

# Effects of Gene Duplication, Positive Selection, and Shifts in Gene Expression on the Evolution of the Venom Gland Transcriptome in Widow Spiders

Robert A. Haney,<sup>1,\*</sup> Thomas H. Clarke,<sup>2,3</sup> Rujuta Gadgil,<sup>1</sup> Ryan Fitzpatrick,<sup>1</sup> Cheryl Y. Hayashi,<sup>3</sup> Nadia A. Ayoub,<sup>2</sup> and Jessica E. Garb<sup>1,\*</sup>

<sup>1</sup>Department of Biological Sciences, University of Massachusetts, Lowell

<sup>2</sup>Department of Biology, Washington and Lee University

<sup>3</sup>Department of Biology, University of California, Riverside

\*Corresponding author: E-mail: robert\_haney@uml.edu; jessica\_garb@uml.edu.

Associate editor: Jay Storz

Accepted: December 15, 2015

**Data deposition:** Final transcriptomes are available at NCBI's Transcript Shotgun Assembly database under accession numbers GBJM00000000, GBJN00000000, and GBJQ00000000. Raw sequencing reads are available at NCBI's Short Read Archive under accession numbers SRR1539523, SRR1539569, and SRR1539570. Venom gland ESTs from *Steatoda grossa* and *Latrodectus geometricus* were deposited at dbEST under accession numbers JZ875857–JZ876560.

## Abstract

Gene duplication and positive selection can be important determinants of the evolution of venom, a protein-rich secretion used in prey capture and defense. In a typical model of venom evolution, gene duplicates switch to venom gland expression and change function under the action of positive selection, which together with further duplication produces large gene families encoding diverse toxins. Although these processes have been demonstrated for individual toxin families, high-throughput multitissue sequencing of closely related venomous species can provide insights into evolutionary dynamics at the scale of the entire venom gland transcriptome. By assembling and analyzing multitissue transcriptomes from the Western black widow spider and two closely related species with distinct venom toxicity phenotypes, we do not find that gene duplication and duplicate retention is greater in gene families with venom gland biased expression in comparison with broadly expressed families. Positive selection has acted on some venom toxin families, but does not appear to be in excess for families with venom gland biased expression. Moreover, we find 309 distinct gene families that have single transcripts with venom gland biased expression, suggesting that the switching of genes to venom gland expression in numerous unrelated gene families has been a dominant mode of evolution. We also find ample variation in protein sequences of venom gland-specific transcripts, lineage-specific family sizes, and ortholog expression among species. This variation might contribute to the variable venom toxicity of these species.

**Key words:** venomics, comparative transcriptomics, black widow, gene family evolution, *Latrodectus*.

## Introduction

Venom, which has functional importance for prey capture, predator defense, and competitor deterrence, is a biochemical adaptation that has evolved multiple times in select lineages of animals (Fry et al. 2009; Casewell et al. 2013). Venom consists of a complex mixture of proteins, peptides, and small molecules, many of which disrupt the nervous system of target organisms (King and Hardy 2013). As venoms are largely protein-based secretions, there is a relatively straightforward

connection between expression of genes in the venom gland and ecologically important venom phenotypes, which has made venoms attractive systems for molecular evolutionary studies (Duda and Palumbi 1999; Kordis and Gubensek 2000; Fry 2005).

The application of high-throughput sequencing technologies to individual spider species has demonstrated the expression of numerous transcripts in spider venom glands with homology to known toxin types (Zhang et al. 2010;

He et al. 2013). Given the lack of comparative data, however, little is known about how venom gland gene expression varies among species. Comparative data on venom gland expression can shed light on mechanisms of venom gene evolution. In the predominant model, a duplicate of a single gene expressed outside of venom gland tissue shifts its expression to the venom gland. Subsequent functional diversification of the venom gland expressed gene occurs through the action of positive selection (Kordis and Gubensek 2000; Fry et al. 2009; Wong and Belov 2012). After several rounds of gene duplication and divergence, a large gene family is generated that encodes functionally diverse toxins with a pattern of venom gland biased expression. This simple model involves a single irreversible recruitment to venom gland expression by a duplicated gene, generally encoding a limited set of protein types that appear to contribute to the toxic constituents of venom across Metazoa (Fry et al. 2009). Yet, recruitment has been found to occur in the opposite direction as well, with families of toxin genes with venom gland biased expression producing genes that revert to expression outside the venom gland, presumably acquiring a nontoxic physiological role (Casewell et al. 2012). Dynamism in the cohort of genes expressed in the venom gland could also allow for frequent recruitments to venom gland biased expression from a broad suite of gene families and functional categories, resulting in gene families with only one or a few members exhibiting venom gland-specific expression. These genes could then provide raw material for the development of new toxins, particularly if they are secreted proteins, resulting in a wider spectrum of proteins contributing to venom than previously proposed (Fry et al. 2009). However, not all venom gland biased genes will become toxins; genes with venom gland biased expression could be involved in other aspects of venom gland physiology, including toxin production.

The ecological and evolutionary mechanisms that have shaped venom gland expression and venom protein composition still require testing with more extensive comparative data on tissue-specific expression. In this study, we focus on three spider species: 1) The Western black widow *Latrodectus hesperus*, 2) the congeneric brown widow (*Latrodectus geometricus*), and 3) a member of a putative sister genus (Arnedo et al. 2004) to *Latrodectus* (*Steatoda grossa*). We use high-throughput sequencing of cDNA (RNA-Seq) to identify venom gland-specific transcripts (VSTs): Transcripts with venom gland exclusive transcription or with expression strongly biased toward the venom gland (Haney et al. 2014) in each species. This venom gland-specific transcriptome should be enriched for venom toxins, both known and novel, and provide key insights into venom gland functionality. We use our widow spider venom gland transcriptomes to test specific hypotheses. We assess the importance of gene duplication in venom evolution by comparing the sizes of gene families containing venom gland specifically expressed transcripts with the sizes of control gene families containing only broadly

expressed transcripts. We also test the widely held view that positive selection is a dominant process driving protein sequence divergence after a shift to venom gland biased expression (Duda and Palumbi 1999; Wong and Belov 2012; Casewell et al. 2013).

Closely related venomous species can differ in toxicity to specific prey items or potential predators (Sanz et al. 2006; Duda and Remigio 2008). The clades to which our focal species belong present distinct venom toxicity phenotypes. Specifically, experimental and clinical data indicate that venoms from species in the “mactans” clade of *Latrodectus* (containing all “black widow” species, including *L. hesperus*) are more toxic to vertebrates and cause a greater amount of uncontrolled neurotransmitter release compared with venom from brown widows (*L. geometricus*) (Muller et al. 1989, 1992; Isbister and Gray 2003; Vetter and Isbister 2008; Isbister and Fan 2011). Members of the genus *Steatoda*, including *S. grossa*, generally have venoms that are far less toxic to vertebrates than are venoms from *Latrodectus* species (Maretic et al. 1964; Cavalieri et al. 1987; Warrell et al. 1991; Muller et al. 1992; Graudins et al. 2012). However, experimental comparisons indicate that venom from some *Steatoda* species may be more toxic to some insects (Cavalieri et al. 1987; Atakuziev et al. 2014). The basis for interspecific variation in toxicity is venom composition, and thus at least in part is related to underlying species differences in gene expression in the venom gland, as well as to differences in the coding regions of homologous venom gland expressed genes. Hence we explore variability among species in the types and functions of transcripts with venom gland biased expression that could contribute to differences in venom toxicity. We explore potential mechanisms for toxicity differences including evolutionary shifts to venom gland expression in functionally diverse gene families, venom gland biased species-specific families and singleton sequences, and transcripts with species-specific upregulation exclusive to the venom gland.

## Materials and Methods

### Transcriptome Assembly and Annotation

Adult female *L. hesperus* were collected in Riverside, CA, in March 2009, July 2010, and August 2011. Adult female *L. geometricus* were collected in San Diego, CA, in July 2010 and Riverside, CA, in August 2011, while adult female *S. grossa* were obtained from Spider Pharm (Yarnell, AZ) in July 2010 and August 2011. Spiders were kept in the laboratory for 3–7 days and fed one cricket 1–5 days prior to dissection. cDNA libraries were prepared using standard Illumina RNA-Seq protocols for 2–3 replicates from multiple tissues including venom gland, ovary, and cephalothorax tissue from all three species, and paired-end Illumina sequencing (100-bp reads) was performed. Trinity assemblies (Grabherr et al.

2011) were performed separately for each tissue in each species, and for each Trinity component only the longest sequence was retained. BLAT (Kent 2002) was used to compare sequences across pruned tissue assemblies within species, and transcripts that were at least 98% identical over the full match length were grouped and each group was assembled into a contig with CAP3 (Huang and Madan 1999). Contigs and singletons resulting from CAP3 assembly were retained to produce a nonredundant draft transcriptome for each species (Clarke et al. 2015). Potential bacterial and cross-species contaminants were purged along with potential chimeric transcripts as described in Clarke et al. (2015) to produce a final reference transcriptome for each species. Additionally, a total of 416 expressed sequence tag (EST) sequences were produced from venom gland cDNA from *L. geometricus* and 567 from *S. grossa* using protocols described in McCowan and Garb (2014), and EST contigs were assembled using CAP3 and compared with the reference transcriptome using BLAT.

