others	
aminotransferase-like venom protein 1 precursor	NP_001155153.1
aminotransferase-like venom protein 2 precursor	NP_001155156.1
antigen 5-like protein 1 precursor	NP_001155154.1
aspartylglucosaminidase precursor	NP_001155157.1
laccase-like precursor	NP_001155159.1
venom laccase precursor	NP_001155158.1
Unknowns	
venom protein D precursor	NP_001155171.1
venom protein F precursor	NP_001155160.1
venom protein G precursor	NP_001164344.1
venom protein H precursor	NP_001155027.1
venom protein I precursor	NP_001164346.1
venom protein J precursor	NP_001164347.1
venom protein K precursor	NP_001155028.1
venom protein L precursor	NP_001155029.1
venom protein M precursor	NP_001155030.1
venom protein N precursor	NP_001164349.1
venom protein O precursor	NP_001155031.1
venom protein Q precursor	NP_001155161.1
venom protein R precursor	NP_001155164.1
venom protein T precursor	NP_001155166.1
venom protein U precursor	NP_001155170.1
venom protein V precursor	NP_001155041.1
venom protein X precursor	NP_001155167.1
venom protein Y precursor	NP_001155168.1
venom protein Z precursor	NP_001155169.1

Table S3. Quantitative real-time PCR Primers

Gene	Forward Primer (3' - 5')	Reverse Primer (3' - 5')	Product Size (bp)
AGA-V	TCAAAGGGTCACGATGGACG	AGCGAATTCAGTCGCGAGAT	133
AGA-L	TTTCAGTGGCTCGGAAGGTC	TCGGTTTGCAAGCTCTCCTT	112
VPR	TACGGCTAAAGCCACGAACA	TAGCCGAAACGATCGCAGAA	116
RPL	GCCCTAACCTTGGCGTGATA	TGCCCCAAGACCAGTAAACC	75