A peer-reviewed version of this preprint was published in PeerJ on 1 June 2018.

View the peer-reviewed version (peerj.com/articles/4691), which is the preferred citable publication unless you specifically need to cite this preprint.

Zobel-Thropp PA, Bulger EA, Cordes MHJ, Binford GJ, Gillespie RG, Brewer MS. 2018. Sexually dimorphic venom proteins in long-jawed orb-weaving spiders (*Tetragnatha*) comprise novel gene families. PeerJ 6:e4691 https://doi.org/10.7717/peerj.4691

Sexually dimorphic venom proteins in long-jawed orb-weaving spiders (*Tetragnatha*) with potential roles in sexual interactions

Pamela A Zobel-Thropp [1] , **Emily A Bulger** [2] , **Matthew H J Cordes** [3] , **Greta J Binford** [1] , **Rosemary G Gillespie** [4] , **Michael S Brewer** [Corresp. 5]

[1] Department of Biology, Lewis and Clark College, Portland, OR, United States
[2] Division of Biological Sciences, University of California, San Diego, San Diego, CA, United States
[3] Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ, United States
[4] Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, CA, United States
[5] Department of Biology, East Carolina University, Greenville, NC, United States

Corresponding Author: Michael S Brewer
Email address: brewermi14@ecu.edu

Venom has been associated with the ecological success of many groups of organisms, most notably reptiles, gastropods, and arachnids. In some cases, diversification has been directly linked to tailoring of venoms for dietary specialization. Spiders in particular are known for their diverse venoms and wide range of predatory behaviors, although there is much to learn about scales of variation in venom composition and function. The current study focuses on venom characteristics in different sexes within a species of spider. We chose the genus *Tetragnatha* (Tetragnathidae) because of its unusual courtship behavior involving interlocking of the venom delivering chelicerae (i.e., the jaws), and several species in the genus are already known to have sexually dimorphic venoms. Here, we use transcriptome and proteome analyses to identify venom components that are dimorphic in *Tetragnatha versicolor*. We present cDNA sequences of unique high molecular weight proteins that are only present in males and that have remote, if any, detectable similarity to known venom components in spiders or other venomous lineages and several have no detectable homologs in existing databases. While the function of these proteins is not known, their presence in association with the cheliceral locking mechanism during mating together with the presence of prolonged male-male mating attempts in a related, cheliceral-locking species (*Doryonychus raptor*) lacking the dimorphism suggests potential for a role in sexual communication.

PeerJ Preprints NOT PEER-REVIEWED

1
2
3
4 **Sexually dimorphic venom proteins in long-jawed orb-weaving spiders (*Tetragnatha*) with**
5 **potential roles in sexual interactions**
6
7 Pamela A. Zobel-Thropp[1], Emily A. Bulger[2], Mathew H.J. Cordes[3], Greta J. Binford[1], Rosemary
8 G. Gillespie[4], Michael S. Brewer[5]
9
10 [1]Department of Biology, Lewis and Clard College, Portland, OR, USA
11 [2]Division of Biological Sciences, University of California, San Diego, San Diego, CA, USA
12 [3]Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ, USA
13 [4]Deparment of Environmental Science, Policy, and Management, University of California,
14 Berkeley, Berkeley, CA, USA
15 [5]Department of Biology, East Carolina University, Greenville, NC, USA
16
17 Corresponding Author:
18 Michael S. Brewer[5]
19
20 Email address: brewermi14@ecu.edu
21
22
23

24 **ABSTRACT**

25 Venom has been associated with the ecological success of many groups of organisms, most
26 notably reptiles, gastropods, and arachnids. In some cases, diversification has been directly
27 linked to tailoring of venoms for dietary specialization. Spiders in particular are known for their
28 diverse venoms and wide range of predatory behaviors, although there is much to learn about
29 scales of variation in venom composition and function. The current study focuses on venom
30 characteristics in different sexes within a species of spider. We chose the genus *Tetragnatha*
31 (Tetragnathidae) because of its unusual courtship behavior involving interlocking of the venom
32 delivering chelicerae (i.e., the jaws), and several species in the genus are already known to have
33 sexually dimorphic venoms. Here, we use transcriptome and proteome analyses to identify
34 venom components that are dimorphic in *Tetragnatha versicolor*. We present cDNA sequences
35 of unique high molecular weight proteins that are only present in males and that have remote, if
36 any, detectable similarity to known venom components in spiders or other venomous lineages
37 and several have no detectable homologs in existing databases. While the function of these
38 proteins is not known, their presence in association with the cheliceral locking mechanism during
39 mating together with the presence of prolonged male-male mating attempts in a related,
40 cheliceral-locking species (*Doryonychus raptor*) lacking the dimorphism suggests potential for a
41 role in sexual communication.

42

43 **INTRODUCTION**

44 Phenotypic differences between sexes are widespread among animals, and can be attributed to
45 sexual and/or natural selection. Sexual selection is the most common explanation for
46 morphological and behavioral differences, which are frequently attributed to intra- or intersexual

PeerJ Preprints NOT PEER-REVIEWED

47 competition for mates. However, natural selection may also lead to differences as a result of
48 bimodal or dimorphic niche separation between the sexes (Berns, 2013). Recent work has
49 highlighted the importance of additional modalities that differ between sexes, including acoustic
50 (Elias & Mason, 2014) and chemical (Wyatt, 2014) traits. Here we focus on sexual differences in
51 chemical traits, and more specifically on venom, in which several studies have highlighted
52 variation according to sex. We examine sexual differences in venom composition of spiders, and
53 consider mechanisms of selection that may have given rise to these differences.

54 Venoms are complex chemical cocktails that attract research attention in both applied and
55 basic sciences and have been characterized in many animals, including mammals (shrews,
56 platypus), toxicoferan reptiles (lizards and snakes), fish, sea anemones, cephalopods, cone snails,
57 insects, centipedes, scorpions, shrews, and arachnids (Fry et al., 2009). They can be impressively
58 complex; among spiders in particular, individual venoms can have 1000s of components
59 (Escoubas, 2006). The complexity typically consists of related sets of molecules within which
60 are components with exquisite functional specificity, and novel activities (reviews in (Kuhn-
61 Nentwig & Stöcklin, 2011; Smith et al., 2013; King, 2015). Activities involve manipulation of
62 physiological processes, particularly neurological, thus they are a source of discovery of
63 components with human applications in pharmaceuticals or insecticides (King, 2015).

64 Venoms frequently exhibit sexual dimorphism (snakes, (Menezes et al., 2006); scorpions,
65 (D'Suze, Sandoval & Sevcik, 2015; Miller et al., 2016); spiders, (Herzig et al., 2008; Binford,
66 Gillespie & Maddison, 2016b)). Because the primary functional roles of venoms in spiders are
67 thought to be predation and defense, in most cases sexual differences in venom composition
68 between adults of many species are hypothesized to result from natural selection optimizing
69 composition based on differences in feeding biology and associated differences in diet

70 composition and/or vulnerability to predation. Thus, hypotheses to explain sexual dimorphism in
71 spider venoms typically center on differential optimization of chemical pools for divergent adult
72 niches. Though largely untested, these hypotheses provide plausible explanations for many
73 known dimorphisms. For example, in Sydney funnel-web spiders (*Atrax robustus*), male venoms
74 are more toxic to mammals than are those of females (Gray & Sutherland, 1978), potentially
75 associated with male *Atrax* being found more often wandering outside of burrows than females
76 (Isbister & Gray, 2004). In contrast, in the theridiid spiders *Latrodectus mactans* and *Steatoda*
77 *paykulliana*, female venoms have higher mammalian neurotoxic activity than male venoms
78 {Maretic:1964um}. These spiders eat vertebrates and bite humans defensively, so the general
79 pattern of increased female potency has been attributed to the shorter lifespan of male spiders
80 and reduced adult foraging of males (Rash, King & Hodgson, 2000).