Transcriptomes were annotated by identification of BLASTX alignments of each transcript to proteins in the National Center for Biotechnology Information (NCBI) *nr* and UniProt databases with a cutoff *e*-value of  $1 \times 10^{-5}$ , allowing for a maximum of ten target sequences. Text searches of tabulated BLAST results were implemented to identify transcripts with homology to known widow spider venom components such as toxins and enzymes. GetORF was used to predict the most plausible open reading frame (ORF) for each transcript, selecting the longest ORF in the same frame as the best BLAST hit, or the longest ORF in the absence of BLAST homology. However, if a predicted protein possessed a methionine start codon and was 5' and 3' bounded by stop codons, it was retained as long as it was at least 75% of the length of the longest ORF (Clarke et al. 2015). InterProScan was used to identify the domain architecture of predicted gene products using the Pfam database. SignalP (Petersen et al. 2011) was used to identify signal peptides in predicted proteins using a two-pass methodology. Signal peptides were predicted from proteins translated as described above, after which proteins were trimmed to the first methionine and the signal peptide prediction process repeated, pooling the results of both passes to account for uncertainty in start codon prediction.

### Identifying Venom Gland Specific and Broadly Expressed Transcripts

In this study, reads from replicates of three individual tissue libraries (venom gland, cephalothorax, ovaries) were mapped back to each species reference transcriptome that was produced in Clarke et al. (2015) for determination of transcript abundance using RSEM (RNA-Seq by Expectation-Maximization) (Li and Dewey 2011). Using edgeR (Robinson et al. 2009) transcripts that did not reach a threshold of at least 1 count per million (CPM) in at least two libraries were removed, and differentially expressed transcripts among tissues

were identified. Specifically, we identified the set of transcripts with highly biased venom gland expression in each species. We performed pairwise comparisons to identify transcripts with significant upregulation in venom gland relative to cephalothorax and ovary tissue at a false discovery rate (FDR) of 0.05. The intersection of these two sets of transcripts with significantly venom gland biased expression is referred to as the set of VSTs for each species. Similarly, we defined a broadly expressed set of transcripts as the intersection of sets of transcripts that were not differentially expressed in any pairwise comparison between tissues in each species. Gene ontology (GO) terms for the top UniProt hit were acquired from UniProt-GoA. GOseq analysis was performed to find functional categories overrepresented among transcripts with venom gland biased expression while accounting for the influence of transcript length in the detection of differential expression (Young et al. 2010).

### Defining Tissue-Specific Gene Families across Species

We defined putative gene families as sets of nonredundant transcripts that form homologous groups among the three species' predicted proteomes using OrthoMCL (Li et al. 2003). This approach takes a set of all-by-all BLASTP results with an *e*-value cutoff of  $1 \times 10^{-5}$  and applies a Markov clustering algorithm to define groups of between-species orthologs and within-species paralogs derived from postspeciation duplications using a reciprocal best-hit criterion. Families are numbered in the order output by OrthoMCL and prefixed with the first four letters of the family name (Theridiidae; e.g., Ther1). We used our tissue-specific expression data to identify three sets of families: 1) VGE families whose member transcripts exclusively exhibited venom gland-specific expression (i.e., all members were VSTs); 2) Venom gland containing (VGC) families whose members included at least one VST but were not exclusively composed of VSTs; and (3) nondifferentially expressed (nonDE) families whose members exclusively exhibited broad expression, to act as a control group for the effects of transitions to venom gland biased expression.

We explored these families to identify species specificity and to test whether gene duplication and retention of functionally diversified duplicates has led to increased family size as expected under a traditional model of venom evolution by comparing VGE and VGC families with nonDE families as a control. We also explored patterns of gains and losses of venom gland biased expression in VGC families, including whether duplications followed a shift to venom gland biased expression. For each of the 36 largest putative VGC gene families, the predicted proteins were aligned with MUSCLE (Edgar 2004). Family phylogenies were inferred using Bayesian analysis of protein alignments in MrBayes 3.2.2 (Ronquist et al. 2012) sampling across fixed amino acid matrices. Analyses were run as two independent sets of 1–12 Markov chain Monte Carlo (MCMC) chains for 1–10 million generations, with sampling

every 500 generations. These run lengths were sufficient to achieve a standard deviation of split frequencies of  $\leq 0.01$ , and potential scale reduction factors of  $\sim 1$  for all parameters, indicating convergence and a sufficient sample from the posterior probability distribution. Consensus trees were built after discarding the first 25% of sampled trees from each run and midpoint rooted. We used SIMMAP 1.5 (Bollback 2006) to map ancestral expression states (venom gland biased or not venom gland biased) on Bayesian phylogenies, using best-fit priors for rate and bias values determined by MCMC analysis.

### Testing for Positive Selection

To gauge selective pressure at the protein level and to test whether selection varied among families based on expression pattern, the ratio of nonsynonymous to synonymous substitution rates ( $dN/dS$ ) averaged across sites and lineages for orthoMCL families in the three categories described above was calculated from model M0 with PAML 4 (Yang 2007) for all families with at least three members. We performed explicit tests of site-based diversifying positive selection in PAML 4 by comparing models M7 and M8 for individual families (Yang et al. 2000). Model M7 allows  $dN/dS$  to vary according to a beta distribution constrained to lie between 0 and 1. Model M8 allows an additional category of sites with unconstrained  $dN/dS$ . These models were compared by likelihood-ratio test with two degrees of freedom. If M8 was a significantly better fit to the data and the extra category of sites had an average  $dN/dS > 1$ , positive selection was inferred. For each VGE, VGC, and nonDE family, the amino acid alignment was used to guide a nucleotide alignment of the underlying coding sequences using PAL2NAL, with gaps removed from the final alignment (Suyama et al. 2006). This codon alignment was used as input for PAML analysis, together with the maximum-likelihood gene tree inferred from the protein alignment using FastTree 2.1.7 (Price et al. 2010) for comparisons of models M7 and M8. FastTree output was parsed for PAML compatibility using a custom Perl script. Only alignments with at least 30 codons after gap removal were retained for further analysis, and  $dN/dS$  values for which  $dS < 0.01$  were excluded from model 0 comparisons.

### Differential Expression of Orthologs along Species Lineages

We extracted the set of 1:1:1 orthologous (single copy in each species) families from the orthoMCL results to explore changes in gene expression along lineages. We focused on 1:1:1 orthologs to ascertain changes strictly related to speciation as opposed to a combined effect of duplication and speciation and to avoid potentially confounding effects of read mapping ambiguity among closely related paralogs. As transcript lengths can vary among species due to biological variation and/or assembly artifacts, length standardized fragments per kilobase of transcript per million mapped reads

(FPKM) values were used to test for differential expression among species. For each ortholog, fragments per kilobase of transcript per million mapped reads (FPKM) values were calculated separately for each venom gland, ovary, and cephalothorax RNA-Seq library in each species using RSEM. These values were used to determine transcripts with species-specific differential expression in the venom gland, as well as in the cephalothorax and ovary, using sRAP (Warden et al. 2013). FPKM values were rounded below the default cutoff and  $\log_2$  transformed, and differentially expressed transcripts among species within tissues were determined by Analysis of Variance (ANOVA) with significance controlled by enforcing an FDR of 0.05. The venom gland differentially expressed set was compared with that of the other two tissues to identify transcripts upregulated in the venom gland along a species lineage, but not in the other two tissues.

## Results

### Hundreds of VSTs in Widow Spider Species

The final species transcriptome assemblies contained 161,843 transcripts in *S. grossa*, 191,314 transcripts in *L. hesperus*, and 152,807 in *L. geometricus*. In *S. grossa*, 27,597 transcripts met the filtering criterion of at least 1 CPM in at least two of any of the venom gland, cephalothorax, and ovary libraries (table 1 and fig. 1). Of these 27,597, 411 were significantly upregulated in venom gland tissue relative to both cephalothorax and ovary and were considered VSTs. In *L. hesperus*, 30,255 transcripts met the filtering criterion and 377 were VSTs, and in *L. geometricus* 29,141 transcripts remained after filtering of which 597 were VSTs (supplementary table S1, [Supplementary Material](#) online). In general, transcripts assembled from short-read data

**Table 1**

Toxins, Enzymes, and Neural Function Proteins Found among VSTs

Species	<i>Steatoda grossa</i>	<i>Latrodectus hesperus</i>	<i>Latrodectus geometricus</i>	Families	VGEs	VGCs
>1 CPM	27,597	30,255	29,141			
VSTs	411	377	597		66	389
Latrotoxins	12	14	17		8	6
Latrodectins	6	5	4		0	6
ICK	3	4	2		1	1
CRISP	1	6	5		2	0
Metalloproteases	3	12	8		2	4
Serine proteases	3	5	8		2	6
Hyaluronidases	1	2	1		0	0
Lipases	3	4	4		1	4
Chitinases	1	4	3		1	0
Neural function <sup>a</sup>	8	16	9		3	5

NOTE.—Homology assignments were made by text search of BLAST results derived from querying transcriptomes against the nr and UniProt databases.

