Integrative Approach Reveals Composition of Endoparasitoid Wasp Venoms

Jeremy Goecks, Emory University
Nathan T. Mortimer, Emory University
James A. Mobley, University of Alabama-Birmingham
Gregory J. Bowersock, University of Alabama-Birmingham
James Taylor, Emory University
Todd A. Schlenke, Emory University

Journal Title: PLoS ONE
Volume: Volume 8, Number 5
Publisher: Public Library of Science | 2013-05-23, Pages e64125-e64125
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1371/journal.pone.0064125
Permanent URL: http://pid.emory.edu/ark:/25593/f7tsk

Final published version: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0064125

Copyright information:
© 2013 Goecks et al.
This is an Open Access article distributed under the terms of the Creative Commons Attribution 2.0 Generic License (http://creativecommons.org/licenses/by/2.0/), which permits making multiple copies, distribution of derivative works, distribution, public display, and publicly performance, provided the original work is properly cited. This license requires credit be given to copyright holder and/or author, copyright and license notices be kept intact.

Accessed January 28, 2019 6:30 PM EST
Integrative Approach Reveals Composition of Endoparasitoid Wasp Venoms

Jeremy Goecks¹,²,³, Nathan T. Mortimer¹,³, James A. Mobley³,⁴, Gregory J. Bowersock³,⁴, James Taylor¹,²*, Todd A. Schlenke¹*

¹ Department of Biology, Emory University, Atlanta, Georgia, United States of America, ² Department of Mathematics and Computer Science, Emory University, Atlanta, Georgia, United States of America, ³ Department of Surgery, University of Alabama-Birmingham, Birmingham, Alabama, United States of America, ⁴ Comprehensive Cancer Center, University of Alabama-Birmingham, Birmingham, Alabama, United States of America

Abstract

The fruit fly Drosophila melanogaster and its endoparasitoid wasps are a developing model system for interactions between host immune responses and parasite virulence mechanisms. In this system, wasps use diverse venom cocktails to suppress the conserved fly cellular encapsulation response. Although numerous genetic tools allow detailed characterization of fly immune genes, lack of wasp genomic information has hindered characterization of the parasite side of the interaction. Here, we use high-throughput nucleic acid and amino acid sequencing methods to describe the venoms of two related Drosophila endoparasitoids with distinct infection strategies, Leptopilina bouardi and L. heterotoma. Using RNA-seq, we assembled and quantified libraries of transcript sequences from female wasp abdomens. Next, we used mass spectrometry to sequence peptides derived from dissected venom gland lumens. We then mapped the peptide spectral data against the abdomen transcriptomes to identify a set of putative venom genes for each wasp species. Our approach captured the three venom genes previously characterized in L. bouardi by traditional CDNA cloning methods as well as numerous new venom genes that were subsequently validated by a combination of RT-PCR, blast comparisons, and secretion signal sequence search. Overall, 129 proteins were found to comprise L. bouardi venom and 176 proteins were found to comprise L. heterotoma venom. We found significant overlap in L. bouardi and L. heterotoma venom composition but also distinct differences that may underlie their unique infection strategies. Our joint transcriptomic-proteomic approach for endoparasitoid wasp venoms is generally applicable to identification of functional protein subsets from any non-genome sequenced organism.


Introduction

The fruit fly Drosophila melanogaster is a genetic model system that has been important in the study of conserved aspects of host innate immune responses against pathogens [1]. This work has proven useful for understanding homologous aspects of human innate immunity, such as the role of Toll-like receptors in mounting humoral immune responses against microbes [2], and has been particularly useful for understanding homologous immune defenses in important insects such as vectors of human disease, crop pollinators, and agricultural pests [3–5]. However, many aspects of innate immunity in this model system remain incompletely characterized, especially with respect to cellular immune responses [6–8].

Some of the most common pathogens of fruit flies are endoparasitoid wasps, which can infect greater than 50% of flies in natural populations [9–11]. These wasps lay their eggs in fruit fly larvae and pupae, and the developing wasp larvae eventually consume and kill their fly hosts. However, wasp infection of D. melanogaster larvae induces a cellular immune response against wasp eggs termed melanotic encapsulation [6,12]. Encapsulation of foreign tissues is an important aspect of parasite resistance for a wide diversity of arthropods, including insect vectors of human disease [13,14]. Furthermore, endoparasitoid wasps produce cocktails of virulence proteins in their venom glands that they inject into the fly body cavity along with their eggs to suppress this fly immune response, leading to an evolutionary arms race between fly immunity and wasp virulence. Much can be learned about host immune systems by studying the mechanisms that pathogens use to suppress them. Given the genetic tools available in D. melanogaster, the identification of venoms from Drosophila endoparasitoids would allow for detailed studies on the molecular biology and evolution of interacting host immune proteins and parasite virulence proteins.

Our goal in this study was to characterize the venom of two well-studied Figitid endoparasitoid wasps, L. bouardi and L. heterotoma. L. bouardi is a specialist on Drosophila species of the melanogaster subgroup while L. heterotoma is more of a generalist of the Drosophila genus [15]. Furthermore, these wasps have widely differing infection strategies: L. bouardi relies on both cloaking its
eggs from host hemocytes and preventing hemocytes from binding to egg surfaces; *D. melanogaster* hosts infected by *L. boulardi* mount strong transcriptional immune responses even during a successful wasp infection. On the other hand, *L. heterotoma* infection relies on host hemocyte destruction; *D. melanogaster* hosts infected by *L. heterotoma* fail to mount any obvious immune transcriptional response at all [15–24].

One approach that has been used to identify venom proteins from Drosophila endoparasitoids is the cloning of cDNAs derived from venom gland mRNAs. This traditional cloning approach has resulted in the identification of four venom genes from Drosophila endoparasitoid wasps: *LbGAP* (a RhoGAP domain containing protein), a serpin, and an extracellular superoxide dismutase from the Figitid *L. boulardi*, and an aspartylglucosaminidase from the Braconid *Asobara tabida* [20,25–27]. *LbGAP* has been shown to alter fly host hemocyte morphology and function, the serpin and superoxide dismutase have been shown to suppress host production of melanin, and the aspartylglucosaminidase has no known function. However, the number of proteins in the venom cocktails of other parasitoid wasps and in a wide range of other venomous animals is on the order of tens or even hundreds [28–31], suggesting we have only scratched the surface of Drosophila endoparasitoid wasp venom gene identification.