81 While sexual differences in venom composition driven by adult niche are likely, venoms
82 are also known to play a role in sexual biology (Polis & Sissom, 1990) and thus may be under
83 the influence of sexual selection. As secreted molecules with intra-individual functionality, they
84 have potential for biological roles in sex (Binford, Gillespie & Maddison, 2016b). The possibility
85 that different components within an individual's venom have different roles and/or contexts of
86 usage, sets up exciting potential for considering competing or complementary evolutionary
87 mechanisms influencing venom phenotypes.

88 The long jawed orb-weaving spiders (Araneae: Tetragnathidae: *Tetragnatha*) provide a
89 compelling context to explore the nature and potential cause of sexual dimorphism in venoms.
90 Members of the genus *Tetragnatha* are broadly distributed with *ca.* 347 species worldwide
91 (World Spider Catalog). The majority of species worldwide are remarkably uniform in
92 appearance, dull brown or olive in color, with long first and second legs, typically long jaws in

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3358v1 | CC BY 4.0 Open Access | rec: 20 Oct 2017, publ: 20 Oct 2017

93 adulthood, and an elongate opisthosoma (Levi, 1981). Their behavior and ecology is also fairly
94 homogeneous as they generally construct a light and fragile orb web with an open center and
95 build the web over water or in other wet places (Gillespie, 1987), although they have undergone
96 adaptive radiation in the Hawaiian Islands, associated with marked shifts in ecology and
97 behavior (Gillespie, 2004a; Blackledge & Gillespie, 2004). One striking aspect of this genus of
98 spiders is their very unusual sexual behavior: While courtship in most spiders involves an
99 elaborate and extended period of vibrational or visual communication, in most *Tetragnatha* there
100 is little evident communication prior to the male and female approaching each other. They
101 connect physically by spreading the chelicerae wide and locking fangs (Figure 1A), involving a
102 dorsal spur on the male chelicerae. The cheliceral locking mechanism apparently precludes the
103 need for epigynal coupling and is associated with secondary loss of a sclerotized epigynum (Levi,
104 1981).

105 Comparisons of crude venoms between sexes of *Tetragnatha* using 1-D protein
106 electrophoresis have identified a particularly striking sexual dimorphism in which males have an
107 abundance of high molecular weight components that are not in females (Binford, Gillespie &
108 Maddison, 2016b). Phylogenetic comparisons across species indicate that these high molecular
109 weight components persist across an evolutionary transition in feeding biology that reduces the
110 differences in adult feeding niches. Specifically, adult males of orb-weaving *Tetragnatha* species
111 do not typically build webs and are functionally wandering predators. However, a lineage of
112 *Tetragnatha* in Hawaii has lost orb-weaving behavior (Gillespie, 2004b; 2005) and both males
113 and females wander in search of prey, thus reducing dimorphism in feeding biology. Males could
114 be more prone to predation and have unique components that function in defense, but if so,
115 increased vulnerability would also affect female wandering *Tetragnatha.* So a defensive role

116 does not seem likely. This pattern, combined with the cheliceral locking behaviors during
117 copulation, lead to a hypothesis that the unique male components in venom play some as yet
118 undescribed role in mating biology (Binford, Gillespie & Maddison, 2016b).

119 A starting place for testing our hypothesis that the unique components in male
120 *Tetragnatha* venoms function in sexual biology, is to characterize the chemicals that contribute
121 to the differences. The goal of this study is to identify the molecules that are sexually dimorphic
122 in venoms of a readily accessible "model" species, *Tetragnatha versicolor*. Using comparative
123 venom gland transcriptomes and proteomes of adult males and females, we identify sequence
124 characteristics of dimorphic components, with particular attention to those unique to males. We
125 infer function preliminarily using homology searching and report the discovery of divergent
126 highly expressed male-specific proteins that support the potential for functioning in mating
127 biology.

128

129 **MATERIAL AND METHODS**

130 ***Collection***. Individuals of *Tetragnatha versicolor* were collected by hand from three populations:
131 along the southern fork of Strawberry Creek on the UC Berkeley Campus (UCB) (37.872ºN,
132 122.262ºW), Little Sugar Creek, Binford Farm (BF), Crawfordsville, Indiana (40.061ºN,
133 86.853ºW), and Greenville, North Carolina (ECU) (35.626ºN, 77.409ºW).

134 ***Venom Extraction***. Live specimens were transported to Lewis & Clark College where venom
135 was extracted by electrostimulation (Binford, 2001). We obtained venom samples from 19
136 females and 16 males from UCB, nine females and two males from BF, and nine females and
137 five males from ECU. To compile sufficient protein amounts for proteomics analyses, venom
138 was pooled within sexes for each of these three populations.

139 ***RNA Isolation*** Trancsriptomic analyses were performed using RNA isolated only from Indiana
140 (BF) specimens. To capture some breadth of transcriptional timing after emptying venom glands,
141 surviving spiders (10 females and two males) were divided into two groups within each sex, and
142 venom glands were isolated from five females and one male two and three days after venom
143 extraction. The glands extracted two and three days after milking were pooled within sexes, and
144 processed and analyzed separately between sexes for all subsequent analyses. To extract glands,
145 spiders were anesthetized with CO_2 and venom glands were removed by dissection and flash-
146 frozen immediately in liquid nitrogen. Total RNA was isolated by grinding tissues in TRIzol®
147 reagent (Life Technologies, Carlsbad, CA), adding chloroform (200 µL per mL of TRIzol®),
148 mixing by inversion, and incubating for 20 min at 4° C. The tube was centrifuged at 4° C for 15
149 min at 14,000 rpm. An equal volume of cold 100% ethanol was added to the RNA-containing
150 upper aqueous phase. The solution was then passed through an RNeasy® Mini Spin Column
151 (Qiagen, Chatsworth, CA) and purified, according to RNeasy® protocols.

152 ***Illumina RNA Sequencing and Quality Control***. RNA extractions from the BF population of *T.*
153 *versicolor* males (2) and females (10) were shipped to the Genomic Services Lab at
154 HudsonAlpha (Huntsville, AL), where cDNA libraries were prepared from total RNA (poly-A
155 isolation, Illumina TruSeq RNA Library Prep Kit v2), and 50 bp paired-end Illumina HiSeq 2500
156 RNA-seq was used to generate sequence reads. All QC trimming and assemblies were done
157 using a pipeline provided by the UC Berkeley Museum of Vertebrate Zoology
158 (https://github.com/MVZSEQ). Quality and GC content of the resulting paired-end reads was
159 assessed using the FastQC v0.10.0 program
160 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). TRIMMOMATIC v0.36 (Bolger,
161 Lohse & Usadel, 2014) and CUTADAPT v1.7.1 (Martin, 2011) were used to clean up the

162 sequence data: nucleotides below a quality threshold of 20 were trimmed from the ends of
163 sequences, and sequences shorter than 36 nucleotides (after trimming) were discarded. The reads
164 were aligned to a custom library of bacterial sequences to remove prokaryotic contamination
165 using BOWTIE2 v2.1.0 (Langmead & Salzberg, 2012). Individual paired-end files were
166 resynchronized, removing any paired-end sequences only present in one of the two files. Left and
167 right reads that overlap were combined into a single longer read to aid in downstream assembly
168 using FLASH v1.2.7 (Magoč & Salzberg, 2011).

169 ***Transcriptome Assembly and ORF Prediction***. The resulting male and female files were
170 assembled separately, by sex, and combined for a general *T. versicolor* venom transcriptome.
171 The processed reads for each sex and the combined read files were assembled using the
172 TRINITY pipeline v2.0.6 (http://trinityrnaseq.sourceforge.net/) with default parameters except
173 the following, group_pairs_distance=999 and min_kmer_cov=2. High-confidence open reading
174 frames (ORFs) (i.e., likely coding sequences), were obtained for each gene in the transcriptome
175 using TRANSDECODER r20140704 (Haas et al., 2013). A minimum protein length of 30 amino
176 acids was used in ORF predictions. The completeness of each assembly was assessed via
177 BUSCO v1.1 (Simão et al., 2015).