<sup>a</sup>Includes sequences with BLAST homology to leucine-rich repeat transmembrane neuronal 4, synaptic vesicle 2-related protein, regulating synaptic membrane exocytosis protein 2-like, synaptic vesicle glycoprotein 2B, and neural cell adhesion molecule L1.

were longer than their EST counterparts, and 90–95% of the assembled ESTs had a BLAT match in their respective species' transcriptome. In *L. geometricus*, 40 (19.9%) EST CAP3 output sequences (of 201 contigs and singletons) with a BLAT alignment were longer than their corresponding reference transcript, while 61 (19.3%) of the 316 CAP3 output sequences in *S. grossa* were longer than the corresponding transcript from the reference.

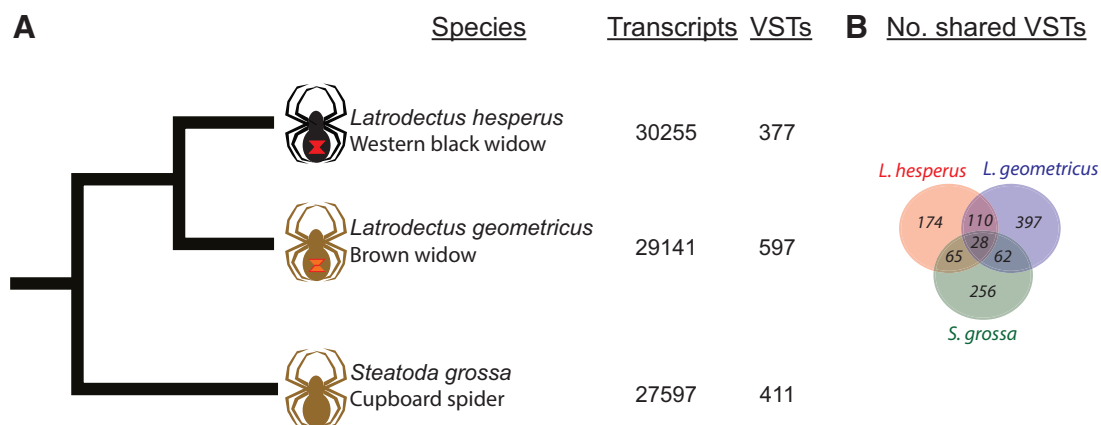
### VSTs Include Toxins Shared across Species and Are Enriched with Divergent Sequences

BLASTX analyses found that a minority of VSTs had a significant hit in each species (nr: *S. grossa* 121 of 411 [29.4%], *L. hesperus* 152 of 377 [40.3%], and *L. geometricus* 184 of 597 [30.8%]; UniProt results differed by no more than 1%; supplementary table S1, [Supplementary Material](#) online). The proportion of VSTs with BLAST hits was substantially lower than the proportion for the entire filtered transcriptome (nr: *S. grossa* 15,011 of 27,597 [54.4%], *L. hesperus* 17,677 of 30,255 [58.4%], and *L. geometricus* 14,749 of 29,141 [50.6%]; UniProt results differed by no more than 1%), suggesting an excess of novel or highly divergent transcripts in the venom gland transcriptomes.

Despite a low percentage of annotated VSTs, transcripts with BLAST hits to several known *Latrodectus* venom components were among the VSTs in each species, including latrotoxins, inhibitory cystine knot (ICK) toxins, cysteine-rich secretory proteins (CRISP), and latroductins ( $\alpha$ -latrotoxin-associated low molecular weight proteins), but the numbers per category varied among species (table 1 and supplementary table S1, [Supplementary Material](#) online). Thirty-five *L. hesperus* VSTs had a BLASTN hit at an e-value of 0.0 to

transcripts encoding the 52 proteins identified in *L. hesperus* venom from mass spectrometry analyses reported in Haney et al. (2014) and are highlighted in supplementary table S1, [Supplementary Material](#) online. Venom-associated enzymes (metalloproteases, serine proteases, lipases, chitinases, hyaluronidases) along with sequences with BLAST homology to nervous system-related proteins (leucine-rich repeat transmembrane neuronal 4, synaptic vesicle 2-related protein, regulating synaptic membrane exocytosis protein 2-like, synaptic vesicle glycoprotein 2B, neural cell adhesion molecule L1) were also in each species' VSTs (table 1 and supplementary table S1, [Supplementary Material](#) online). Although these neural function proteins are not known to function as toxins, they share venom gland-specific expression across the three species, and we have previously shown the presence of leucine-rich repeat transmembrane neuronal 4 in *L. hesperus* venom with mass spectrometry (Haney et al. 2014).

A minority of VSTs in each species had at least one protein domain prediction in the Pfam database (*S. grossa*: 89 (21.6%); *L. geometricus*: 147 (24.6%); *L. hesperus*: 121 (32.1%)). For each species, the proportion of VSTs with a domain prediction was substantially lower than the proportion for all transcripts above threshold expression (*S. grossa*: 12,651 (45.8%); *L. geometricus*: 12,302 (42.2%); *L. hesperus*: 15,049 (49.7%)). The top 20 most highly represented VST protein domains across all species included ankyrin repeat motifs, latrotoxin C-terminal domains, low-density lipoprotein receptor class A domains, and leucine-rich repeats. Other domains common among VSTs included M13 metalloprotease and serine protease domains, lipase domains, and cysteine-rich secretory family domains (supplementary table S2 and supplementary fig. S1, [Supplementary Material](#) online).



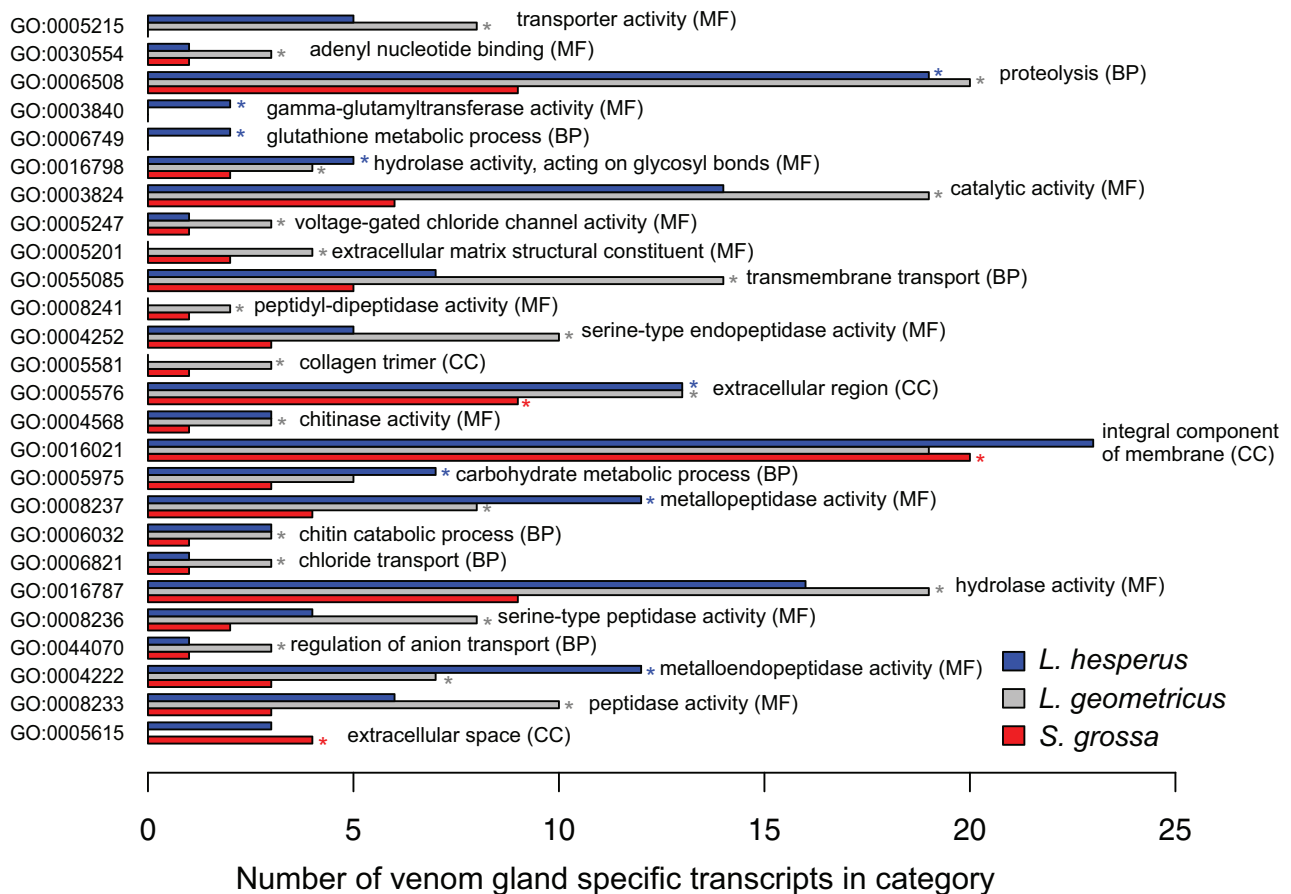
**FIG. 1.**—Phylogeny of species used in this study showing interspecific variation in number of VSTs. For each species, RNA-Seq was used to produce nonredundant sets of transcripts and to determine tissue-specific gene expression. (A) Column 2 shows the number of transcripts that met an expression threshold  $\geq 1$  read CPM, and Column 3 shows the number of transcripts that had exclusive or highly biased expression in the venom gland of each species (i.e., VSTs). (B) It shows the number of VSTs unique to each species and shared between species as a Venn diagram. Shared VSTs are defined as members of the same orthoMCL family that have venom gland biased expression in each species.