In order to devise an efficient strategy for identification of total venom protein content in Drosophila endoparasitoid wasps, we first considered the strategies used in other wasp venom studies. Venom proteins from the endoparasitoid wasp *Peromatus puparum* were identified by mass spectrometry [32], but due to the absence of *P. puparum* transcript sequence data only 12 of the 56 putative venom proteins could be positively identified, based on homology to known proteins. To get around this problem, partial transcript sequences (ESTs) were used in conjunction with proteomics data to investigate the venom of the endoparasitoid wasp *Chelonus inanitus* [33]. This approach identified 29 venom genes, including several genes unique to *C. inanitus* as well as several genes that have no known homology. Similarly, next generation sequencing of venom gland RNAs combined with a low throughput proteomics approach led to the identification of 8 full-length venom transcripts from the wasp *Microtus hyperodas* [34]. Finally, results from mass spectrometry of *Nasonia vitripennis* venom proteins were mapped back to the completed *N. vitripennis* genome leading to identification of 79 venom genes [35]. Altogether, these studies show that a joint nucleic acid and amino acid sequencing approach is invaluable, and that high throughput sequencing techniques result in a greater ability to identify venom genes.

Here, we took a new high-throughput approach to identify venom proteins from two non-genome sequenced Drosophila endoparasitoid wasps. In our approach, we collected pure venom proteins from gut, fat body, body muscle, and reproductive tract tissues in addition to all the transcripts from the venom gland itself. Total RNA was isolated from abdomen samples, poly(A) mRNAs were pulled down with the use of oligo(dT) beads, and mRNAs were processed into cDNAs for sequencing. We obtained two sets of sequences from each sample, first using an Illumina GA-II to produce 50 bp single-end reads and then using an Illumina Hi-Seq 2000 to produce 2×100 bp paired-end reads. This sequencing produced 16.5 Gb of data for *L. boulardi* and 21.9 Gb for *L. heterotoma*.

Because there is no reference genome for *L. boulardi* or *L. heterotoma*, we performed de novo assembly of transcripts using Trinity version r2011-10-29 [36] to create an abdomen transcriptome from females of each wasp species. Trinity assemblies contig via overlaps amongst reads. Initial assemblies for both wasps produced a very large number of potential transcripts, almost 50,000 for *L. boulardi* and more than 270,000 for *L. heterotoma* (Table 1). Assembly statistics show relatively large N50 values of 2142 bp for *L. boulardi* and 1653 bp for *L. heterotoma* (meaning that half of the total assembled sequence is found in transcripts of at least that length) but small mean lengths, indicating that the assemblies include a large number of short transcripts. Mapping the RNA-seq reads back to the assemblies revealed that the assemblies incorporated most of the read data. For *L. boulardi*, 96% of single-end reads and 88% of paired-end reads mapped back to the assembly; for *L. heterotoma*, 95% of single-end reads and 86% of paired-end reads mapped back to the assembly.

To filter out assembly artifacts and incomplete/short transcripts, we calculated expression levels for each contig using RSEM [37] and removed transcripts that were not sufficiently expressed (Table 1). RSEM probabilistically maps reads to transcripts, calculates expression based on the mapped reads, and reports expression in units of transcripts per million (TPM). Only paired-end reads were used with RSEM because they can be mapped more accurately and hence yield more accurate expression levels. Using the recommended threshold of 1 TPM, a set of expressed transcripts was identified: 43% of *L. boulardi* transcripts and only 12% of *L. heterotoma* transcripts were sufficiently expressed, indicating that many of the original contigs were of low quality. Filtering the assemblies to include only expressed transcripts markedly improved the assemblies, as N50 and mean length both improved dramatically, especially for *L. heterotoma* (Table 1). Thus, expression filtering provided a concise library of putative transcripts from female wasp abdomens. We then translated each transcript sequence from the abdomen transcript libraries of each wasp species in all possible frames, to be used as database against which our proteomics data could be compared.

We used liquid chromatography-tandem mass spectrometry to identify peptide sequences from pools of venom purified from the lumens of wasp venom glands. This procedure resulted in 2,917 *L. boulardi* peptide sequences and 3,625 *L. heterotoma* peptide sequences.

### Results

Integrating Transcriptome and Proteome Data to Find a Putative Set of Venom Proteins

We pursued a joint transcriptomic and proteomic strategy for analyzing *L. boulardi* and *L. heterotoma* venom content (Fig. 1). We first performed RNA-seq on female abdomens of each wasp species. The abdomen contains the venom gland and at approximately 1 mm long is the smallest tissue containing the venom gland that can be easily collected in mass quantities. However, abdomen tissue samples will still contain transcripts from gut, fat body, body muscle, and reproductive tract tissues in addition to all the transcripts from the venom gland itself. Total RNA was isolated from abdomen samples, poly(A) mRNAs were pulled down with the use of oligo(dT) beads, and mRNAs were processed into cDNAs for sequencing. We obtained two sets of sequences from each sample, first using an Illumina GA-II to produce 50 bp single-end reads and then using an Illumina Hi-Seq 2000 to produce 2×100 bp paired-end reads. This sequencing produced 16.5 Gb of data for *L. boulardi* and 21.9 Gb for *L. heterotoma*.

Because there is no reference genome for *L. boulardi* or *L. heterotoma*, we performed de novo assembly of transcripts using Trinity version r2011-10-29 [36] to create an abdomen transcriptome from females of each wasp species. Trinity assemblies contig via overlaps amongst reads. Initial assemblies for both wasps produced a very large number of potential transcripts, almost 50,000 for *L. boulardi* and more than 270,000 for *L. heterotoma* (Table 1). Assembly statistics show relatively large N50 values of 2142 bp for *L. boulardi* and 1653 bp for *L. heterotoma* (meaning that half of the total assembled sequence is found in transcripts of at least that length) but small mean lengths, indicating that the assemblies include a large number of short transcripts. Mapping the RNA-seq reads back to the assemblies revealed that the assemblies incorporated most of the read data. For *L. boulardi*, 96% of single-end reads and 88% of paired-end reads mapped back to the assembly; for *L. heterotoma*, 95% of single-end reads and 86% of paired-end reads mapped back to the assembly.

To filter out assembly artifacts and incomplete/short transcripts, we calculated expression levels for each contig using RSEM [37] and removed transcripts that were not sufficiently expressed (Table 1). RSEM probabilistically maps reads to transcripts, calculates expression based on the mapped reads, and reports expression in units of transcripts per million (TPM). Only paired-end reads were used with RSEM because they can be mapped more accurately and hence yield more accurate expression levels. Using the recommended threshold of 1 TPM, a set of expressed transcripts was identified: 43% of *L. boulardi* transcripts and only 12% of *L. heterotoma* transcripts were sufficiently expressed, indicating that many of the original contigs were of low quality. Filtering the assemblies to include only expressed transcripts markedly improved the assemblies, as N50 and mean length both improved dramatically, especially for *L. heterotoma* (Table 1). Thus, expression filtering provided a concise library of putative transcripts from female wasp abdomens. We then translated each transcript sequence from the abdomen transcript libraries of each wasp species in all possible frames, to be used as database against which our proteomics data could be compared.