178 ***Read mapping to identify transcriptome dimorphisms.*** The processed reads from each sex were
179 mapped against the combined assembly to identify genes that are unique to either sex, with
180 particular emphasis placed on male-only transcripts. The mapping was performed using STAR
181 v2.4.2a (Dobin et al., 2012) and default parameters. A custom python script was used to generate
182 a BED file from the combined transcriptome and BEDTOOLS v2.18.1 "multicov" (Quinlan &
183 Hall, 2010) was used to generate counts of reads from each sex mapping to combined assembly
184 transcripts. An additional round of mapping with BOWTIE v1.1.1 (Langmead et al., 2009)

185 followed by GFOLD v1.1.4 (Feng et al., 2012) analysis identified differentially expressed
186 transcripts between males and females (GFOLD cutoff = 2, which approximates a $\log_2$ fold
187 change of two).

188 ***Functional annotation and GO enrichment analyses.*** The male, female, and combined
189 assemblies were annotated via the TRINOTATE pipeline (Haas et al., 2013). This approach
190 comprises the following steps. All contigs (BLASTX) and predicted amino acid sequences
191 (BLASTP) were searched against the Swissprot database (downloaded 23-iv-2015). Protein
192 domains were identified by running a HMMER v3.1b2 (Finn, Clements & Eddy, 2011) search
193 against the PFAM ((Bateman et al., 2004); downloaded 23-iv-2015) database, and signal
194 peptides, indicating secreted proteins, were discovered with SignalP v4.1
195 (http://www.cbs.dtu.dk/services/SignalP/). TMHMM v2.0c (Krogh et al., 2001) was used to
196 annotate transmembrane domains. Finally, RNAMMER v1.2 (Lagesen et al., 2007) identified
197 rRNA transcripts. Results from database searches were loaded into a sqlite database, and GO
198 terms were applied and used in downstream analyses.

199 GO enrichment analyses were performed on the combined assembly using the subset of
200 genes endemic to male spiders, as verified by the proteomic analyses outlined below. Two data
201 sets were analyzed, 1) all male specific proteins and 2) high molecular weight (>43 kDa) male
202 proteins. This was done using scripts provided with the TRINITY and TRINOTATE software
203 and the R Bioconductor package "GOseq" v1.18.0 (Young et al., 2010).

204 ***Gene family reconstruction via Markov clustering.*** Predicted ORFs from the male, female, and
205 combined assemblies were combined into a single FASTA file, a BLAST database was created,
206 and the sequences were searched (BLASTP; e-value cutoff = $e10^{-5}$) against themselves (i.e., an
207 ALLvsALL BLAST). The results were clustered into putative gene families using the Markov

208 Clustering Algorithm (MCL v14.137; (Enright, Van Dongen & Ouzounis, 2002)) with default
209 parameters and an inflation value of 2.0. The resulting e-values were negative log transformed,
210 and the results were separated using a heuristically chosen cutoff (-1.91) and results were
211 visualized in CYTOSCAPE v3.0.1 (Shannon et al., 2003). Clusters representing putative gene
212 families were used in subsequent analyses. Initial clustering identified a single cluster of
213 particular interest, cluster six, comprising high molecular weight proteins that was subsequently
214 subclustered (negative log transformed and heuristically chosen edge weight cutoff = -0.93).
215 ***Proteomic analyses***. Crude venoms were dissolved in a standard buffer (5 mM $CaCl_2$/50 mM
216 Tris, pH 8), pooled by sex, and shipped to the Arizona Proteomics Consortium. Venom-
217 expressed proteins were separated by size using SDS-PAGE (12%, Fig 1C). To increase
218 resolution, each lane was divided into three sections and digested with trypsin followed by a
219 clean-up step using C18 ZipTips (Millipore). Tryptic peptides were analyzed using an LTQ
220 Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) and the resulting MS/MS data were
221 searched using SEAQUEST on DISCOVERER (Thermo Fisher Scientific, v 1.3.0.339) against
222 masses of theoretical fragments from a database that included our translated transcriptome
223 sequences, and all chelicerate sequences in NCBI (downloaded 4/23/2015), totaling 171,068
224 sequences. Matches required a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance
225 of 10.0 ppm; oxidation of methionine and carbamidomethyl of cysteine were specified in
226 Seaquest as variable modifications. The Seaquest output was organized in Scaffold (Proteome
227 Software Inc, v 4.4.3). Peptides were identified with 90% minimum threshold and 0% false
228 discovery rate (FDR) and proteins were identified with 100% minimum threshold and 0% FDR.
229
230 **RESULTS**

231 ***Data archiving*** All raw read data are available through the NCBI short read archive (SRA
232 accession number SRP118124). Results and data files associated with proteomics
233 (https://dx.doi.org/10.6084/m9.figshare.5378308;
234 https://dx.doi.org/10.6084/m9.figshare.5378299), gene family clustering
235 (https://dx.doi.org/10.6084/m9.figshare.5420986), gene ontology/annotations
236 (https://dx.doi.org/10.6084/m9.figshare.5421655), and differential expression
237 (https://dx.doi.org/10.6084/m9.figshare.5421517) are available via Figshare (project URL:
238 https://figshare.com/account/home#/projects/24559.
239 ***Transcriptome*** The final trimmed read files comprised 43,550,457 ♀ (20,370,960 left;
240 20,370,960 right; 2,808,555 merged), 48,457,662 ♂ (22,783,475 left; 22,783,475 right;
241 2,890,675 merged), and 92,008,100 combined (male and female reads added together) reads. The
242 Trinity assemblies produced 16,799 ♀ (N50 = 745; 9,904,215 total bases), 24,351 ♂ (N50 = 464;
243 10,664,050 total bases), and 38,021 combined (N50 = 661; 20,752,384 total bases) contigs.
244 BUSCO estimations of completeness show the combined assembly out performs the sex specific
245 assemblies in capturing core single copy orthologs (% missing: 87 ♀; 91 ♂; 79 combined). These
246 completeness results likely have such high percentages of missing core orthologs due to
247 sequencing material from such highly specific venom gland tissues. Unless otherwise noted, the
248 combined transcriptome was used in the remainder of analyses. ORF predictions produced
249 23,624 putative peptides, and functional annotations were obtained for 8,075 out of 28,241
250 Trinity identified "genes" in the combined assembly, not to be confused with contigs that include
251 isoforms and alleles of individual genes.
252 A flow chart summary of steps we used to identify the set of transcripts ("entities") that
253 are venom-expressed (detected in the proteome) and dimorphic is presented in Figure 1B. We

254 use "entities" to refer to the distinct unique sets of venom components, and define entities as the
255 total number of clusters (homologous groups) and singleton sequences. MCL analysis identified
256 10,581 entities in the combined transcripts. Of these 87% are dimorphic - 53% are found only in
257 males and 47% in females. GFOLD analysis identified 3,800 (out of 23,634 – 16.07%)
258 differentially expressed transcripts, indicating sexual dimorphisms in a wide range of gene
259 including, but not limited to, venom cocktail peptides/proteins (1,430 female upregulated; 2,370
260 male upregulated).

261 ***Proteome*** Crude venom separations of male and female venoms show that the profiles of
262 expressed proteins are quite different (Figure 1C), evidence that is supported by proteomics. LC-
263 MS/MS produced 3,205 spectra that corresponded to 62 distinct proteins in 31 clusters, which
264 correspond to 31 entities (Figure 1B), only 0.29 % of total entities. Of these, nine are male-
265 specific, non-metabolic proteins, at least eight of which have no significant homology ($e \leq 10^{-5}$)
266 with any sequences in databases searched including NCBI and Arachnoserver (Figure 1C, Table
267 1A).

268 The results below detail the entities that were dimorphic and confirmed as venom
269 components through detection in proteomes.