Signal peptides were more commonly predicted in VSTs than in all transcripts (18.5% of 411 *S. grossa* VSTs vs. 8.6% of all 27,597 *S. grossa* transcripts; 19.3% of 377 *L. hesperus* VSTs vs. 8.7% of all 30,255 transcripts; 15.2% of 597 *L. geometricus* VSTs vs. 7.7% of all 29,141 transcripts).

We determined GO terms overrepresented in the VSTs compared with all transcripts that met the expression threshold as a summary of venom gland function (fig. 2). At an FDR of 0.1, there were no underrepresented GO terms for any species, and none overrepresented for *S. grossa* VSTs, 14 overrepresented for *L. geometricus* VSTs, and 7 for *L. hesperus* VSTs. At an FDR of 0.2, there were again no underrepresented terms for any species, and 3 overrepresented terms in *S. grossa*, 21 in *L. geometricus*, and 8 in *L. hesperus*. Only the GO cellular component category extracellular region was overrepresented among VSTs in all species, while proteolysis, hydrolase activity, metallopeptidase activity, and metalloendopeptidase activity were shared overrepresented GO categories in the *Latrodectus* species (fig. 2).

### Contrasting Effects of Venom Gland Expression on Family Size

Of the 1,385 VSTs identified across the 3 species, 684 (49.4%) were members of gene families, as determined by OrthoMCL analyses of translated proteins, while 701 were singletons (50.6%) that did not group with other sequences. A smaller percentage of broadly expressed (nonDE) transcripts were singletons (16,380 of 51,384: 31.9%). Of the 684 VSTs belonging to a gene family, 172 belonged to 66 VGE families in which all transcripts were venom gland specifically expressed (i.e., were VSTs), while 512 belonged to 389 VGC families, in which at least one but not all transcripts were VSTs (supplementary tables S3–S5, [Supplementary Material](#) online). BLAST hits were recovered for transcripts in 30 of the 66 VGE families, indicating that most (20 of 30) consist of known toxins, enzymes (chitinases, lipases, proteases) that may act as toxins or toxin spreading factors, or transcripts with BLAST homology to neural function proteins, found in the venom of *L. hesperus*, which may represent putative toxins (table 1 and



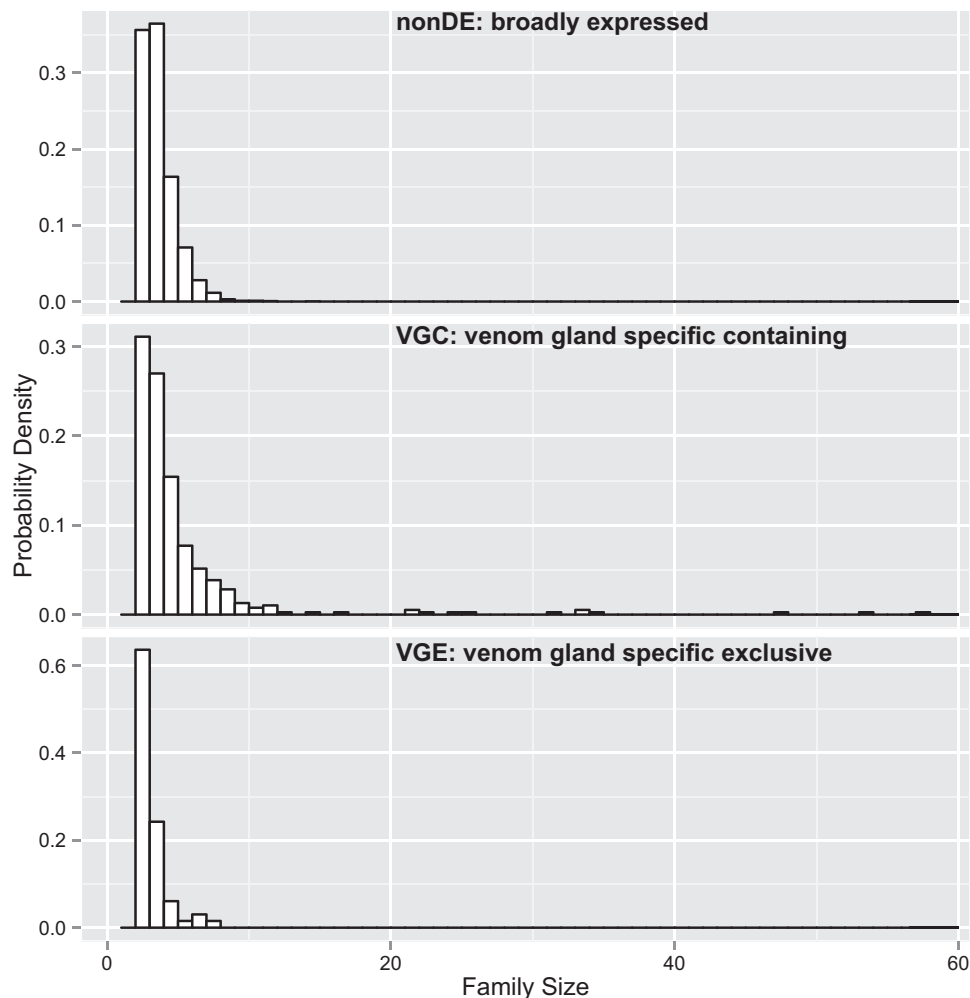
**Fig. 2.**—Similarities and differences among GO categories for VSTs across species. All GO categories overrepresented in at least one species are shown, along with the number of transcripts showing significantly biased venom gland expression in this category for each species. The y-axis is labeled with GO term identifiers and a description of the term is to the right of the bars, followed in parentheses by the ontology. \*Indicates that a term is significantly overrepresented at an FDR < 0.2 in a species. CC, cellular component ontology; MF, molecular function; BP, biological process.

supplementary table S3, [Supplementary Material](#) online). BLAST hits were recovered for transcripts in 190 of 389 VGC families (supplementary table S4, [Supplementary Material](#) online). Only 32 families showed homology to known venom toxins, enzymes, or neural function proteins, and instead a more diverse functional range of proteins was represented. However, of the remaining families, 92 contained 118 VSTs that encode a protein with a signal peptide, suggesting that the gene product may be secreted.

Despite our expectation that recruitment to venom gland expression should lead to increased family size through neofunctionalization and retention of gene duplicates (Kordis and Gubensek 2000; Fry 2005; Wong and Belov 2012), we found that the 66 VGE families were on average significantly smaller in number of contained sequences than the 4,108 nonDE families, which are composed of transcripts that exclusively exhibited broad expression (mean VGE family size = 2.61

sequences; mean nonDE = 3.11; median VGE = 2; median nonDE = 3; Wilcoxon rank-sum test:  $P = 5.66 \times 10^{-6}$ ) (fig. 3). If we restrict the VGE families to those with transcripts that have venom gland-specific expression and additional evidence of being a venom component (homology to a known toxin in BLAST, or possessing a signal peptide, or both:  $P = 17$ ), we found that the mean size (3.00) was still less than that of the broadly expressed control set of families.

In contrast to VGE families, VGC families were significantly larger on average than the nonDE families (mean VGC = 4.68; mean nonDE = 3.11; median VGC = 3; median nonDE = 3; Wilcoxon rank-sum test:  $P = 1.48 \times 10^{-8}$ ). To test whether the larger sizes of VGC families compared with nonDE families were related to an increase in the number of VSTs potentially occurring after a shift to venom gland expression, we performed a regression of VGC family size on the number of VSTs per family. Although the test result was significant



**Fig. 3.**—Family sizes for transcripts with specific patterns of expression. The y-axis is the probability density for a given family size in each of the three categories for pattern of expression. nonDE, broadly expressed control families; VGC, families containing, but not exclusively composed of, venom gland-specific transcripts; VGE, families composed exclusively of venom gland-specific transcripts.

( $P = 0.0018$ ), the effect size was low ( $R^2 = 0.025$ ). In fact, the majority of VGC families had only a single VST (308 of 389), and of the largest VGC families (>7 members), 36 of 53 (68%) had only a single VST.

### Gains and Losses of Venom Gland Biased Expression in Larger VGC Families

In the 36 largest VGC families (7 or more members) with only a single VST, examination of ancestral states of expression phenotype showed that the most common pattern was an unambiguous single gain of venom gland biased expression with no subsequent duplication. This pattern occurred in 25 families, in which the basal expression state had a probability ( $p > 0.65$ ) of not being venom gland biased, as did the ancestral state of all internal nodes in the phylogeny (fig. 4a and supplementary fig. S2, [Supplementary Material](#) online). For the remaining single VST families, internal node probabilities of nonvenom gland biased expression between 0.5 and 0.65 make it difficult to determine whether a single gain has occurred, and to place that gain relative to putative duplication events.