We used liquid chromatography-tandem mass spectrometry to identify peptide sequences from pools of venom purified from the lumens of wasp venom glands. This procedure resulted in 2,917 *L. boulardi* peptide sequences and 3,625 *L. heterotoma* peptide sequences.
sequences (Table 2). The program SEQUEST was then used to query these peptide sequences against the wasp abdomen translated transcript databases. Approximately 90% of all venom peptides from each wasp species had unique hits to the abdomen transcript databases and also hit transcripts that were hit by at least one other peptide (Table 2). We considered transcript sequences that were hit by more than one of the filtered venom peptides to code for genuine venom proteins, and this analysis yielded 129 putative venom proteins for *L. boulardi* and 176 venom proteins for *L. heterotoma* (Datasets S1–S4). These putative venom proteins were hit by an average of approximately 20 peptide sequences and average peptide coverage across the full protein sequences was approximately 23% for both wasp species (Table 2). The fact that thousands of peptide sequences queried against tens of thousands of transcript sequences repeatedly hit only 129 and 176 *L. boulardi* and *L. heterotoma* transcript sequences, respectively, indicates that the proteins collected from wasp venom glands represent a unique and specific subset of the total set of proteins made in female wasp abdomens, as expected.

### Expression Specificity of Putative Venom Transcripts

As one means of validating the putative venom proteins, we assayed the tissue specificity of their transcription using RT-PCR. We reasoned that most venom genes would be expressed only in female wasps and specifically in venom glands, although some previously identified Hymenopteran venoms, such as Calreticulin and Heat-shock protein 70 [38,39] are homologous to essential proteins.

#### Table 1. Assembly statistics for the abdomen transcriptomes of each species.

<table>
<thead>
<tr>
<th>Species</th>
<th># Transcripts &gt;100 bp</th>
<th>N50 (bp)</th>
<th>Mean Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. boulardi</em></td>
<td>Initial Assembly</td>
<td>49,972</td>
<td>2142</td>
</tr>
<tr>
<td></td>
<td>Expression-filtered Assembly</td>
<td>21625 (43% of initial set)</td>
<td>3856</td>
</tr>
<tr>
<td><em>L. heterotoma</em></td>
<td>Initial Assembly</td>
<td>269,692</td>
<td>1653</td>
</tr>
<tr>
<td></td>
<td>Expression-filtered Assembly</td>
<td>32293 (12% of initial set)</td>
<td>5674</td>
</tr>
</tbody>
</table>

---

1. Use RNA-seq to sequence the complete wasp abdomen. This produces an RNA-seq library for the wasp abdomen.

2. Because there is no reference genome, assemble RNA-seq data using a de novo approach. This produces an initial transcriptome of the abdomen.

3. Improve initial transcriptome assembly by quantifying transcript expression and removing transcripts that are not expressed at a biologically meaningful level. This produces a final transcriptome of the abdomen.

4. Perform mass spectrometry on venom extracted from venom gland, mapping spectral data onto the abdomen transcriptome. This step yields a set of putative venom proteins.

5. Analyze putative venom proteins by (a) annotating transcripts with function and analyzing via gene ontology enrichment and (b) homology analysis between species.

---

![Figure 1. Our method for identifying endoparasitoid wasp venoms.](doi:10.1371/journal.pone.0064125.g001)

---

Table 1. Assembly statistics for the abdomen transcriptomes of each species.
genes from other organisms and therefore may be globally expressed.

We randomly selected ten putative venom genes from each of *L. boulardi* and *L. heterotoma*, along with a panel of four non-venom control genes (Tables 3 and 4), and attempted to amplify them from adult male and female wasp cDNAs (Fig. 2, top panel in A and B). In both cases, 8 out of 10 venom genes were specifically expressed in females, suggesting that they are linked to parasitism. In both wasps, the functions of the non-female specific genes (arginine kinase and glucose dehydrogenase in *L. boulardi*, and arginine kinase and insulin-like growth factor binding protein in *L. heterotoma*), as well as a lack of other homologous sequences in the respective transcriptomes, suggests they may be essential, and therefore unlikely to be venom specific, although this does not exclude them from having a virulence function.

We next amplified the panel of genes from cDNAs made from dissected venom glands and non-venom 'body' samples (female carcasses following venom gland dissection) (Fig. 2, bottom panel in A and B). Surprisingly, all of the *L. boulardi* venom genes were expressed in both venom gland and body samples, including the two previously identified venom genes LbGAP and serpin. This suggests that these *L. boulardi* venom genes may serve distinct roles in other tissues but behave as virulence factors when expressed in the venom gland. It is also possible that *L. boulardi* virulence genes may be transcribed outside of the venom gland, for example in the ovaries, similar to another Drosophila parasitoid, *Atheta japonica* [40].

From the *L. heterotoma* samples, all 8 female specific genes were found to be specifically expressed in the venom gland (note there is a weak venom gland band in sample 10). Of the two *L. heterotoma* venom genes expressed in both males and females, one was expressed in both female body samples and in the venom gland (insulin-like growth factor binding protein, sample 1), while the second (arginine kinase, sample 6) was found in the female body sample only. However arginine kinases are found in the venoms of several wasp species including *L. boulardi*, *P. puparum* and *Cyphononyx dorsalis* [32,41] and represented one of the more abundant proteins in our mass spectrometry data (Table 4), so it remains possible that the arginine kinase is a venom protein despite its apparent absence from our venom gland RT-PCR assay. Finally, all control wasp genes showed expression in both males and females, and most showed expression in venom glands as well. Overall, this expression data is consistent with the idea that most of our predicted venom genes show specific expression, either in venom glands (*L. heterotoma*) or in female wasps (*L. boulardi*).