270 ***Cluster 6*** The most abundant, dimorphic cluster, the sixth most highly represented in the
271 combined transcriptome (66 homologous polypeptides, 38,922 mapped reads), included the
272 proteins with the highest number of distinct polypeptides (12) and represent 30.7 % of MS
273 spectra in the male proteome (Table 1). Transcripts in this cluster appear to code for proteins of
274 sizes consistent with the large proteins unique to male venoms (Fig 1C). While none of the
275 assembled transcripts are full length (initiating methionine through stop codon) individual
276 transcripts in this cluster translate into polypeptides ranging from 41-1093 aa. Multiple sequence

PeerJ Preprints NOT PEER-REVIEWED

277 alignment of these polypeptides generates a consensus sequence of 1158 aa in length, and the
278 longest single transcript in the alignment (1093 aa) has a predicted MW of 128.95 kDa. The
279 homologs in this cluster are grouped by MCL into eight sub-clusters and five singletons (Figure
280 2).

281 Proteins in cluster 6 contain multiple repeating units, each of which has ~150 aa and a
282 conserved pattern of 10 cysteine residues (Figure 3). Submission of a multiple alignment of these
283 repeats to the Fold and Function Assignment (FFAS) server (http://ffas.sanfordburnham.org/ffas-
284 cgi/cgi/ffas.pl; PMID 15980471) returns strong evidence of distant homology to Argos, a 223-
285 residue antagonist of epidermal growth factor receptor signaling with a known structure (PDB ID
286 3c9a). Argos contains 3 small β-sheet rich domains, the first two of which correspond to one of
287 the sequence repeats present in the cluster 6 proteins, with a similar pattern of 10 cysteine
288 residues making five disulfide bonds (Figure 3B). The third Argos domain corresponds to an
289 extra half-repeat with six of the 10 cysteines and three disulfide bonds (Figure 3B). Argos uses
290 contacts from multiple domains to bind and sequester small protein ligands of EGF receptors.
291 Based on the FFAS score (-15), it is quite likely that the cluster 6 proteins have a similar fold,
292 disulfide pattern, and domain organization to Argos; however, the sequence homology to Argos
293 is so distant (<20% sequence identity between Argos domains 1 and 2 and any sequence repeat
294 in the *Tetragnatha* proteins as shown in Figure 3A) that a functional similarity is much less
295 certain.

296 ***Other male-specific venom proteins*** Seven additional male-specific clusters include
297 transcripts that code for polypeptides that span 20-37 kDa. These are sizes that correspond to
298 predicted molecular weights of full-length proteins within each cluster (Fig 1C). Cluster numbers
299 are labeled next to individual bands in Figure 1C and correspond to the relative rankings based

300 on representation in the combined transcriptome: cl-2448 (>36 kDa), cl-217 (37 kDa), cl-218 (35
301 kDa), cl-1113 (26 kDa), cl-9248 (>26 kDa), cl-7374 (>24 kDa), and cl-2111 (21 kDa). All were
302 detected in the proteome, though not as abundantly as cluster 6. Polypeptides encoded by genes
303 in all of these clusters correspond to novel gene families without homologous sequences present
304 in existing databases, including GenBank and Arachnoserver. Proteins in each of these clusters
305 have multiple cysteine residues (8-21), and with the exception of clusters 217 and 218, they have
306 at least two CxC motifs.

307 In addition to identifying novel families of proteins in these venoms, we also identified
308 several sequences with evidence of homology to known enzymes. Venom polypeptides in the
309 male venom hit three large peptidases: M13 metalloendopeptidase (specifically neprilysin, ~90
310 kDa), M14 carboxypeptidase M (sequence not full length, but estimated to be >55 kDa), and S8
311 peptidase (specifically neuroendocrine convertase, ~72 kDa). Only one large peptidase - M13
312 metalloendopeptidase - was identified in the venom proteome of females; however, homologous
313 sequences to each were found in both transcriptomes. Both S8 and M14 peptidase family
314 members are known to be involved with activation and processing/regulation of hormones,
315 respectively, whereas M13 peptidases are comprised of GluZincins, a superfamily of peptidases
316 that act on molecules <~40 aa (MEROPS peptidase database,
317 http://merops.sanger.ac.uk/index.shtml) and have also been reported in venom of a trapdoor
318 spider (Undheim et al., 2013). The degradative enzymes hyaluronidase and phospholipase A2
319 (PLA2) were identified in the venom of females, constituting ~1% of the proteome. Cluster 19
320 contains peptides corresponding to various hemocyanin subunits, and represents the most
321 abundant set of proteins detected in the female venom proteome (Table 1B). Each subunit varies
322 in size, and the most prominent was subunit G, which is predicted to be ~72 kDa (Figure 1C).

323 The female venom is also rich in small cysteine-rich peptides, corresponding to clusters 2335,
324 846, and 8293 identified in the proteome (Table 1A).

325 Within gene families observed only in male proteomes, we recovered interesting patterns
326 of sexually dimorphic expression and potential functions. Despite some mRNA from females
327 mapping to these transcripts, their peptides were not detected in the female proteome. This could
328 be due to a lack of translation following transcription or perhaps the proteins are not present in
329 the venom. GO enrichment analyses performed on all male-only peptides, based on the
330 proteomics analysis, indicated potential non-feeding functions of these proteins. Four GO terms
331 were significantly enriched, two relating to hormone functions – GO:0016486 (BP peptide
332 hormone processing; FDR p-value = 0.0143), GO:0008237 (MF metallopeptidase activity; FDR
333 p-value = 0.0349), GO:0042445 (BP hormone metabolic process; FDR p-value= 0.0349), and
334 GO:0006518 (BP peptide metabolic process; FDR p-value= 0.0497). Half of the enriched GO
335 terms were specifically associated with hormone functions, supporting our hypothesis that this
336 venom-based sexual dimorphism could be involved in sexual communication.

337

338 **DISCUSSION**

339 The results we present provide a first identification and characterization of unique and sexually
340 dimorphic components in venoms of *Tetragnatha versicolor*. Combined proteomics and
341 transcriptomics identify proteins that are expressed in venoms, and confirm the presence of
342 sexually dimorphic expression of particular components. The majority of components we
343 identify are sufficiently different from proteins in databases to be able to annotate with
344 confidence using homology searching. Recovering a low annotation percentage and high protein-
345 coding compliment (in terms of genes and isoforms) are both consistent with previous genomic

346 studies of spider taxa (Croucher et al., 2013; Sanggaard et al., 2014; Brewer et al., 2014) and
347 illustrates the early nature and promise of spider genome biology. Additionally, it is likely that
348 tetragnathid spiders will have many novel genes and gene families, as this family has not
349 previously been the subject of deep sequencing efforts.

350 Interestingly, this study shows that the majority of proteins identified in the *T. versicolor*
351 venom proteome are sexually dimorphic (~87 %), with 4,908 distinct proteome entities only in
352 adult males (Figure 1). Most of these had no detected corresponding transcripts in female venom
353 gland tissues and are not present in female venom cocktails. However, a small number of
354 sequence reads from females map to transcripts of male-only peptides, (and vice versa)
355 indicating these may be expressed in females but not translated or not incorporated into the
356 female venom cocktail. Of the "high molecular weight", male-only proteins in the venom
357 proteome, only three of 23 corresponding transcripts are not differentially expressed between the
358 sexes, as indicated by non-significant GFOLD values. While the males have more unique
359 components, there are 4,269 unique female proteome entities, including small number of unique
360 low molecular weight peptides.

