Supplementary Material - file 1

The first venomous crustacean revealed by transcriptomics and functional morphology: remipede venom glands express a unique toxin cocktail dominated by enzymes and a neurotoxin

Research Article

Björn M. von Reumont*[1], Alexander Blanke[2], Sandy Richter[3], Fernando Alvarez[4], Christoph Bleidorn[3], Ronald A Jenner*[1]

The supplementary PDF files include:

Supplementary_file1: Supplementary Text, References, Supplementary Figures 1-2
Supplementary_file2: Supplementary Figures 3
Supplementary_file3: Supplementary Figures 4-6
Supplementary_file4: Supplementary Figures 7-9
Supplementary_file5: Supplementary Tables 1 and 2

Supplementary Text

Metalloproteinases

Metalloproteinases are a diverse group of proteolytic enzymes that are involved in many biological processes. Members of the M12 family of metalloproteinases are important components of snake venoms (Brust et al. 2013), especially in viperid snakes where they constitute the largest component of venom toxins (Casewell et al. 2009; Rokyta et al. 2011). A variety of metalloproteinases are also expressed in the venoms of several species of spiders, scorpions, centipedes, cephalopods, cone snails, the platypus, and in *Hydra* nematocysts (Fernandes-Pedrosa et al. 2008; Jiang and al. 2010; Ruiming et al. 2010; Whittington et al. 2010; Morgenstern et al. 2011; Undheim and King 2011; Almeida et al. 2012; Balasubramanian et al. 2012; Rendón-Anaya et al. 2012; Terrat et al. 2012; Ruder et al. 2013). The remipede venom glands express 6 distinct transcript types of venom-relevant metalloproteinases, with the M14 family represented by five transcripts, and the M13 familiy represented by a single transcript. Together these transcripts represent over 5.5 % of venom gland toxin diversity, and they comprise just over 1% of the total number of expressed toxin reads (Figure 2). Interestingly, we find in all reconstructed trees (Supplementary Figures 6a, 6b, and 6c) distinct clades that are composed of transcripts deriving from complete animal tissue. Given the different effects that metalloproteinases can have in different venoms, including degradation of extracellular matrix, prevention of blood clotting, inflammation, skin damage, and myonecrosis, and the roles they play in normal cellular processes, the precise roles of the remipede venom metalloproteinases remain unclear.

Cysteine Rich Secretory Proteins (CRISPs)

CRISPs are found in the venoms of a variety of animals, including cone snails, centipedes, hymenopterans, and the duck-billed platypus, but they are particularly important components of snake venoms (Fry et al. 2009; Cardoso et al. 2010; Undheim and King 2011). The hymenopteran venom allergens, such as antigen 5 (aka allergen 5) are also classified in this group. Venom CRISPs are known to have a variety of activities,

including ion channel blocking, the inhibition of smooth muscle contraction, and causing myonecrosis. The remipede venom glands express two distinct CRISP transcripts that together with all but one CRISP sequence from the complete animal library group together in a remipede-specific clade (Supplementary Figure 7a). This clade is sister group to a clade of hymenopteran venom allergens 5 and 3. Venom allergens are typical components of hymenopteran venoms, but they have recently also been detected in the venom of several species of centipedes (Liu et al. 2012). The pattern of cysteine residues in the remipede CRISPs has diverged, however, from that of the hymenopterans, and some of the remipede sequences also show several insertions not present in the hymenopterans sequences (see Supplementary Figure 7b). The biological role of the remipede venom CRISPs remains unknown.

Venom serine carboxypeptidases

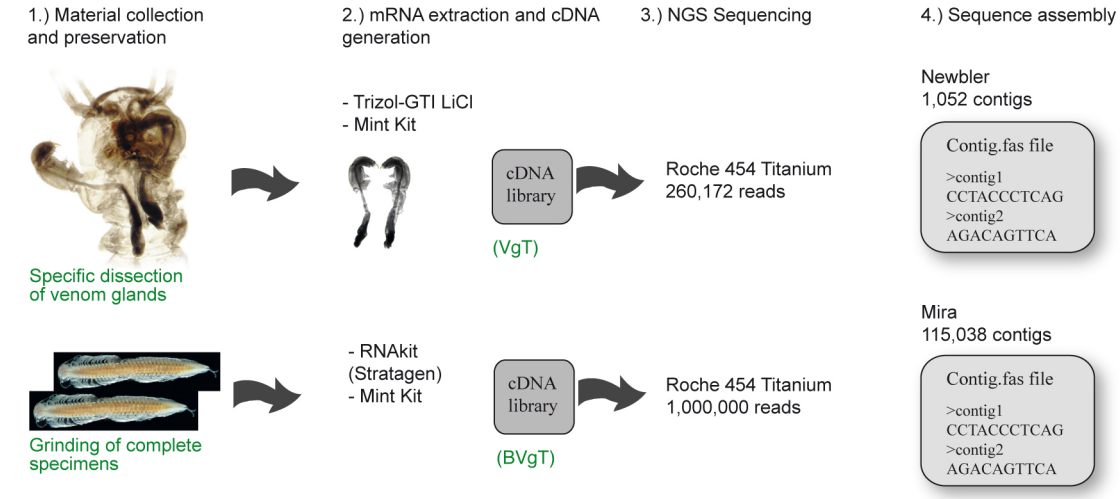
The third most highly expressed toxin in the remipede venom glands shows strong similarities to the venom serine carboxypeptidases found in several hymenopteran venoms (Supplementary Figure 8). They are members of the peptidase S10 family, and the remipede venom gland transcripts contain the conserved Ser-Asp-His catalytic triad characteristic of this family (alignment positions 105-476-541, see alignment data in DRYAD link given in the acknowledgements). The biological role of this toxin in remipede venom is unclear, but given that members of the peptidase S10 family are active at acidic pH, they may perform their function in the venom before it is injected into the prey.