In the 17 families with at least 7 members and multiple VSTs, only 2 families (ther1047: 15-hydroxyprostaglandin dehydrogenase, ther1530: gamma glutamyl peptidase) had a single unambiguous gain followed by an increase in transcript number potentially due to gene duplication (fig. 4b and c). A loss of venom gland biased expression occurred in one family (ther1596; fig. 4d), but sequences in this family do not have significant BLAST hits in the *nr* database. In ther1596, the basal family state and all internal nodes indicate venom gland biased expression, with reversion to nonvenom gland biased expression in a single transcript. Only a single family (ther40: metalloendopeptidase-like) showed clear evidence of multiple gains of venom gland biased expression, which appear to have been followed in two instances by an increase in transcript numbers within *Latrodectus* species that could be due to duplication (fig. 5).

Fourteen multiple VST families had one or more internal nodes with ambiguous expression state, leading to uncertainty in inferring gains or losses of venom gland biased expression, although most had at least one gain, as the basal state was not venom gland biased (supplementary fig. S3a, [Supplementary Material](#) online). In other families, the basal state was ambiguous (supplementary fig. 3b, [Supplementary Material](#) online).

### Positive Selection on Venom Gland Expressed Families

Values of  $dN/dS$  from PAML model M0 were elevated in VGE ( $N = 23$ , median  $dN/dS = 0.137$ ) and VGC ( $N = 224$ , median  $dN/dS = 0.135$ ; supplementary table S6, [Supplementary Material](#) online) families when compared with the broadly expressed (nonDE) control set ( $N = 2,496$ , median  $dN/dS = 0.074$ ), with both pairwise comparisons significant by

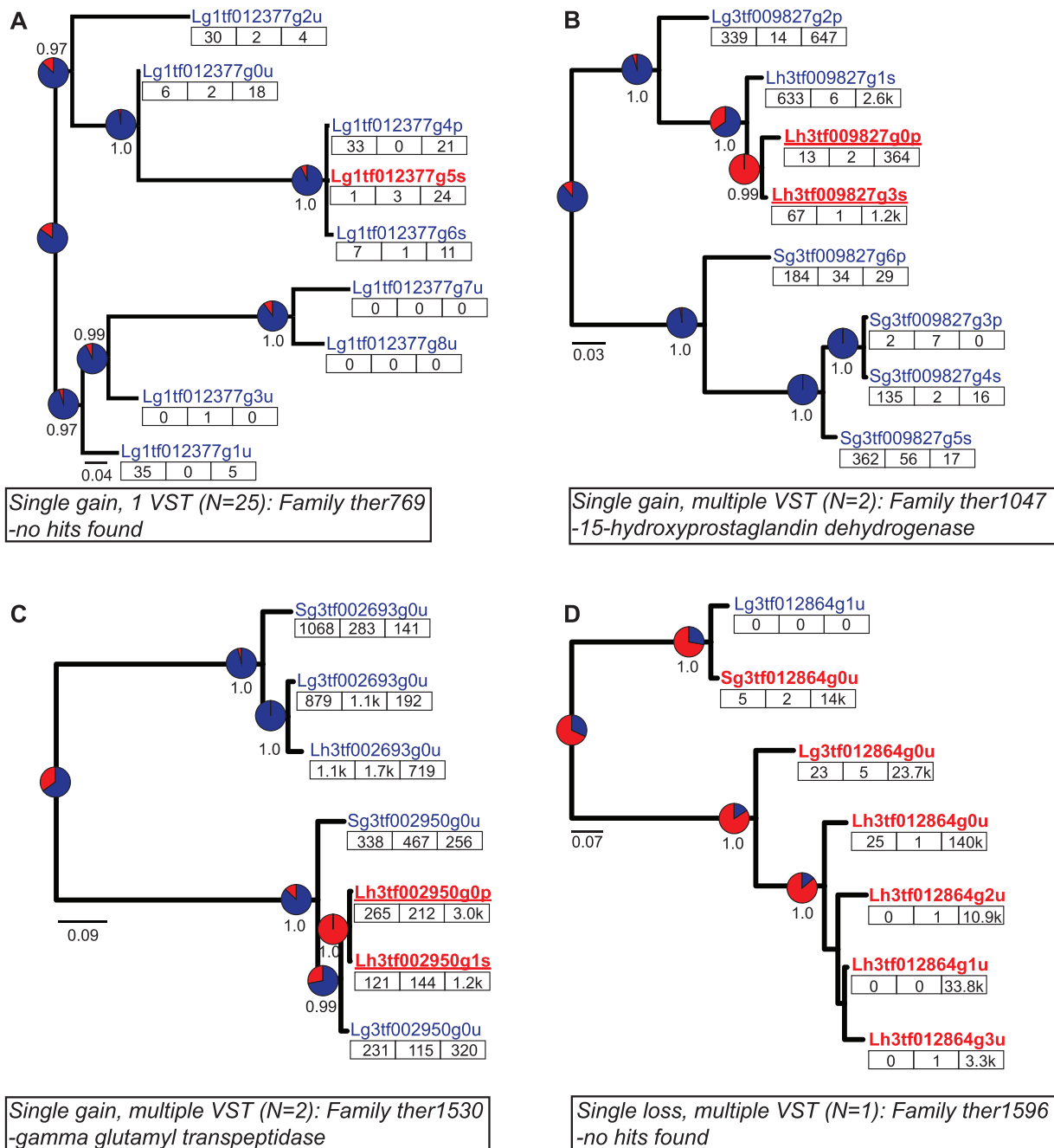
Wilcoxon rank-sum test (VGE vs. nonDE:  $P = 7.42 \times 10^{-5}$ ; VGC vs. nonDE:  $P = 2.20 \times 10^{-16}$ ). A comparison of  $dN/dS$  distributions between the VGE and VGC families yielded a nonsignificant result (Wilcoxon rank-sum test:  $P = 0.658$ ). Elevated  $dN/dS$  in VGE and VGC resulted from an increase in nonsynonymous substitutions: The median tree lengths for  $dN$  were higher for VGE (0.188) and VGC (0.176) than for nonDE families (0.070).

Among a total of 2,893 families tested, there were a total of 426 (14.7%) families that showed evidence for positive selection in a significantly better fit of PAML's Model M8 versus Model M7, with a category of sites with an average  $dN/dS > 1$  (table 2 and supplementary table S7, [Supplementary Material](#) online). By family type there were an excess of positive tests in VGE families, with 6 findings of positive selection from 24 tests (25%), while VGC families had 27 positives from 242 families (11.1%) and nonDE families had 393 from 2,627 (15%). However, this finding was neither significant in a three-way test of independence ( $G = 4.49$ ,  $df = 2$ ,  $P = 0.106$ ) nor in pairwise comparisons (VGE vs. VGC:  $G = 3.18$ ,  $df = 1$ ,  $P = 0.075$ ; VGE vs. nonDE:  $G = 1.62$ ,  $df = 1$ ,  $P = 0.203$ ). Of the six VGE families with evidence for positive selection in their evolution, three had no BLAST homology, and the other three families included transcripts with BLAST homology to two toxin types (CRISPs and latrotoxins) while the third had homology to leucine-rich repeat proteins. Ten positively selected VGC families had BLAST hits, of which five were to hypothetical or uncharacterized proteins, whereas the others were to iduronate 2-sulfatase-like, putative vacuolar amino acid transporter, histone-lysine N-methyltransferase SETMAR-like, UPF0462 protein C4orf33 homolog and NMD protein affecting ribosome stability and mRNA decay, putative.