361 The unique peptides in female venoms are homologous to other spider venom peptides,
362 range in size from 5.9-7.9 kDa, and have ICK motifs (–C6C-CC-C-C-) that are consistent with
363 them functioning as neurotoxins involved in prey immobilization. The biased presence of
364 potentially toxic peptides in female venoms is consistent with observations of differences
365 between males and females in concentration of low molecular weight components (Binford,
366 Gillespie & Maddison, 2016, Figure 1C). This pattern mirrors within sex, among species
367 differences in Hawaiian *Tetragnatha* that have evolved differences in feeding biology.
368 Specifically, as part of an adaptive radiation within Hawaiian *Tetragnatha,* a clade lost web-

369 building behavior and evolved to be wandering foragers with an associated shift in dietary niche.
370 With that evolutionary transition to wandering the lineage underwent a coincident reduction in
371 low molecular weight venom peptides (Binford, 2001). Given that evolutionary shifts in low
372 molecular weight peptides in venoms appear to occur in association with shifts in feeding
373 biology, the lack of detection of venom peptides in males may be best explained by differences
374 in adult niche that lead to a reduction in male reliance on venom peptides for prey
375 immobilization (Binford, Gillespie & Maddison, 2016).

376 The more striking dimorphism that is less easily explainable by differences in dietary
377 niche is in the male specific novel polypeptides, the "cluster 6" proteins. These belong to a single
378 gene family with estimated molecular weights corresponding to proteins detected with 1-D
379 protein gels across a comparative sampling of *Tetragnatha* (Fig 1, (Binford, Gillespie &
380 Maddison, 2016b)). The rationale for suggesting a possible role beyond feeding is that these
381 "cluster 6" components comprise a high proportion of the male-specific proteins (12 of 23
382 unique polypeptides; next largest family 2 of 23; Table 1A), suggesting an important functional
383 role unique to males. While BLAST searches of these male-specific proteins did not detect
384 homology to known sequences, predicted structural homology to Argos proteins that bind
385 ligands to epidermal growth factors suggests potential for binding to small proteins. Moreover,
386 functional annotations and GO enrichment analyses suggest hormone-related functions, either of
387 which could feasibly play a role in sexual signaling/mate recognition. Due to the degree of
388 similarity in motifs, and likely homology, between Argos and the novel male-only "cluster 6"
389 proteins discovered in *T. versicolor*, we propose the name *Argonuiino* (*Argo* for the Argos
390 protein, *nui* is Hawaiian for "large", and *ino* is Hawaiian for "venom") for this high-molecular
391 weight venom gene family that may function in sexual biology.

392 The high molecular weight components in male *T. versicolor* venom may be
393 characteristic of the genus *Tetragnatha*, based on previous 1-D gel studies of venom peptide
394 diversity (Binford, Gillespie & Maddison, 2016b). The origin of the male-specific proteins
395 appears to have been coincidental with the origin of the unusual premating cheliceral-locking
396 behavior by which these spiders intertwine their fangs while mating (Fig 1a). Members of the
397 family Tetragnathidae are secondarily haplogyne, having lost much of the complexity in male
398 and female genitalia that often functions in maintaining species boundaries and mate recognition,
399 evolving via sexual selection by female choice. The presence of cheliceral locking during mating
400 in many tetragnathid species provides an alternative mechanism upon which sexual selection and
401 mate recognition may have evolved. This has been demonstrated in the tetragnathid species
402 *Leucauge venusta* where tactile stimulation of females is accomplished via specialized male
403 setation during cheliceral locking (Aisenberg, Barrantes & Eberhard, 2015). Herein, we propose
404 a similar mechanism in the genus *Tetragnatha* using chemical communication
405 and/orstimmulation rather than, or in addition to, mechanical interactions. The hints of function
406 emerging from similarity to protein-binding proteins (Argos) and/or hormone-related functions
407 indicate these polypeptides may have evolved functions related to this sexual interaction.

408 There are two pieces of evidence that support the sexual roles for the high molecular
409 weight venom components. First, the possibility that the two groups of high molecular weight
410 polypeptides may be involved in communicating sexual state, availability, or identity, is
411 reinforced by the significant homology of these components to hormone processing peptidases
412 (BLAST e-value ~ 0.0), which suggests that these components may be involved in mate
413 recognition. If this were the case, then we would expect to find these high molecular weight
414 polypeptides in all spiders that show cheliceral-locking behavior, but not in those without; while

415 preliminary data for a small number of Hawaiian and mainland *Tetragnatha* support this
416 hypothesis (Binford, Gillespie & Maddison, 2016a), clearly more data are needed. Second, if the
417 high molecular weight components in the venom are playing a sexual function, then we might
418 expect that recognition could be compromised at some level in taxa that display cheliceral-
419 locking behavior but do not have high molecular weight polypeptides. Here again, an intriguing
420 observation in support of this argument is the finding of prolonged male-to-male cheliceral
421 locking and mating attempts (Gillespie, 1991) in the tetragnathid spider *Doryonychus raptor*, a
422 species which lacks dimorphic venom (Binford, Gillespie & Maddison, 2016a), suggesting that
423 recognition is less complete than in most other taxa that employ this mating strategy. Thus, the
424 high molecular weight polypeptides in males may serve to signal sexual identity and reduce the
425 chances of wasted mating opportunities. Chemical signals to reduce same sex mating attempts
426 have been documented in arthropod taxa (Scharf & Martin, 2013), including orthopterans
427 (repelling pheromone (Seidelmann & Ferenz, 2002)), hemipterans (alarm pheromone, (Ryne,
428 2009)), and dipterans (inhibitory pheromones, (Curcillo & Tompkins, 1987) and (Schlein, Galun
429 & Ben-Eliahu, 1981)). If demonstrated, *Tetragnatha* spiders could be the first non-insect
430 arthropods to utilize chemical signals and the first arthropod to employ venoms to reduce same
431 sex mating attempts.
432
433 **CONCLUSIONS**
434 While the results are still preliminary, we have shown that the venom of *Tetragnatha*
435 spiders can potentially serve in in both mate recognition as well as adaptive specialization for
436 prey, in which case venom could provide insights into the genomic underpinnings of adaptive
437 radiation as well as the interplay between plasticity and variability in fostering species

438 proliferation. *T. versicolor* displays a striking sexual dimorphism in venom compositions,
439 including proteins and peptides comprising a wide array of molecular weights, as well as many
440 novel compounds. A high molecular weight gene family, deemed *Argonuiino*, is associated with
441 gene ontologies relating to hormone processing and regulation and have FFAS indicated
442 homology to the *Drosophila* protein *Argos*, an epidermal growth factor associated protein. A
443 related species that lacks the dimorphic components engages in prolonged male-to-male mating
444 attempts, and EGF manipulation has been shown to elicit reproductive receptivity in female
445 rodents, supporting our hypothesis of sexual communication related functions of these male-
446 specific proteins.

447

448 **ACKNOWLEDGEMENTS**

449 The authors would like to thank David and Diana Binford for help collecting on the farm; Linda
450 Breci and Cynthia David for proteomic work at the RDI Analytical/Biological Mass
451 Spectrometry Core, University of Arizona; and Chase Magsig, Aayushi Patel, and Augustus
452 Floyd for help collecting North Carolina Spiders.

453
454 **REFERENCES**

455 Aisenberg A, Barrantes G, Eberhard WG 2015. Hairy kisses: tactile cheliceral courtship affects
456 female mating decisions in Leucauge mariana (Araneae, Tetragnathidae). *Behavioral*
457 *ecology and …*. DOI: 10.1007/s00265-014-1844-2.

458 Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths Jones S, Khanna A, Marshall M,
459 Moxon S, Sonnhammer EL 2004. The Pfam protein families database. *Nucleic Acids*
460 *Research* 32:D138–D141.