Homology to Known Venoms

Another means of validating the putative venom proteins that we identified is to examine homology between our venom proteins and known Hymenopteran venom proteins. Based on the existing

<table>
<thead>
<tr>
<th>Table 2. Proteomics results.</th>
</tr>
</thead>
<tbody>
<tr>
<td># Peptides</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td><em>L. boulardi</em></td>
</tr>
<tr>
<td><em>L. heterotoma</em></td>
</tr>
</tbody>
</table>

Figure 2. RT-PCR of putative venom genes from *L. boulardi* (A) and *L. heterotoma* (B). Sex- (top panel) and tissue- (bottom panel) specific RT-PCRs of venom (1–10) and control (c1–c4) genes. Gene ID numbers correspond to those listed in Tables 3 and 4. doi:10.1371/journal.pone.0064125.g002
Hymenopteran venom literature, we would expect a significant degree of overlap among venom proteins of different species. To test this, we blasted wasp transcript libraries corresponding both to the identified venom proteins and the non-venom 'body' proteins using BLASTx to a database of known Hymenopteran venom proteins extracted from GenBank (868 proteins from 89 different species, database available upon request) and plotted the E-value distributions. We found that venom transcripts from *L. boulardi* and *L. heterotoma* have significantly lower (more significant) BLASTx E-values when queried against the venom database than their respective body transcripts (Fig. 3A–D). This analysis shows that *L. boulardi* and *L. heterotoma* venoms have homology to other Hymenopteran venoms, and provides further evidence that our approach has led to the successful identification of venom proteins.

To further characterize the overlap between venoms, we used the conserved domain database (CDD) [42] to identify putative functional domains in the venom proteins of *L. boulardi*, *L. heterotoma*, and the Hymenopteran venom database. We identified 69 unique conserved domains in the 129 *L. boulardi* venom proteins, 90 unique conserved domains in the 176 *L. heterotoma* venom proteins, and 65 unique conserved domains in the 868 venom proteins that make up the Hymenopteran venom database. There is a great deal of overlap among these conserved domain predictions (Fig. 3E), both among all three lists, and between *L. boulardi* and *L. heterotoma* specifically. Interestingly, there are far fewer domains found in the venom database that are shared with only one of *L. boulardi* or *L. heterotoma*. This suggests that there is a conserved subset of venom proteins (and likely venom activities) shared across Hymenopteran species, but that many venom activities are also unique to species.

**Analysis of Putative Venom Functions**

Wasp venom gland cells secrete venom into the lumen of the gland in preparation for laying eggs in a host, and hence we expected that putative venom proteins would often include secretory signals in their amino acid sequence. We used the SignalP program [43] to find secretory signals in the open reading frames of venom and body proteins (Table 5). For both wasp species, putative venom proteins included secretory signals much more frequently than did body proteins. In *L. boulardi*, 30% of venom proteins included secretory signals while only 5% of body proteins included secretory signals (chi square P = 10\(^{-4}\)). Similarly, 45% of *L. heterotoma* venom proteins and 6% of body proteins included secretory signals (chi square P = 10\(^{-4}\)). The significantly heightened presence of secretory signals in our lists of putative proteins (Table 5) suggests that these wasps secrete proteins that are not only toxic, but also serve to aid in the propagation of the species.
venom proteins is further validation that our venom identification strategy was successful.

To better understand functions of the putative venom proteins, we performed an enrichment analysis using gene ontology (GO) annotations (Tables 5 and S1). Whereas a majority of venom transcripts could be assigned to conserved protein families (avg. 72%) this was only possible for a minority of the total transcripts (avg. 39%). This difference was expected given that venom transcripts have all been verified using protein spectral data. For genes assigned to protein families, similar proportions of venom transcripts and total transcripts could be assigned a GO annotation (Table 5).

The known venom proteins were not included within the enriched GO terms in *L. boulardi* venom. This suggests that rather than clustering into specific functions, virulence genes may encode a broad range of activities, and thus target multiple host mechanisms. Interestingly we found that *L. boulardi* venom is enriched in proteins classified as having ‘antioxidant activity’ (GO:0016209) (Table 6). *D. melanogaster* larvae are thought to generate oxygen radicals as part of the immune response [44–47],
and perhaps the antioxidant proteins in *L. boulardi* venom counteract this defense.

Other parasitoid wasps have been shown to alter host physiology and metabolism [48,49]. We found that the only enriched GO term shared between *L. boulardi* and *L. heterotoma* venom is 'glycolysis' (GO:0006096)(Table 7), and that additionally *L. boulardi* venom is also enriched in 'lipid transporter activity' (GO:0005319). It seems unlikely that glycolytic enzymes or lipid transport molecules would play a role in parasitoid virulence, but their presence supports the idea that *L. boulardi* and *L. heterotoma* venom proteins may similarly regulate host metabolism.

We found four related enriched GO terms in *L. heterotoma* venom, ‘guanyl ribonucleotide binding’ (GO:0032561), ‘GTP binding’ (GO:0005525), ‘guanyl nucleotide binding’ (GO:0019001) and ‘GTPase activity’ (GO:0003924). The GTPase activity category is composed of 12 genes, including homologs of *L. heterotoma* Rho GTPases in the fly-wasp interaction cannot be encapsulated of wasp eggs [50–54]. Although the exact roles of *L. heterotoma* venom GTPases in the fly-wasp interaction cannot be known without experimentation, their identification shows the utility of our high throughput, unbiased venom identification approach, and this result provides a specific basis for further mechanistic study.

### Comparison of Species Venom Transcripts

Our transcriptome data allows us to investigate the amount of sequence homology between *L. boulardi* and *L. heterotoma*, and particularly the homology between venom genes. We aligned the venom and body transcriptome subsets of each wasp to the venom, body and complete transcriptomes of the other using tBLASTx and compared the distributions of E-values. We found that *L. boulardi* and *L. heterotoma* venom sequences are significantly more

### Table 6. *L. boulardi* venom proteins with annotated 'antioxidant' activity (GO:0016209).

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Sequence ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peroxiredoxin-1</td>
<td>comp193_c0_seq1</td>
</tr>
<tr>
<td>Protein disulfide-isomerase</td>
<td>comp629_c0_seq1</td>
</tr>
<tr>
<td>Peroxiredoxin-6</td>
<td>comp3099_c0_seq1</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0064125.t006

### Table 7. Wasp venom proteins with annotated ‘glycolysis’ activity (GO:0006096).

<table>
<thead>
<tr>
<th>Wasp species</th>
<th>Annotation</th>
<th>Sequence ID</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. boulardi</em></td>
<td>Fructose-bisphosphate aldolase</td>
<td>comp413_c1_seq2</td>
</tr>
<tr>
<td><em>L. boulardi</em></td>
<td>Pyruvate kinase</td>
<td>comp431_c0_seq1</td>
</tr>
<tr>
<td><em>L. boulardi</em></td>
<td>Enolase</td>
<td>comp464_c0_seq1</td>
</tr>
<tr>
<td><em>L. boulardi</em></td>
<td>Phosphoglycerate kinase</td>
<td>comp1646_c0_seq1</td>
</tr>
<tr>
<td><em>L. heterotoma</em></td>
<td>Fructose-bisphosphate aldolase</td>
<td>comp307_c0_seq1</td>
</tr>
<tr>
<td><em>L. heterotoma</em></td>
<td>Pyruvate kinase</td>
<td>comp576_c0_seq1</td>
</tr>
<tr>
<td><em>L. heterotoma</em></td>
<td>Enolase</td>
<td>comp644_c0_seq1</td>
</tr>
<tr>
<td><em>L. heterotoma</em></td>
<td>Phosphoglycerate kinase</td>
<td>comp2171_c0_seq1</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0064125.t007
similar to each other (lower tBLASTx E-values, blue bars in Fig. 4A and 4D) than the body sequences are to venom sequences (green bars in Fig. 4A and 4D), and found large tails (>20%) of venom sequence comparisons in the highest homology (lowest E-score) category. These data confirm that our venom gland lumen proteomic approach led to the identification of similar, functionally relevant subsets of venom genes from each species, and that there is a high degree of sequence homology among Leptopilina venoms.