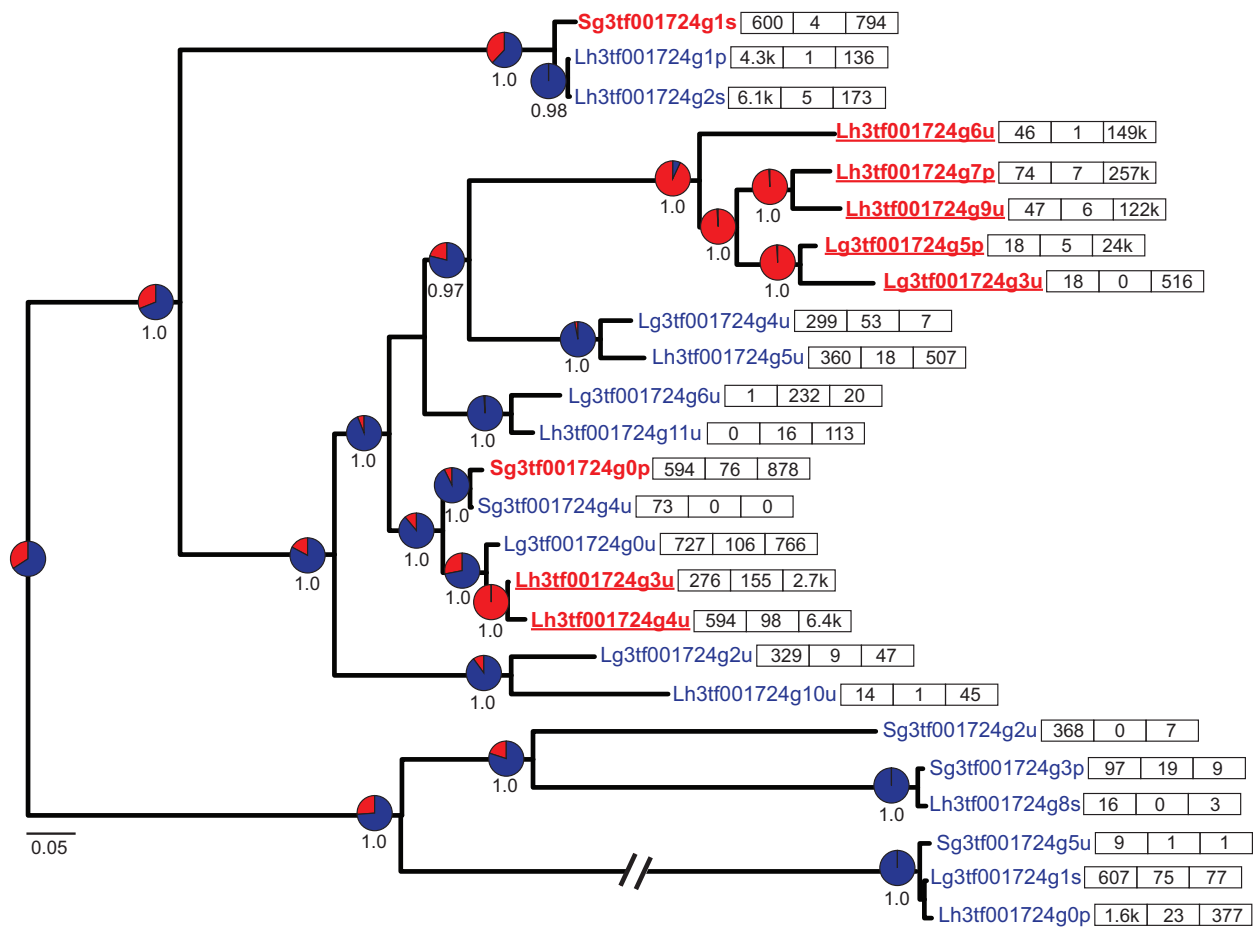
### Species-Specific Families and Singleton Transcripts Are Enriched in Venom Glands

Among VGE families, only 16 of the 66 included transcripts from all three species, while 32 were restricted to one of our three species (*S. grossa* = 7; *L. geometricus* = 14; *L. hesperus* = 11). These species-specific families were small (2–3 transcripts), and generally lacked BLAST homology (number of families containing transcripts with BLAST hits: *S. grossa* = 1; *L. geometricus* = 2; *L. hesperus* = 1). The exceptions had members with BLAST hits to latrotoxins in *L. geometricus*, metalloproteases in *L. hesperus*, and a hypothetical protein in *S. grossa* with a predicted signal peptide (supplementary table S3, [Supplementary Material](#) online). The VGC families tended to have a broader phylogenetic distribution, with 192 of the 389 having transcripts from all three species, while 117 were species specific, with *L. hesperus* having the fewest of these families (*S. grossa* = 44; *L. geometricus* = 59; *L. hesperus* = 14). Species-specific VGC families had a much wider range in size than species-specific VGE families





**FIG. 4.**—Gains and losses in venom gland biased expression in exemplar VGC families. Shown are midpoint rooted majority rule consensus trees generated from a minimum of 3,002 trees sampled after burn-in in Bayesian analysis. Numerical values at nodes are posterior probabilities. Transcripts are labeled by species abbreviation (Sg, *Steatoda grossa*; Lg, *Latrodectus geometricus*; Lh, *Latrodectus hesperus*), and VSTs are in bold and shaded red. Annotation of group transcripts is by BLASTX homology. Pie charts at nodes indicate the probability of the ancestral state being venom gland-specific expression (red) or not venom gland-specific expression (blue). Boxes show RSEM read counts averaged across replicate libraries for each transcript in each tissue in the following order: Cephalothorax, ovary, and venom gland. Underlined VSTs are putatively the result of duplication after a shift to venom gland biased expression.



**Fig. 5.**—Multiple gains in venom gland biased expression in a large VGC family with BLAST homology to membrane metalloendopeptidase-like proteins. Shown is a midpoint rooted majority rule consensus tree generated from 3,002 trees sampled after burn-in in Bayesian analysis. Numerical values at nodes are posterior probabilities. Transcripts are labeled by species abbreviation (Sg, *Steatoda grossa*; Lg, *Latrodectus geometricus*; Lh, *Latrodectus hesperus*), and VSTs are in bold and shaded red. Annotation of group transcripts is by BLASTX homology. Pie charts at nodes indicate the probability of the ancestral state being venom gland-specific expression (red) or not venom gland-specific expression (blue). Boxes show RSEM read counts averaged across replicate libraries for each transcript in each tissue in the following order: Cephalothorax, ovary, and venom gland. Underlined VSTs are putatively the result of duplication after a shift to venom gland biased expression. Slashes on branch represent shortening due to space considerations.

(2–33 members), but similarly to species-specific VGE family members, the species-specific VGC family members typically lacked BLAST homology (number of families with BLAST hits: *S. grossa* = 7; *L. geometricus* = 3; *L. hesperus* = 0; supplementary table S4, [Supplementary Material](#) online). Nine of the species-specific VGC families had a total of 12 VSTs with a signal peptide, suggesting a secreted gene product. Yet only one of these families corresponded to a known toxin (*ther7382*: Latrotoxin).

The size of species-specific VGE families did not vary significantly among species (mean size: *S. grossa* = 2; *L. geometricus* = 2.1; *L. hesperus* = 2.1), but species-specific VGC families are shifted toward larger sizes in

**Table 2**

PAML Results for VGE, VGC, and NonDE Transcript Families

Family Type	No. Families	No. of positive tests ( $P < 0.01$ )
VGE	24	6 (25%)
VGC	242	27 (11.1%)
NonDE	2627	393 (15%)

NOTE.—For each set of families, the number of families analyzed under Models M7 and M8 are shown together with the percentage of tests that indicated positive selection (M8 a significantly better fit at  $P = 0.01$  or better and average dN/dS for category of unconstrained sites > 1).

**Table 3**

Wilcoxon Rank-Sum *P* Values for Tests Comparing Distributions of Number of Transcripts in Each Species for Lineage-Specific and Three Species Shared Families

	<i>Steatoda grossa</i> vs. <i>Latrodectus geometricus</i>	<i>Steatoda grossa</i> vs. <i>Latrodectus hesperus</i>	<i>Latrodectus geometricus</i> vs. <i>Latrodectus hesperus</i>
Lineage-specific VGE families	0.3407	0.4941	0.7337
Lineage-specific VGC families	0.9452	0.0584	0.0537
Three species VGE families	0.3593	0.6115	0.7144
Three species VGC families	0.4076	0.1159	0.4930

**Table 4**

Transcripts with Species-Specific Upregulation Exclusively in the Venom Gland

	Total Sequences Upregulated in Species	VSTs Upregulated in Species	Total Sequences Upregulated with BLAST Hit	No. of GO Terms for All
<i>Latrodectus hesperus</i>	64	7	57	197
<i>Latrodectus geometricus</i>	10	0	9	26
<i>Steatoda grossa</i>	16	2	11	23

NOTE.—From 6,509 1:1:1 orthologous families, each species had transcripts that were significantly more abundant in that species venom gland compared with the other two species.

*L. hesperus* (mean size: *S. grossa* = 3.3; *L. geometricus* = 2.8; *L. hesperus* = 5.3), although these differences are only marginally significant (table 3). For VGE or VGC families with members from all three species, there is no significant difference in the distribution of the number of transcripts per species (table 3).

Each species also had a number of transcripts that did not cluster into a family with any other sequences, yet showed venom gland-specific expression. In all three species, VSTs were more likely to be singletons (701 of 1,385: 50.6%) than broadly expressed transcripts (16,380 of 51,384: 31.9%), and this difference was significant ( $\chi^2 = 215.41$ , *df* = 1, *P* < 0.0001). These VST singletons were on average 83 amino acids long, only 33% the length of VSTs that grouped into families (242 amino acids). The number of these singleton VSTs varied to some degree among species (*S. grossa*: 229 (55.7% of VSTs); *L. geometricus*: 297 (49.7%); *L. hesperus*: 175 (46.4%)). Thus, there are many distinct unknown transcripts in each species with venom gland-specific expression, as a reduced percentage of these species-specific singletons had significant BLAST hits to *nr* when compared with VSTs as a whole (*S. grossa*: 15.3%; *L. geometricus*: 16.5%; *L. hesperus*: 17.1%). However, BLAST hits indicated homology to toxins and enzymes (supplementary table S1, [Supplementary Material](#) online), and 45 sequences (*S. grossa*: 18; *L. geometricus*: 18; *L. hesperus*: 9) had predicted signal peptides.