461 Berns CM 2013. The Evolution of Sexual Dimorphism: Understanding Mechanisms of Sexual

462 Shape Differences N2 . In: Moriyama H ed. *Sexual Dimorphism*. Sexual Dimorphism.
463 Rijeka: InTech, Ch. 01. DOI: 10.5772/55154.
464 Binford GJ, Wells MA 2003. The phylogenetic distribution of sphingomyelinase D activity in
465 venoms of Haplogyne spiders. *Comparative Biochemistry and Physiology Part B:*
466 *Biochemistry and Molecular Biology* 135:25–33. DOI: 10.1016/S1096-4959(03)00045-9.
467 Binford GJ, Gillespie RG, Maddison WP 2016a. Sexual dimorphism in venom chemistry in
468 Tetragnatha spiders is not easily explained by adult niche differences. *Toxicon* 114:45–52.
469 Binford GJ, Gillespie RG, Maddison WP 2016b. Sexual dimorphism in venom chemistry in
470 Tetragnatha spiders is not easily explained by adult niche differences. *Toxicon* 114:45–52.
471 DOI: 10.1016/j.toxicon.2016.02.015.
472 Blackledge TA, Gillespie RG 2004. Convergent evolution of behavior in an adaptive radiation of
473 Hawaiian web-building spiders. *Proceedings of the National Academy of Sciences*
474 101:16228–16233. DOI: 10.1073/pnas.0407395101.
475 Bolger AM, Lohse M, Usadel B 2014. Trimmomatic: a flexible trimmer for Illumina sequence
476 data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.
477 Brewer MS, Carter RA, Croucher PJP, Gillespie RG 2014. Shifting habitats, morphology, and
478 selective pressures: Developmental polyphenism in an adaptive radiation of Hawaiian
479 spiders. *Evolution* 69:162–178. DOI: 10.1111/evo.12563.
480 Croucher PJP, Brewer MS, Winchell CJ, Oxford GS, Gillespie RG 2013. De novo
481 characterization of the gene-rich transcriptomes of two color-polymorphic spiders, Theridion
482 grallator and T. californicum (Araneae: Theridiidae), with special reference to pigment genes.
483 *Bmc Genomics* 14:862. DOI: 10.1186/1471-2164-14-862.
484 Curcillo PG, Tompkins L 1987. The ontogeny of sex appeal inDrosophila melanogaster males.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3358v1 | CC BY 4.0 Open Access | rec: 20 Oct 2017, publ: 20 Oct 2017

485 *Behavior genetics* 17:81–86.
486 D'Suze G, Sandoval M, Sevcik C 2015. Characterizing Tityus discrepans scorpion venom from a
487 fractal perspective: Venom complexity, effects of captivity, sexual dimorphism, differences
488 among species. *Toxicon* 108:62–72. DOI: 10.1016/j.toxicon.2015.09.034.
489 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
490 TR 2012. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. DOI:
491 10.1093/bioinformatics/bts635.
492 Elias DO, Mason AC 2014. The role of wave and substrate heterogeneity in vibratory
493 communication: practical issues in studying the effect of vibratory environments in
494 communication. *Studying vibrational communication*. DOI: 10.1007/978-3-662-43607-3_12.
495 Enright AJ, Van Dongen S, Ouzounis CA 2002. An efficient algorithm for large-scale detection
496 of protein families. *Nucleic Acids Research* 30:1575–1584.
497 Escoubas P 2006. Molecular diversification in spider venoms: A web of combinatorial peptide
498 libraries. *Molecular Diversity* 10:545–554. DOI: 10.1007/s11030-006-9050-4.
499 Feng J, Meyer CA, Wang Q, Liu JS, Shirley Liu X, Zhang Y 2012. GFOLD: a generalized fold
500 change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics*
501 28:2782–2788. DOI: 10.1093/bioinformatics/bts515.
502 Finn RD, Clements J, Eddy SR 2011. HMMER web server: interactive sequence similarity
503 searching. *Nucleic Acids Research*:gkr367.
504 Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, Nevalainen TJ, Norman
505 JA, Lewis RJ, Norton RS, Renjifo C, la Vega de RCR 2009. The Toxicogenomic Multiverse:
506 Convergent Recruitment of Proteins Into Animal Venoms. *Annual Review of Genomics and*
507 *Human Genetics* 10:483–511. DOI: 10.1146/annurev.genom.9.081307.164356.

508 Gillespie R 2004a. Community assembly through adaptive radiation in Hawaiian spiders.
509 *Science* 303:356–359.
510 Gillespie R 2004b. Community assembly through adaptive radiation in Hawaiian spiders.
511 *Science* 303:356–359. DOI: 10.1126/science.1091875.
512 Gillespie RG 1987. The Mechanism of Habitat Selection in the Long-Jawed Orb-Weaving
513 Spider Tetragnatha elongata (Araneae, Tetragnathidae). *Journal Of Arachnology* 15:81–90.
514 DOI: 10.2307/3705511?ref=no-x-route:542c24af2497b197f3b3f4b547bf353c.
515 Gillespie RG 1991. Hawaiian spiders of the genus Tetragnatha: I. Spiny leg clade. *Journal Of*
516 *Arachnology* 19:174–209.
517 Gillespie RG 2005. Evolution of Hawaiian Spider Communities. *American Scientist* 93:122–131.
518 Gray MR, Sutherland SK 1978. Venoms of Dipluridae. In: *Arthropod venoms*. Springer, 121–
519 148.
520 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D,
521 Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman
522 R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A 2013. De novo
523 transcript sequence reconstruction from RNA-seq using the Trinity platform for reference
524 generation and analysis. *Nature Protocols* 8:1494–1512. DOI: 10.1038/nprot.2013.084.
525 Herzig V, Khalife AA, Chong Y, Isbister GK, Currie BJ, Churchill TB, Horner S, Escoubas P,
526 Nicholson GM, Hodgson WC 2008. Intersexual variations in Northern (Missulena pruinosa)
527 and Eastern (M. bradleyi) mouse spider venom. *Toxicon* 51:1167–1177. DOI:
528 10.1016/j.toxicon.2008.02.001.
529 Isbister GK, Gray MR 2004. Bites by Australian mygalomorph spiders (Araneae,
530 Mygalomorphae), including funnel-web spiders (Atracinae) and mouse spiders

531 (Actinopodidae: Missulena spp). *Toxicon* 43:133–140. DOI: 10.1016/j.toxicon.2003.11.009.
532 King G 2015. Venoms to drugs: venom as a source for the development of human therapeutics.
533 Royal Society of Chemistry.
534 Krogh A, Larsson B, Heijne von G, Sonnhammer ELL 2001. Predicting transmembrane protein
535 topology with a hidden markov model: application to complete genomes11Edited by F.
536 Cohen. *Journal of molecular biology* 305:567–580. DOI: 10.1006/jmbi.2000.4315.
537 Kuhn-Nentwig L, Stöcklin R 2011. Venom composition and strategies in spiders: is everything
538 possible? *Advances in Insect Physiology* 60:1–86. DOI: 10.1016/B978-0-12-387668-
539 3.00001-5.
540 Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW 2007. RNAmmer:
541 consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research* 35:3100–
542 3108.
543 Langmead B, Salzberg SL 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*
544 9:357–359. DOI: 10.1038/nmeth.1923.
545 Langmead B, Trapnell C, Pop M, Salzberg SL 2009. Ultrafast and memory-efficient alignment
546 of short DNA sequences to the human genome. *Genome biology* 10:R25. DOI: 10.1186/gb-
547 2009-10-3-r25.
548 Levi HW 1981. American orb-weaver genera Dolichognatha and Tetragnatha north of Mexico
549 (Araneae: Araneidae, Tetragnathinae).
550 Magoč T, Salzberg SL 2011. FLASH: fast length adjustment of short reads to improve genome
551 assemblies. *Bioinformatics* 27:2957–2963. DOI: 10.1093/bioinformatics/btr507.
552 Martin M 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.
553 *EMBnet journal* 17:pp–10.