There are also significant differences in E-value distributions when comparing the venom versus body transcript subsets of one wasp to body sequences of the other (Fig. 4B and 4E), and when comparing venom versus body transcript subsets of one wasp to the full transcriptome of the other (Fig. 4C and 4F). Venom sequences show higher homology than body sequences to whatever transcript set they are blasted against. These data suggest that body transcripts, which have not been validated by being matched to wasp protein sequences, may include a significant fraction of erroneous contigs. Consistent with this hypothesis, our transcript assembly and expression validation pipeline yielded one third more L. heterotoma transcripts than L. boulardi transcripts (Table 1), suggesting the L. heterotoma library may contain more false contigs, and blast analyses using L. heterotoma as the query species (Fig. 4E, 4F, and 4F) tended to show greater tails of sequence comparisons in the lowest homology (greatest E-score) categories. However, the existence of erroneous contigs in the body transcript subsets cannot alone explain the high homology observed between L. boulardi and L. heterotoma venom transcripts, as the differences in venom versus body E-value distributions are much greater in venom comparisons (Fig. 4A and 4D) than in other comparisons (Fig. 4B, 4C, 4E, and 4F).

Identification of Known Leptopilina Venom Proteins

To further evaluate the predicted venom proteins, we used BLASTp to compare our predicted venom proteins with previously known Leptopilina venom proteins. Our approach successfully identified all three known L. boulardi venom proteins (Table 8): LbGAP (comp433_c0_seq1), Serpin (comp138_c0_seq1), and Extracellular superoxide dismutase 3 (comp937_c0_seq1) [20,25,26]. Furthermore, we found that L. boulardi has four additional RhoGAP domain containing proteins in its venom that show strong homology to LbGAP, and a second previously unidentified venom Serpin protein (Fig. 5). This suggests that RhoGAP and Serpin proteins play an important role in L. boulardi virulence.

There were no previously characterized L. heterotoma venom proteins, but interestingly, we found that L. heterotoma venom contains homologs of two of the known L. boulardi venom proteins: three LbGAP-like proteins and one Serpin-like protein (Table 9, Fig. 5). Besides LbGAP and Serpin, L. boulardi and L. heterotoma share many other venom protein molecular functions (Tables 8 and 9). For example, both venoms contain an Arginine kinase, a protein type demonstrated to have paralytic activity in the parasitic spider wasp [41], and Calreticulin, which has been shown to inhibit hemocyte spreading and encapsulation in a butterfly wasp endoparasitoid [39]. However, we found that L. heterotoma has approximately 25% more venom genes than does L. boulardi (Table 2), twelve of which are GTPases (Tables 5 and S1), and these extra venom proteins may enable L. heterotoma to infect a greater diversity of host species than L. boulardi [13] and may also be responsible for their differences in infection characteristics in their common natural host, D. melanogaster [15–24]. Altogether, our data suggest that there are a common set of venom protein functionalities that most parasitic wasps tend to use, but that even closely related species such as L. boulardi and L. heterotoma can evolve unique venom activities suited to their particular niches.

**Discussion**

Our approach integrated high-throughput transcriptomic and proteomic data to investigate the venoms of two endoparasitoid wasp species, L. boulardi and L. heterotoma. We first assembled the transcriptomic data and then filtered the assembly based on expression levels; next, we used the filtered assembly as a library for mapping mass spectrometry data to identify a list of putative venom proteins. The putative venom transcripts we identified included the three previously characterized L. boulardi venom proteins, and we further validated our venom gene identification approach using experimental and computational techniques. RT-PCR results showed female- and venom gland-specific transcription of putative venom genes, the putative venom proteins from L. boulardi and L. heterotoma showed significant homology to each other and to other known Hymenopteran venom proteins, and the putative venom proteins were significantly more likely to contain secretion signal sequences.

Our venom identification approach yielded many more proteins than lower throughput methods previously used for analyzing wasp venoms. So far, traditional cloning methods have only identified 4 venom proteins from Drosophila endoparasitoids [19,20,25–27], but we identified 304 potential venom proteins in two wasp species. Our RT-PCR results showed that 19 of 20 putative venom genes were expressed in venom glands (often specifically in venom glands or females), suggesting a true positive rate of ~95%, so by extrapolation we conclude that our approach yielded about 122 venom genes for L. boulardi and 167 venom genes for L. heterotoma. Our venom identification method is applicable to general protein identification in any tissue sample or cell type, but will likely be most useful for subsets of proteins that are spatially sequestered. Whereas transcriptomics alone is likely sufficient for general protein identification in particular tissues, if subsets of proteins are sought, and the protein subsets can be isolated (such as in an organelle or gland), a joint transcriptomic and proteomics approach is ideal for specifically identifying those protein subsets.

Our analysis of venoms from the D. melanogaster wasp endoparasitoids L. boulardi and L. heterotoma provides a foundation for future functional studies of the infection strategies of these wasps. For example, we have identified 9 novel homologs of known L. boulardi venom genes across both wasp species (Fig. 5), we have identified numerous L. boulardi and L. heterotoma protein types that are commonly found in venoms of other Hymenopterans (Table 8 and 9), and we have identified a potentially novel type of venom genes, an expanded family of GTPases, in L. heterotoma. Learning how Drosophila parasites evade and suppress the cellular encapsulation response will further our understanding of innate immunity in this model organism, especially the weak links in innate immunity that parasites tend to exploit. The difference in venom contents and effects between the two closely related wasps L. boulardi and L. heterotoma also suggests that venom identification in further species, genera, and families of Drosophila endoparasitoids will continue to provide novel material for elucidating patterns of parasite virulence evolution.
Integrative Approach Identifies Wasp Venoms