#### An Excess of Upregulated Transcripts in the *Latrodectus hesperus* Venom Gland

Among 6,509 1:1:1 orthologs, we identified species-specific increases in expression that were exclusive to the

venom gland (table 4) and that could affect venom toxicity, which varies among species. *Latrodectus hesperus* had by far the most orthologs upregulated in venom glands relative to the other species (table 4 and supplementary table S8, [Supplementary Material](#) online). Few species-specific venom gland upregulated transcripts were VSTs (9 total, 7 of which were from *L. hesperus*) but most had BLAST hits (table 4), which indicated that transcripts with species-specific venom gland upregulation vary in identity and putative function (supplementary table S8, [Supplementary Material](#) online). Of the 246 GO terms matched to proteins in the venom gland upregulated sets of the three species (table 4), only three (GO: 0003676 nucleic acid binding; GO: 0046872 metal ion binding; GO: 0016021 integral component of membrane) were shared among all three species (supplementary table S9, [Supplementary Material](#) online). A total of 12 of these transcripts possessed a predicted signal peptide (3 *S. grossa*, 1 *L. geometricus*, and 8 *L. hesperus*).

As *L. hesperus* had the largest number of upregulated transcripts in the venom gland, we explored their possible functions via GO terms. The top two GO terms, associated with 10 and 7 transcripts, were both in the cellular component ontology: “Membrane” and “integral to membrane.” Of these transcripts, two were also VSTs with BLAST homology to a leucine-rich repeat transmembrane protein (Lh3tf007815g0u) and a putative vacuolar amino acid transporter (Lh3tf004605g1u). Non-VST transcripts included those with BLAST homology to other membrane components that have potential nervous system-related functions, including an inward rectifier K<sup>+</sup> channel, and glutamate (N-methyl-D-aspartate: NMDA) receptor subunit 3A.

## Discussion

Our comparative multitissue transcriptomic analyses of widow spider species identified numerous transcripts with venom gland biased expression (VSTs) and highlighted an array of evolutionary processes that have contributed to diversity in venom gland expressed transcripts. The first of these is positive selection, which acting in conjunction with gene duplication to generate protein sequence diversity and modified functions plays an important role in models of venom gland evolution, especially for toxin sequences involved in a predator–prey arms race. This can generate large families of functionally diverse toxins (Olivera 1997; Duda and Palumbi 1999). However, although positive selection affects a few VGE families (with BLAST hits mostly to known or putative toxins or spreading factors), we did not find a significant enrichment of positive selection in these families. Furthermore, all three expression categories include families with evidence for adaptive evolution. Although it is possible that positive selection has less of a role in the evolution of theridiid toxins than in other groups, we suspect that positive selection is more extensive for our VGE families than we uncovered because the power of tests was limited by the small family sizes (Anisimova et al. 2001, 2002). Of course, even if not pervasive, positive selection on specific families with venom gland specifically expressed transcripts could affect venom phenotypes by influencing protein function via changes in amino acid sequence, although functional differences among family members, and in some cases their toxicity, remain to be demonstrated.

Acting with selection to produce large, diverse toxin families in the traditional model of venom toxin gene evolution is gene duplication. After recruitment to venom gland expression, VGE families, although enriched in known or potential toxins, do not appear to have expanded substantially, as expected if retention of gene duplicates was a driving force. Yet family size assessment could be confounded by high rates of protein sequence divergence (Chang and Duda 2012). If toxin transcripts evolve rapidly and become highly diverged, they may not be grouped together by the orthoMCL algorithm, leading to increased numbers of smaller families. There is some evidence that this process may be contributing to the results for VGE families, as transcripts with BLAST hits to one of the several classes of toxin that are believed to have a common origin (ICKs, CRISPs, latrotoxins) do not form single VGE families (supplementary table S3, [Supplementary Material](#) online). Although it is necessary to use a single broad criterion for family definition in genome-wide studies, more focused studies involving manual curation should be performed to test the influence of duplication and selection in specific theridiid toxin families.

A second issue potentially confounding our results on duplication is fragmentation of transcriptome assemblies. In

conjunction with sequence divergence, incomplete transcript assembly may particularly affect the latrotoxins, as complete latrotoxin proteins are long (1,200–1,400+ amino acids), and despite structural similarities (Orlova et al. 2000), paralogs can be highly divergent (Garb and Hayashi 2013; Haney et al. 2014). No complete latrotoxins were recovered from the current transcriptome assemblies. Many transcripts with homology to latrotoxins consisted of fragments, and latrotoxins were dispersed across numerous small families. More complete transcripts, such as those generated from third generation sequencing technologies (Schadt et al. 2010), may improve transcriptome assembly and gene family definition, as would completed genome assemblies of these species. Fragmentation of longer transcripts may also affect family definition in each of the three categories, but would not necessarily bias toward smaller family sizes in the VGE category alone. Neither excess sequence divergence nor incomplete assembly obviates our finding that gene family expansion after a shift to venom gland biased expression is only one contributor to venom evolution. In addition, VSTs often originate within larger families after family expansion, a process complementary to the traditional model in which substantial gene duplication occurs after a shift to venom gland expression, which has been previously demonstrated for spider toxin families (Binford et al. 2009; Pineda et al. 2014).

Together with findings from other recent studies, our results indicate a richer set of scenarios for the evolution of venom and its constituent molecules (Casewell et al. 2012; Hargreaves et al. 2014), and suggest that a greater number of families than previously recognized may potentially contribute to venom diversity. Although the species studied here possess a broadly similar suite of toxins, enzymes, and other proteins with strongly venom gland biased expression, they varied in the number of toxins of a given type, the most common domains, and GO terms of their VSTs (fig. 2). These differences may be important in terms of potential influence on venom toxicity phenotypes, which vary among our focal species (Muller et al. 1992; Isbister and Gray 2003; Vetter and Isbister 2008; Graudins et al. 2012). Yet interspecific differences in VSTs traced to gains or losses of expression in VGC families involved few known toxins (supplementary table S4, [Supplementary Material](#) online). This is also true of the numerous species-specific VGE and VGC families, and venom gland-specific singleton transcripts, most of which have no BLAST homology. We also identified transcripts upregulated exclusively in the venom gland in a species-specific manner (table 4), although not necessarily tissue specific in expression, as most were not VSTs. Thus they provide a set of potential contributors to variation in venom phenotypes that is complementary to the VSTs. The majority of the transcripts identified as venom gland upregulated were identified in *L. hesperus*, and the most frequent GO terms were membrane and

integral component of membrane. These terms are also associated with the large neurotoxins (latrotoxins) that are a hallmark of widow venom, and which form pores in presynaptic membranes in the process of inducing massive neurotransmitter exocytosis (Orlova et al. 2000; Ushkaryov et al. 2004). Although latrotoxins are found in all three species, and thus predate divergence of the genera *Latrodectus* and *Steatoda*, it is possible that modulation of toxicity could be achieved by other proteins targeted to the neuronal cell membrane. Indeed, two of the *L. hesperus* transcripts upregulated in the venom gland had BLAST homology to integral membrane proteins involved in neural transmission, inward rectifying potassium channels, and NMDA receptors (Tempel et al. 1988; Van Dongen 2009).

We currently do not know the full extent to which VSTs or venom gland species-specific upregulated transcripts produce toxins and influence venom toxicity in our three focal species. These transcripts could also be involved in other processes important for general venom gland function or toxin processing, which requires further study. However, in addition to a strong bias in venom gland expression, many VST proteins possess signal peptides and may be secreted. Although not an exhaustive characterization, mass spectrometry data from *L. hesperus* venom (Haney et al. 2014) showed that it contains novel proteins (lacking BLAST hits or with hits to leucine-rich repeat proteins) in addition to known toxins (latrotoxins, inhibitory cysteine-knot toxins, CRISPs), and enzymes (proteases, lipases, hyaluronidases). In this study, transcripts in leucine-rich repeat protein families do show shifts to venom gland biased expression among our species (supplementary table S4, [Supplementary Material](#) online), and upregulation in the venom gland (supplementary table S8, [Supplementary Material](#) online). Further study is needed to determine if any other toxin candidate gene products are found in venom, and new candidate pharmacologically active molecules need to be verified by functional testing. It is also important from an evolutionary standpoint to understand how differences in venom composition relate to differences in prey taken, as *Latrodectus* females capture and consume small vertebrates in addition to arthropod prey (McCormick and Polis 1982; Hódar and Sánchez-Piñero 2002).

Gene expression is a dynamic process, and can change over short timescales in magnitude or pattern, in response to environmental cues, potentially including recency of prey capture. Levels of expression observed in this study reflect some combination of genetic changes that lead to hardwired patterns, and plastic responses to the environment, and thus the results of our gene expression analyses must be interpreted cautiously, particularly when making evolutionary inferences (Hodgins-Davis and Townsend 2009). However, we restricted our study to adult females kept in the laboratory for at least 3 days and fed an identical food item 1–5 days prior to

dissection, and included replicates to control for within-species variation for each tissue in each species. It is also clear from analyses of VGE families that some homologs can retain tissue-biased expression over evolutionary timescales. Nevertheless, evidence as to whether venom composition or its transcriptional underpinnings change in response to environment is generally lacking in spiders, and common garden experiments that discriminate between genetic and plastic responses are of vital importance. Finally, the importance of enhanced expression in the venom gland for venom toxicity needs to be tested, in terms of correlation between the level of transcription and protein abundance in the venom secretion, and the physiological effects of differing concentrations of a given toxin on prey.