554 Menezes MC, Furtado MF, Travaglia-Cardoso SR, Camargo ACM, Serrano SMT 2006. Sex-
555 based individual variation of snake venom proteome among eighteen Bothrops jararaca
556 siblings. *Toxicon* 47:304–312. DOI: 10.1016/j.toxicon.2005.11.007.
557 Miller DW, Jones AD, Goldston JS, Rowe MP, Rowe AH 2016. Sex Differences in Defensive
558 Behavior and Venom of The Striped Bark Scorpion Centruroides vittatus(Scorpiones:
559 Buthidae). *Integrative and Comparative Biology* 56:1022–1031. DOI: 10.1093/icb/icw098.
560 Polis GA, Sissom WD 1990. Life history. *The biology of scorpions*:161–223.
561 Quinlan AR, Hall IM 2010. BEDTools: a flexible suite of utilities for comparing genomic
562 features. *Bioinformatics* 26:841–842. DOI: 10.1093/bioinformatics/btq033.
563 Rash LD, King RG, Hodgson WC 2000. Sex differences in the pharmacological activity of
564 venom from the white-tailed spider (Lampona cylindrata). *Toxicon* 38:1111–1127.
565 Ryne C 2009. Homosexual interactions in bed bugs: alarm pheromones as male recognition
566 signals. *Animal Behaviour* 78:1471–1475.
567 Sanggaard KW, Bechsgaard JS, Fang X, Duan J, Dyrlund TF, Gupta V, Jiang X, Cheng L, Fan D,
568 Feng Y, Han L, Huang Z, Wu Z, Liao L, Settepani V, gersen IBTO, Vanthournout B, Wang
569 T, Zhu Y, Funch P, Enghild JJ, Schauser L, Andersen SU, Villesen P, Schierup MH, Bilde T,
570 Wang J 2014. Spider genomes provide insight into composition and evolution of venom and
571 silk. *Nature Communications* 5:1–11. DOI: 10.1038/ncomms4765.
572 Scharf I, Martin OY 2013. Same-sex sexual behavior in insects and arachnids: prevalence,
573 causes, and consequences. *Behavioral Ecology and Sociobiology* 67:1719–1730.
574 Schlein Y, Galun R, Ben-Eliahu MN 1981. Receptors of sex pheromones and abstinons inMusca
575 domestica andGlossina morsitans. *Journal of chemical ecology* 7:291–303.
576 Seidelmann K, Ferenz H-J 2002. Courtship inhibition pheromone in desert locusts, Schistocerca

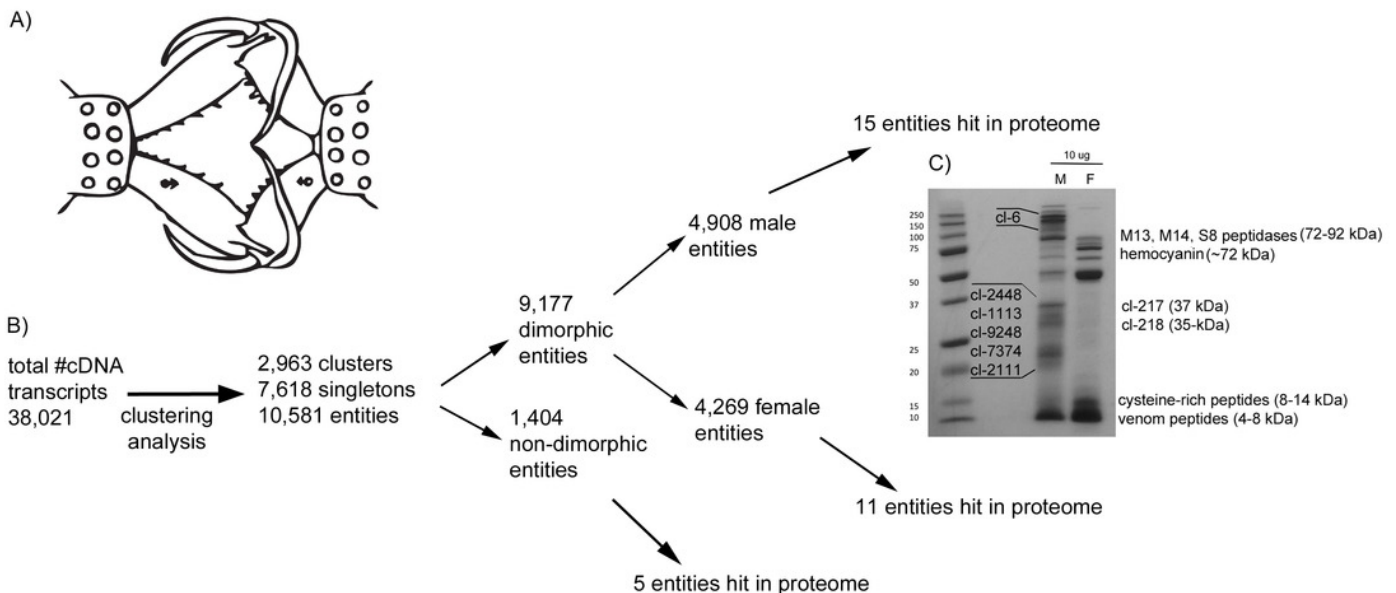
577 gregaria. *Journal of Insect Physiology* 48:991–996.
578 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B,
579 Ideker T 2003. Cytoscape: a software environment for integrated models of biomolecular
580 interaction networks. *Genome Research* 13:2498–2504.
581 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM 2015. BUSCO:
582 assessing genome assembly and annotation completeness with single-copy orthologs.
583 *Bioinformatics* 31:3210–3212. DOI: 10.1093/bioinformatics/btv351.
584 Smith JJ, Herzig V, King GF, Alewood PF 2013. The insecticidal potential of venom peptides.
585 *Cellular and Molecular Life Sciences* 70:3665–3693. DOI: 10.1007/s00018-013-1315-3.
586 Undheim EAB, Sunagar K, Herzig V, Kely L, Low DHW, Jackson TNW, Jones A, Kurniawan N,
587 King GF, Ali SA, Antunes A, Ruder T, Fry BG 2013. A proteomics and transcriptomics
588 investigation of the venom from the Barychelid spider *Trittame loki* (brush-foot trapdoor).
589 *Toxins* 5:2488-2503.
590 World Spider Catalog, 2015. World spider catalog, version 16.
591 Wyatt TD 2014. *Pheromones and Animal Behavior*. Cambridge University Press.
592 Young MD, Wakefield MJ, Smyth GK, Oshlack A 2010. Gene ontology analysis for RNA-seq:
593 accounting for selection bias. *Genome Biology* 11:R14.
594
595

PeerJ Preprints NOT PEER-REVIEWED

Figure 1

Comparative analysis of *Tetragnatha* male and female venom gland transcriptomes

A) diagram shows cheliceral locking between male and female spiders during mating (courtesy Kaston, 1948). B) flowchart shows the pipeline of transcriptome clustering analysis yielding sexually dimorphic entities detected in both transcriptomes and proteomes. C) SDS-PAGE (12%) of *T. versicolor* crude venom from male (M) and female (F) spiders; proteins that correspond to identified clusters are labeled with “cl-“ followed by the cluster number and assigned molecular weight sizes are based on predictions made from full-length amino acid sequences using the compute pI/MW tool (http://web.expasy.org/compute_pi/).