A

E-value ($10^0$) tBLASTx vs. *L. heterotoma* venom

B

E-value ($10^0$) tBLASTx vs. *L. heterotoma* body

C

L. *boulardi* venom vs. body sequences:

- *L. heterotoma* venom: $\chi^2 = 2826.633$, p < 0.001
- *L. heterotoma* body: $\chi^2 = 21.9267$, p = 0.015
- *L. heterotoma* all: $\chi^2 = 37.4445$, p < 0.001

D

E-value ($10^0$) tBLASTx vs. all *L. heterotoma*

E

E-value ($10^0$) tBLASTx vs. *L. boulardi* body

F

L. *heterotoma* venom vs. body sequences:

- *L. boulardi* venom: $\chi^2 = 2707.669$, p < 0.001
- *L. boulardi* body: $\chi^2 = 38.4963$, p < 0.001
- *L. boulardi* all: $\chi^2 = 60.2229$, p < 0.001
Blasted against L. boulardi is isolation of 840 ng of (Invitrogen) according to manufacturer's directions. This led to RNAs were purified using the Dynabeads mRNA Direct kit EDTA, 5 mM DTT, and 1% LiDS (pH 7.5) solution. Poly(A) resuspended in a 100 mM Tris-HCl, 500 mM LiCl, 10 mM u

Methods

Wasp Strains

L. boulardi strain Lh17 and L. heterotoma strain Lh14 were collected in Winters, California in 2002 and have been previously described [15]. Laboratory cultures of these wasps were maintained on D. melanogaster and are available upon request.

mRNA Isolation and CDNA Preparation

Abdomens of approximately 200 female wasps aged 2–7 days were dissected into Trizol reagent (Invitrogen) and incubated at 4°C for 24 hrs. Total RNA was isolated according to manufacturer's directions with the exception that RNA was mixed with 5 µg glycogen (Invitrogen) prior to precipitation with 1 ml isopropanol per 1 ml Trizol overnight at −20°C. RNA was resuspended in a 100 mM Tris-HCl, 500 mM LiCl, 10 mM EDTA, 5 mM DTT, and 1% LiDS (pH 7.5) solution. Poly(A) RNAs were purified using the Dynabeads mRNA Direct kit (Invitrogen) according to manufacturer's directions. This led to isolation of 840 ng of L. boulardi mRNA (from 25.2 µg of total RNA, 3.3% of total) and 1.2 µg of L. heterotoma mRNA (from 32.3 µg of total RNA, 3.7% of total).

cDNAs were synthesized from the isolated wasp mRNAs using the SuperScript II double-stranded cDNA synthesis kit (Invitrogen) according to manufacturer's directions with the following exceptions. First-strand synthesis was primed by random hexamers (Invitrogen) and the SuperScript II reverse transcription reaction was incubated at 42°C for 1 hr. Following second-strand synthesis, cDNA samples were incubated with 1 µl of 10 mg/ml RNase A (Novagen) at 37°C for 1 hr. The cDNA samples were then extracted with phenol:chloroform and precipitated with 7.5 M ammonium acetate and ethanol. Pellets were washed three times with 70% ethanol and resuspended in 1X TE buffer.

Venom Protein Isolation

Venom glands from 50 female wasps were dissected into 1X PBS supplemented with 0.5 mM EDTA and Complete Protease Inhibitor Cocktail (Roche) on ice. Tissues were then homogenized under nonlysing conditions and incubated on ice for a further 5 min. Cells were pelleted by centrifugation at 12000 g at 4°C and venom containing supernatant was removed and combined with 2X sample buffer (4% SDS, 20% glycerol, 120 mM Tris pH 6.8, 0.002% bromophenol blue, 200 mM DTT) for proteomic analysis.

Tissue Specific RT-PCR

Primers were designed in Primer3 (SDSC Workbench) to amplify control and putative venom genes. Control primers were made to target wasp homologs of RNA polymerase II, elav, Histone2A, and collagen IV, and venom primers were made against 10 putative venom genes identified by mass spectrometry of purified venom. Primers used are listed in Tables S2 and S3.

To make tissue specific cDNAs, venom glands from 25 female wasps were dissected in PBS and the venom glands and venom gland-less carcasses were incubated in Trizol reagent for 24 hrs at 4°C. Total RNA was isolated as described above, and reverse transcribed using the Quantitect reverse transcription kit (Qiagen). Targeted venom and control genes were then amplified from these tissue specific cDNAs. Amplification conditions for venom primers were 95°C for 3 min, 25 cycles of 95°C for 15 sec, 55°C for 30 sec, 72°C for 1 min before a final incubation at 72°C for 1 min, and were the same for control primers except 35 cycles of amplification were used. Amplification products were analyzed on 1% agarose gels.

Blast Searches

Blast searches were performed using the command line NCBI-BLAST package (version 2.2.5) [55]. The Hymenopteran venom database was assembled from proteins annotated as 'venom' and restricted to Hymenoptera on GenBank (as of 05/06/2012). Venom and body transcript subsets were both aligned to the venom database using the BLASTx algorithm with default parameters except that the 'max target seqs' parameter was set to 1 to get only the single best blast hit. L. boulardi and L. heterotoma

Figure 5. Novel homologs of known L. boulardi venom proteins in L. boulardi and L. heterotoma. (A) % identity and % similarity of novel homologs of the venom LbGAP in L. boulardi and L. heterotoma venom within the RhoGAP domain (green box) as indicated. (B) % identity and % similarity of novel homologs of the venom Serpin in L. boulardi and L. heterotoma venom within the serpin domain (blue box) as indicated. Signal peptides are shown in red. doi:10.1371/journal.pone.0064125.g005
Table 8. *L. boulardi* venom proteins showing homology to known venom proteins.