In conclusion, we found that extensive diversification after a shift to venom gland expression may be complemented by other processes, including limited gene expansions postrecruitment, and shifts to venom gland-specific expression in diverse gene families. Together with species-specific orthologs upregulated in the venom gland identified in this study, the substantial variation in protein sequences of VSTs, and lineage-specific families and singletons, these shifts to venom gland biased expression in functionally diverse protein families provide candidate mechanisms for interspecific variation in venom toxicity.

## Supplementary Material

Supplementary figures S1–S3 and tables S1–S9 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

We thank Sandra Correa-Garhwal for assistance in spider collection and gland dissection, and Caryn McCowan for aid in acquiring EST data for *Latrodectus geometricus*. We thank Crystal Chaw and Kerry Gendreau for helpful comments on the manuscript. This work was supported by the National Science Foundation (IOS-0951886 to N.A.A., IOS-0951061 to C.Y.H.) and the National Institutes of Health (1F32GM083661-01, 1R15GM097714-01, and 2R15GM097714-02 to J.E.G.).

## Literature Cited

- Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol.* 18:1585–1592.
- Anisimova M, Bielawski JP, Yang Z. 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Mol Biol Evol.* 19:950–958.
- Arnedo MA, Coddington J, Agnarsson I, Gillespie RG. 2004. From a comb to a tree: phylogenetic relationships of the comb-footed spiders (Araneae, Theridiidae) inferred from nuclear and mitochondrial genes. *Mol Phylogenet Evol.* 31:225–245.

- Atakuziev BU, Wright CE, Graudins A, Nicholson GM, Winkel KD. 2014. Efficacy of Australian red-back spider (*Latrodectus hasselti*) antivenom in the treatment of clinical envenomation by the cupboard spider *Steatoda capensis* (Theridiidae). *Toxicon* 86:68–78.
- Binford GJ, et al. 2009. Molecular evolution, functional variation, and proposed nomenclature of the gene family that includes sphingomyelinase D in sicariid spider venoms. *Mol Biol Evol.* 26:547–566.
- Bollback JP. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7:88 doi: 10.1186/1471-2105-7-88.
- Casewell NR, Huttley GA, Wüster W. 2012. Dynamic evolution of venom proteins in squamate reptiles. *Nat Commun.* 3:1066.
- Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG. 2013. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol Evol.* 28:219–229.
- Cavaliere M, et al. 1987. Characterization and some properties of the venom gland extract of a theridiid spider (*Steatoda paykulliana*) frequently mistaken for black widow spider (*Latrodectus tredecimguttatus*). *Toxicon* 25:965–974.
- Chang D, Duda TF. 2012. Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. *Mol Biol Evol.* 29:2019–2029.
- Clarke TH, Garb JE, Hayashi CY, Arensburger P, Ayoub NA. 2015. Spider transcriptomes identify ancient large-scale gene duplication event potentially important in silk gland evolution. *Genome Biol Evol.* 7:1856–1870.
- Duda TF, Palumbi SR. 1999. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc Natl Acad Sci U S A.* 96:6820–6823.
- Duda TF, Remigio EA. 2008. Variation and evolution of toxin gene expression patterns of six closely related venomous marine snails. *Mol Ecol.* 17:3018–3032.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Fry BG. 2005. From genome to “venome”: molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res.* 15:403–420.
- Fry BG, et al. 2009. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu Rev Genomics Hum Genet.* 10:483–511.
- Garb JE, Hayashi CY. 2013. Molecular evolution of alpha-Latrotoxin, the exceptionally potent vertebrate neurotoxin in black widow spider venom. *Mol Biol Evol.* 30:999–1014.
- Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29:644–652.
- Graudins A, et al. 2012. Cloning and activity of a novel  $\alpha$ -latrotoxin from red-back spider venom. *Biochem Pharmacol.* 83:170–183.
- Haney RA, Ayoub NA, Clarke TH, Hayashi CY, Garb JE. 2014. Dramatic expansion of the black widow toxin arsenal uncovered by multi-tissue transcriptomics and venom proteomics. *BMC Genomics* 15:366.
- Hargreaves AD, Swain MT, Hegarty MJ, Logan DW, Mulley JF. 2014. Restriction and recruitment-gene duplication and the origin and evolution of snake venom toxins. *Genome Biol Evol.* 6:2088–2095.
- He Q, et al. 2013. The venom gland transcriptome of *Latrodectus tredecimguttatus* revealed by deep sequencing and cDNA library analysis. *PLoS One* 8:e81357.
- Hódar JA, Sánchez-Piñero F. 2002. Feeding habits of the black widow spider *Latrodectus lilianae* (Araneae: Theridiidae) in an arid zone of south-east Spain. *J Zool.* 257:101–109.
- Hodgins-Davis A, Townsend JP. 2009. Evolving gene expression: from G to E to G×E. *Trends Ecol Evol.* 24:649–658.
- Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Res.* 9:868–877.
- Isbister GK, Fan HW. 2011. Spider bite. *Lancet* 378:2039–2047.
- Isbister GK, Gray MR. 2003. Latrodectism: a prospective cohort study of bites by formally identified redback spiders. *Med J Aust.* 179:88–91.
- Kent WJ. 2002. BLAT—The BLAST-like alignment tool. *Genome Res.* 12:656–664.
- King GF, Hardy MC. 2013. Spider-venom peptides: structure, pharmacology, and potential for control of insect pests. *Annu Rev Entomol.* 58:475–496.
- Kordis D, Gubensek F. 2000. Adaptive evolution of animal toxin multigene families. *Gene* 261:43–52.
- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–2189.
- Maretic Z, Levi HW, Levi LR. 1964. The theridiid spider *Steatoda paykulliana*, poisonous to mammals. *Toxicon* 2:149–154.
- McCormick S, Polis GA. 1982. Arthropods that prey on vertebrates. *Biol Rev.* 57:29–58.
- McCowan C, Garb JE. 2014. Recruitment and diversification of an ecdysozoan family of neuropeptide hormones for black widow spider venom expression. *Gene* 536:366–375.
- Muller G, et al. 1992. Comparison of the toxicity, neurotransmitter releasing potency and polypeptide composition of the venoms from *Steatoda foravae*, *Latrodectus indistinctus*, and *L. geometricus* (Araneae: Theridiidae). *S Afr J Sci.* 88:113–116.
- Muller GJ, Koch HM, Kriegler A, van der Walt BJ, van Jaarsveld P. 1989. The relative toxicity and polypeptide composition of the venom of two Southern African widow spider species: *Latrodectus indistinctus* and *Latrodectus geometricus*. *S Afr J Sci.* 85:44–46.
- Olivera BM. 1997. E.E. Just Lecture, 1996. *Conus* venom peptides, receptor and ion channel targets, and drug design: 50 million years of neuropharmacology. *Mol Biol Cell.* 8:2101–2109.
- Orlova EV, et al. 2000. Structure of alpha-Latrotoxin oligomers reveals that divalent cation-dependent tetramers form membrane pores. *Nat Struct Biol.* 7:48–53.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 8:785–786.
- Pineda SS, et al. 2014. Diversification of a single ancestral gene into a successful toxin superfamily in highly venomous Australian funnel-web spiders. *BMC Genomics* 15:177.
- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490.
- Robinson MD, McCarthy DJ, Smyth GK. 2009. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
- Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61:539–542.
- Sanz L, Gibbs HL, Mackessy SP, Calvete JJ. 2006. Venom proteomes of closely related *Sistrurus* rattlesnakes with divergent diets. *J Proteome Res.* 5:2098–2112.
- Schadt EE, Turner S, Kasarskis A. 2010. A window into third-generation sequencing. *Hum Mol Genet.* 19:R227–R240.
- Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34:W609–W612.
- Tempel BL, Jan YN, Jan LY. 1988. Cloning of a probable potassium channel gene from mouse brain. *Nature* 332:837–839.
- Ushkaryov YA, Volynski KE, Ashton AC. 2004. The multiple actions of black widow spider toxins and their selective use in neurosecretion studies. *Toxicon* 43:527–542.
- Van Dongen AM, editors 2009. *Biology of the NMDA receptor*. CRC Press: Boca Raton (FL).

- Vetter RS, Isbister GK. 2008. Medical aspects of spider bites. *Annu Rev Entomol.* 53:409–429.
- Warden CD, Yuan YC, Wu X. 2013. Optimal calculation of RNA-Seq fold-change values. *Int J Comput Bioinforma In Silico Model.* 2:285–292.
- Warrell DA, Shaheen J, Hillyard PD, Jones D. 1991. Neurotoxic envenoming by an immigrant spider (*Steatoda nobilis*) in southern England. *Toxicon* 29:1263–1265.
- Wong ES, Belov K. 2012. Venom evolution through gene duplications. *Gene* 496:1–7.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
- Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.
- Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-Seq: accounting for selection bias. *Genome Biol.* 11:R14.
- Zhang Y, et al. 2010. Transcriptome analysis of the venom glands of the Chinese wolf spider *Lycosa singoriensis*. *Zoology* 113:10–18.