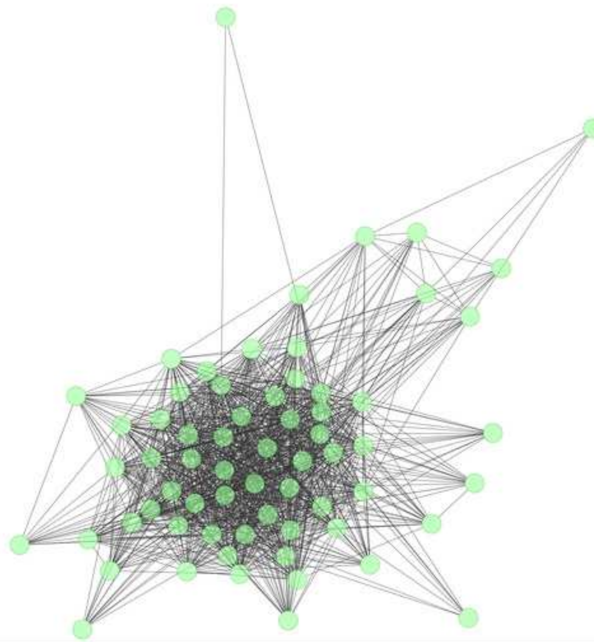
PeerJ Preprints NOT PEER-REVIEWED

Figure 2

High molecular weight, male-only “gene family” and subclustering results from MCL analysis

A) Similarities between components of the high-molecular weight family of male-specific components, and B) subclustering of the same family. All members are present in both networks. These components show no similarities to known venom or toxin genes, bug gene ontology (GO) enrichment tests indicate a role in hormone signaling/transport.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3358v1 | CC BY 4.0 Open Access | rec: 20 Oct 2017, publ: 20 Oct 2017



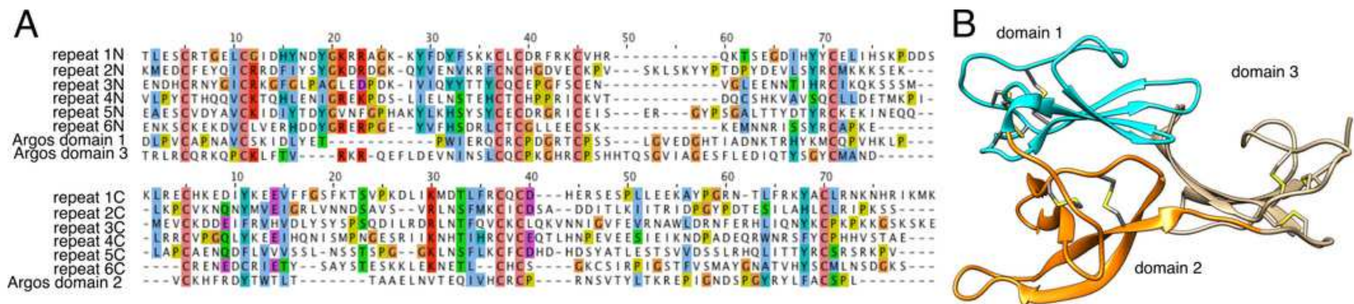
PeerJ Preprints NOT PEER-REVIEWED

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3358v1 | CC BY 4.0 Open Access | rec: 20 Oct 2017, publ: 20 Oct 2017

Figure 3

Remote homology of cluster 6 proteins to *Drosophila* Argos

A) Sequence alignments of Argos to each 6 repeats within a single cluster 6 protein from *Tetragnatha*; at top is the N-terminal half of each repeat aligned to domains 1 and 3 of Argos, while at bottom is the C-terminal half of each repeat aligned to domain 2 of Argos, B) Ribbon diagram of Argos structure (PDB ID 3c9a), colored by domain with disulfide bonds shown.



PeerJ Preprints NOT PEER-REVIEWED

Table 1(on next page)

Sexually dimorphic components of *T. versicolor* venom

Proteomics results are separated into four categories based on general function prediction hits from homology searches: unknown function, potentially toxic/defense proteins, potentially toxic peptides are presented in Panel A, and housekeeping/metabolism proteins are in panel B. Black boxes indicate an absence of a protein in the secreted venom. The symbol (•) indicates <1% of total #spectra. The total number of homologous polypeptides in each cluster is listed, along with the number of corresponding transcripts in the transcriptome. The symbol (^) indicates a hit in the tryptic peptide search against the NCBI database.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3358v1 | CC BY 4.0 Open Access | rec: 20 Oct 2017, publ: 20 Oct 2017

PeerJ Preprints | NOT PEER-REVIEWED

A.

general function prediction based on homology searches [cluster#]	#distinct polypeptides detected in venom		#spectra detected in proteome (% total)		# homologous polypeptides in transcriptome cluster [e<10^{-5}] (#transcripts)		top hit species in homology searches (e-value)
	m	**f**	**m**	**f**	**m**	**f**	
unknown function							
new family (high MW) [6]	12		398 (30.7)		66 (38922)	0 (35)	-
new family [1113]	2		25 (1.9)		3 (2639)		-
new family [9248]	1		10 (✐		1 (229)		-
new family [7374]	1		7 (✐		1 (661)		-
new family [2111]	1		33 (2.5)		2 (1382)		-
new family [2448]	1		8 (✐		2 (453)		-
new family [217]	1		16 (1.2)		7 (6492)	1 (195)	*Culex* (e10^{-5})
new family [218]	1		28 (2.2)		8 (15274)	2 (3)	-
S8 peptidase[889]	1		14 (1.1)		1 (1107)	3 (886)	*Stegodyphus* (e0.0)
M14 carboxypeptidase M [116]	1	1	6 (✐		7 (751)	6 (766)	*Stegodyphus* (e0.0)
M13 metalloendopeptidase [11]	5	1	501 (38.7)	80 (4.2)	48 (19836)	13 (11358)	*Stegodyphus* (e0.0)
potentially toxic/defense proteins							
chitinase [337]	1	1	77 (5.9)	19 (10.1)	6 (5557)	1 (2974)	*Araneus* (e0.0)
venom allergen 5 [843]	1	1	34 (2.6)	157 (8.2)	2 (5401)	2 (86981)	*Stegodyphus* (e10^{-99})
hyaluronidase [10277]		1		6 (✐	0 (62)	1 (690)	*Brachypelma* (e10^{-97})
phospholipase A2 (PLA2) [10291]		1		20 (1.0)	1 (95)	0 (73)	*Stegodyphus* (e10^{-52})
potentially toxic peptides							
venom peptide [2335]		1		22 (1.2)		2 (2043)	*Nephila* BLTX631 (e10^{-135})
venom peptide [846]		2		270 (14.1)	1 (139)	3 (43266)	*Nephila* BLTX100 (e10^{-24})
venom peptide [8293]		1		187 (9.8)	0 (40)	1 (13297)	-

B.

general function prediction based on homology searches [cluster#]	# distinct polypeptides detected in venom		# spectra detected in proteome (% total)		# homologous polypeptides in transcriptome cluster [e<10^{-5}] (#transcripts)		top hit species in homology searches (e-value)
	m	**f**	**m**	**f**	**m**	**f**	
housekeeping/metabolism proteins							
hemocyanin (subunits A,B,C,D,G) [19]		10		835 (43.7)	24 (6550)	14 (16316)	*Stegodyphus* (e0.0)
hemocyanin (subunit D)		1		32 (1.7)	n/a	n/a	*Latrodectus*^
alpha amylase [4700]		1		20 (1.0)	0 (41)	1 (1777)	*Stegodyphus* (e10^{-144})
alpha amylase [10595]		1		37 (1.9)	0 (2)	1 (514)	*Lithobius* (e10^{-98})
alpha-2 macroglobulin [453]	1		6 (✐		3 (1006)	4 (997)	*Hasarius* (e0.0)
acetylcholinesterase [831]	1		15 (1.1)		2 (9789)	1 (27)	*Pardosa* (e10^{-159})
triacylglycerol lipase [75]		1		33 (1.7)	9 (1636)	6 (14009)	*Stegodyphus* (e10^{-149})
protein tyrosine phosphatase rec. [231]	1	1	19 (1.5)	12 (✐	3 (5714)	5 (3671)	*Homo* (e10^{-70})
corticotropin releasing factor [4347]	1		10 (✐		1 (138)	1 (1002)	*Tribolium* (e10^{-82})
G-protein coupled rec. (GPRmth5) [117]	1		76 (5.9)		10 (2294)	1 (3)	*Pediculus* (e10^{-28})
beta casein	1		6 (✐		n/a	n/a	*Bos*^
casein	1		6 (✐		n/a	n/a	*Bos*^
slit-like protein (leu-rich domain) [6912]		1		7 (✐	0 (2)	1 (100)	*Stegodyphus* (e10^{-20})