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Sequence ID</th>
<th>Best blast hit (species)</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known <em>L. boulardi</em> venom proteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extracellular superoxide dismutase 3</td>
<td>comp937_c0_seq1</td>
<td><em>L. boulardi</em></td>
<td>9.00E-99</td>
</tr>
<tr>
<td>RhoGAP</td>
<td>comp433_c0_seq1</td>
<td><em>L. boulardi</em></td>
<td>2.00E-126</td>
</tr>
<tr>
<td>Serpin</td>
<td>comp138_c0_seq1</td>
<td><em>L. boulardi</em></td>
<td>3.00E-155</td>
</tr>
<tr>
<td>Other venom proteins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RhoGAP</td>
<td>comp233_c0_seq1, comp516_c0_seq1, comp3196_c0_seq1, comp779_c0_seq1</td>
<td><em>L. boulardi</em></td>
<td>5.00E-46</td>
</tr>
<tr>
<td>Serpin</td>
<td>comp552_c0_seq2</td>
<td><em>L. boulardi</em></td>
<td>8.00E-66</td>
</tr>
<tr>
<td>Arginine kinase</td>
<td>comp500_c0_seq1</td>
<td>Cyphononyx dorsalis</td>
<td>0.0</td>
</tr>
<tr>
<td>Calreticulin</td>
<td>comp263_c0_seq1</td>
<td>Nasonia vitripennis</td>
<td>2.00E-157</td>
</tr>
<tr>
<td>Cysteine-rich venom protein 6</td>
<td>comp242_c2_seq28</td>
<td>Pimpla hypochondriaca</td>
<td>9.00E-09</td>
</tr>
<tr>
<td>Glucose dehydrogenase venom protein</td>
<td>comp2100_c0_seq1, comp2409_c0_seq1</td>
<td>Nasonia vitripennis</td>
<td>4.00E-98</td>
</tr>
<tr>
<td>Inosine-uridine preferring nucleoside hydrolase</td>
<td>comp3592_c0_seq1</td>
<td>Nasonia vitripennis</td>
<td>4.00E-57</td>
</tr>
<tr>
<td>Lipase A</td>
<td>comp1645_c0_seq14, comp1645_c0_seq15</td>
<td>Nasonia vitripennis</td>
<td>1.00E-35</td>
</tr>
<tr>
<td>Tyrosine 3/tryptophan 5-monooxygenase</td>
<td>comp31_c0_seq1</td>
<td>Eumenes pomiformis</td>
<td>1.00E-71</td>
</tr>
<tr>
<td>Venom metalloprotease</td>
<td>comp3680_c0_seq1</td>
<td>Euolopus pennicornis</td>
<td>4.00E-11</td>
</tr>
<tr>
<td>Venom acid phosphatase</td>
<td>comp9544_c0_seq1</td>
<td>Bombus impatiens</td>
<td>1.00E-69</td>
</tr>
<tr>
<td>Venom chitinase</td>
<td>comp435_c0_seq1</td>
<td>Chelonus inanitus</td>
<td>1.00E-111</td>
</tr>
<tr>
<td>Venom protein 1</td>
<td>comp6910_c0_seq1</td>
<td>Microctonus hyperodae</td>
<td>6.00E-15</td>
</tr>
<tr>
<td>Venom protein 8</td>
<td>comp953_c0_seq1</td>
<td>Microctonus hyperodae</td>
<td>3.00E-14</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0064125.t008

Table 9. *L. heterotoma* venom proteins showing homology to known venom proteins.

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Sequence ID</th>
<th>Best blast hit (species)</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arginine kinase</td>
<td>comp1824_c0_seq1</td>
<td>Cyphononyx dorsalis</td>
<td>0.0</td>
</tr>
<tr>
<td>Aspartylglucosaminidase</td>
<td>comp18_c0_seq1, comp6288_c0_seq1</td>
<td>Nasonia vitripennis</td>
<td>3.00E-93</td>
</tr>
<tr>
<td>Calreticulin</td>
<td>comp326_c0_seq1</td>
<td>Nasonia vitripennis</td>
<td>7.00E-153</td>
</tr>
<tr>
<td>Glucose dehydrogenase venom protein</td>
<td>comp4489_c2_seq4</td>
<td>Nasonia vitripennis</td>
<td>9.00E-121</td>
</tr>
<tr>
<td>Inosine-uridine preferring nucleoside hydrolase</td>
<td>comp3755_c0_seq1</td>
<td>Nasonia vitripennis</td>
<td>9.00E-64</td>
</tr>
<tr>
<td>Lipase A</td>
<td>comp2257_c0_seq1</td>
<td>Nasonia vitripennis</td>
<td>7.00E-34</td>
</tr>
<tr>
<td>Metalloprotease</td>
<td>comp1346_c0_seq1</td>
<td>Euolopus pennicornis</td>
<td>9.00E-20</td>
</tr>
<tr>
<td>Metalloproteinase</td>
<td>comp3216_c0_seq1</td>
<td>Euolopus pennicornis</td>
<td>9.00E-20</td>
</tr>
<tr>
<td>Tyrosine 3/tryptophan 5-monooxygenase</td>
<td>comp455_c2_seq1</td>
<td>Eumenes pomiformis</td>
<td>3.00E-71</td>
</tr>
<tr>
<td>RhoGAP</td>
<td>comp155_c0_seq1, comp155_c0_seq3, comp1758_c1_seq3</td>
<td><em>L. boulardi</em></td>
<td>1.00E-25</td>
</tr>
<tr>
<td>Venom acid phosphatase</td>
<td>comp1442_c0_seq1, comp1442_c0_seq2, comp1602_c1_seq12, comp2614_c0_seq1, comp2636_c0_seq1, comp3064_c0_seq1</td>
<td>Nasonia vitripennis</td>
<td>5.00E-78</td>
</tr>
<tr>
<td>Venom allergen 3</td>
<td>comp1331_c0_seq1</td>
<td>Nasonia vitripennis</td>
<td>3.00E-37</td>
</tr>
<tr>
<td>Venom carboxylesterase-6</td>
<td>comp255_c0_seq1</td>
<td>Apis florea</td>
<td>9.00E-118</td>
</tr>
<tr>
<td>Venom protein 1</td>
<td>comp1040_c0_seq26</td>
<td>Microctonus hyperodae</td>
<td>5.00E-54</td>
</tr>
<tr>
<td>Venom protein 5</td>
<td>comp1564_c0_seq1</td>
<td>Brachymyrmex patagonicus</td>
<td>3.00E-51</td>
</tr>
<tr>
<td>Venom protein 8</td>
<td>comp719_c0_seq1</td>
<td>Microctonus hyperodae</td>
<td>4.00E-14</td>
</tr>
<tr>
<td>Venom protein Ci-95</td>
<td>comp128_c2_seq1</td>
<td>Chelonus inanitus</td>
<td>1.00E-28</td>
</tr>
<tr>
<td>Venom serine protease</td>
<td>comp513_c0_seq1</td>
<td>Nasonia vitripennis</td>
<td>4.00E-117</td>
</tr>
<tr>
<td>Serpin</td>
<td>comp224_c1_seq1</td>
<td><em>L. boulardi</em></td>
<td>2.00E-62</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0064125.t009
transcriptomes and transcript subsets were compared using the tBLASTn algorithm with default parameters except that the ‘max target seq’ parameter was again set to 1. The distributions the single best-hit E-values of venom and body transcript subsets blasted to a common target database were compared to each other by Chi square tests in R (version 2.15.0).

1D PAGE Molecular Weight Fractionation and in-gel Tryptic Digestion

Protein extracts were quantified with a BCA Protein Assay Kit (Thermo Scientific). 10 µg protein aliquots from each wasp species were denatured and reduced using Novex Tris-Glycine SDS sample buffer, and separated by molecular weight on a 4–12% Novex Tris-Glycine gel at a constant 125 V for 45 minutes. Gels were stained overnight with a Colloidal Blue Staining Kit at room temperature with gentle agitation, and destained in MilliQ water for more than 3 hrs at room temperature the following day. Each lane was excised into eight molecular weight fractions containing nearly equal staining densities across each band, and in-gel tryptic digestion was carried out with Trypsin Gold following the manufacturer instructions. Digests were reduced in volume to near dryness in a SpeedVac concentrator (Savant) and brought up to 20 µL using a 5% acetonitrile and 0.1% formic acid solution.

Nano LC-MS²

Peptide digests (10 µL = ~300 ng, controls calculated at ~50% efficiency in digestion and peptide recovery) were injected two separate times onto a Surveyor HPLC plus machine (Thermo Scientific, San Jose CA) using a split flow configuration on the back end of a 100 µm internal diameter × 13 cm pulled tip C18 column (Jupiter C-18 300 Å, 5 µm, Phenomenex). This system was run in-line with a Thermo LTQ XL ion trap mass spectrometer equipped with a nano-electrospray source (Thermo Scientific, San Jose CA), and all data were collected in CID mode. Buffers used were solvent A (0.1% formic acid in ddH2O) and solvent B (0.1% formic acid in acetonitrile). The HPLC gradient setting was as follows; 0–10 min using 2% solvent B (98% solvent A), 10–15 min using 12% solvent B, 15–31 min using 30% solvent B, 31–36 min using 36% solvent B, 36–41 min using 45% solvent B, 41–48 min using 63% solvent B, 48–56 min using 95% solvent B, 56–57 min using 2% solvent B, 57–82 min using 2% solvent B. The LTQ collected data on the top 3 ions using a triple play method from 10 to 57 minutes, with a flow rate of 0.6 µL/min. Before and after the analysis, the spray voltage was 0.0 kV and the flow rate was 3 µl/min. During data collection, the LTQ XL was configured as follows: spray voltage 2.2 kV, capillary temperature 170 °C, 1 microscan with a maximum inject time of 25 ms for all modes. For the triple play setup, the SIM scans were obtained in zoom mode using a ±2.5 Da window, and triggered on a signal threshold of 15,000 counts. The MS/MS scans were obtained in normal mode with a minimum signal threshold of 500 counts based on the SIM scan. The activation settings were charge state 5, isolation width 5 m/z, normalized collision energy 30.0, activation Q 0.250, and activation time 50.00 ms. For the dependent scans, charge state screening was enabled, and the dynamic exclusion was enabled with the following settings: repeat count 3, repeat duration 20.0 s, exclusion list size 500, and exclusion duration 90.0 s. XCalibur RAW files outputted from the mass spectrometer were collected in profile mode, centroided and converted to MzXML using ReAdW v. 3.5.1 (Institute for Systems Biology). mgf files were then created using MzXML2Search (included in TPP v. 3.5, Institute for Systems Biology) for all scans with a precursor mass between 400 Da and 20,000 Da.

Protein Identifications and Quality Control

The libraries of female wasp abdomen transcripts were translated in each of three frames in both orientations for each transcript sequence, for a total of six amino acid sequences per mRNA sequence, using the EMBOSS getorf tool, version 6.5.0. The peptide sequence data were then queried against a database containing these translated wasp abdomen transcripts along with common contaminant sequences such as digestion enzymes and human keratin, using SEQUEST (v.27 rev 12,.dta files). The program was set for “no-enzyme”, a precursor mass window of 0.45 Da, trypsin digestion, static modification C at 57.0293, variable modification M at 15.9949, and a fragment-ion mass tolerance of 0.8 Da. The SEQUEST results files were first combined for each of 8 fractions per lane, then filtered using peptide and protein probability [56], grouped by top scoring protein ID, and finally quantified by normalized spectral counts (label free) using ProteoIQ (NuSep, Athens, GA) [57,58]. The filter cut-off values were set with peptide length (>4 AA), peptide probability (>0.5), peptides per protein (≥2 peptides), and protein probability (≥0.7), and peptides with +1 charge were also excluded from any further analyses. This protocol resulted in a list of venom protein identifications with ≥99.0% confidence.

Gene Ontology Analysis of Venom Proteins

For all expressed transcripts, the longest open reading frame was found using the EMBOSS getorf tool, version 6.5.0, annotated with PFAM protein families using HMMer version 3.0, and then PFAM annotations were converted to corresponding GO categories using InterPro2GOMapping [59], version date 08/11/2012. The open source GOA tools (https://github.com/tanghaibo/goatools/) were used for the enrichment analysis. Results obtained are shown in Table S1.

Data Access

All mRNA-seq data is available at the DDBJ Sequence Read Archive [http://trace.ddbj.nig.ac.jp/dra/index_e.shtml] under accession number SRA058494. The assembled wasp transcripts were deposited in GenBank under Transcriptome Shotgun Assembly accession numbers GJA00000000 and GJC00000000. All raw mass spectrometry data is available at http://www.taylorlab.org/~jeremy/wasp-venom-proteomic-data/. Code bundle at https://bitbucket.org/bxlab/wasp-venom provides data, methods, and original results for transcriptome assembly and quantification, secretion signal sequence analysis, and functional annotation and enrichment.

Supporting Information

Table S1 Results from gene ontology analysis. Dots preceding category ID indicate category level relative to top-level category, GO:0003674, molecular function. pBonferroni <0.05 (DOCX)

Table S2 PCR primers used to amplify L. boulardi genes. (DOCX)

Table S3 PCR primers used to amplify L. heterotoma genes. (DOCX)

Dataset S1 All identified L. boulardi venom cDNAs in fasta format. (FASTA)
Dataset S2 All identified L. heterotoma venom cDNAs in Fasta format. (FASTA)

Dataset S3 All identified L. boulardi venom proteins in Fasta format. (FASTA)

Dataset S4 All identified L. heterotoma venom proteins in Fasta format. (FASTA)

References


Acknowledgments

We would like to thank Aurore Dubuffet and two anonymous reviewers for comments on the manuscript.

Author Contributions

Conceived and designed the experiments: JG NTM JT TAS. Performed the experiments: JG NTM JAM GBJ. Analyzed the data: JG NTM JAM GBJ JT TAS. Contributed reagents/materials/analysis tools: JAM JT TAS. Wrote the paper: JG NTM JAM JT TAS.
ized by *Cotesia flavipes* (Cameron) (Hymenoptera: Braconidae). Biological Control 45: 103–110.