Serine protease inhibitors (Serpin, Kunitz)

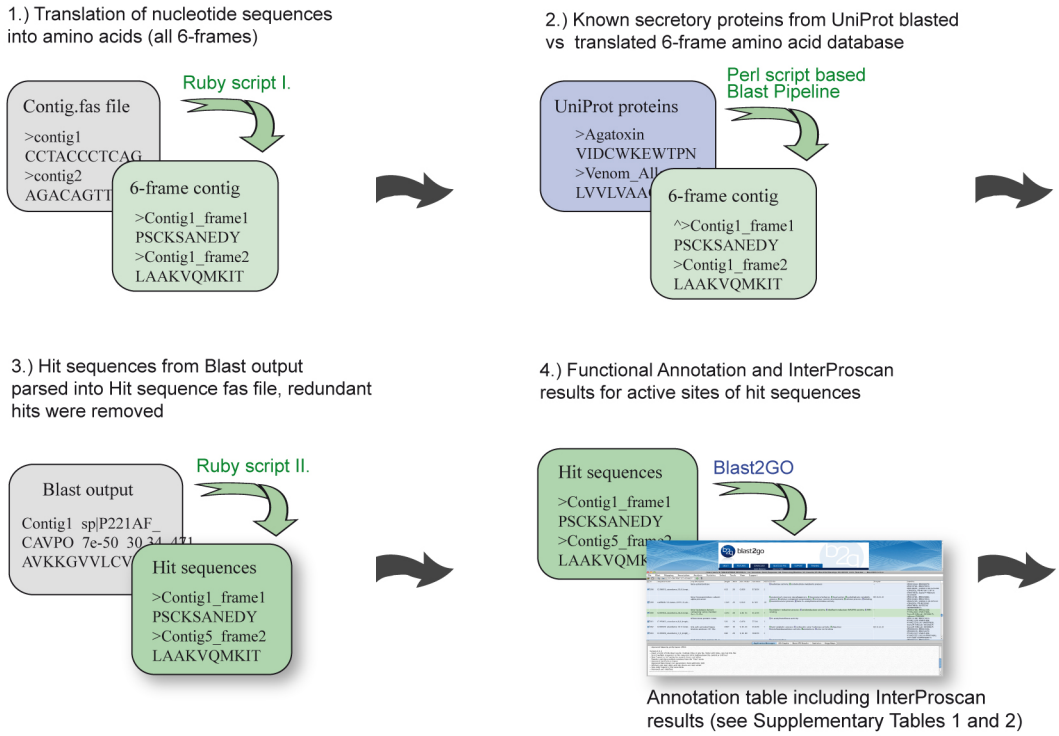
Serpins are serine protease inhibitors that have been found in a variety of venoms (especially the Kunitz-type serpins), including those of snakes, cnidarians, cone snails, platypus (Whittington et al. 2010; Balasubramanian et al. 2012; Terrat et al. 2012; Brust et al. 2013; Ruder et al. 2013), and hymenopterans (Asgari and Rivers 2011; Colinet et al. 2013), as well as the salivary secretions of hematophagous insects and leeches (Fry et al. 2009; Min et al. 2010). Depending on the type of serpin and the taxon in which it occurs they can have different effects, including ion channel blocking, interference with blood

coagulation, and causing hypertension. The remipede serine protease inhibitors are expressed at low levels in the venom glands. In addition to the venom gland transcripts a variety of different non-Kunitz serpin transcripts are expressed in the complete animal library, placed in three different clades in our tree (Supplementary Figure 9a). However, the support for these clades is extremely low, leaving the precise pattern of serpin diversification uncertain. The function of the remipede non-Kunitz serpins also remains unknown. One Kunitz-type serpin is expressed in the remipede venom gland. The nearest non-remipede sequence in our tree is a scorpion Kunitz toxin that has potassium channel inhibiting activity. However, there is no statistical support for the structure in our tree due to the great sequence diversity of Kunitz sequences (Supplementary Figure 9b). Consequently, the function of the remipede Kunitz remains unknown.

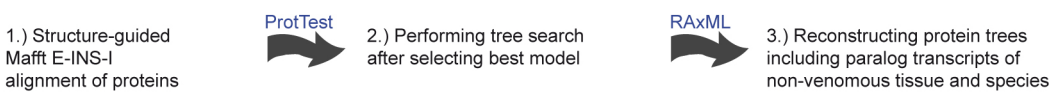
A.) Laboratory work and cDNA generation to gain assembled NGS data



B.) Processing pipeline to identify venom proteins in assembled transcriptome data

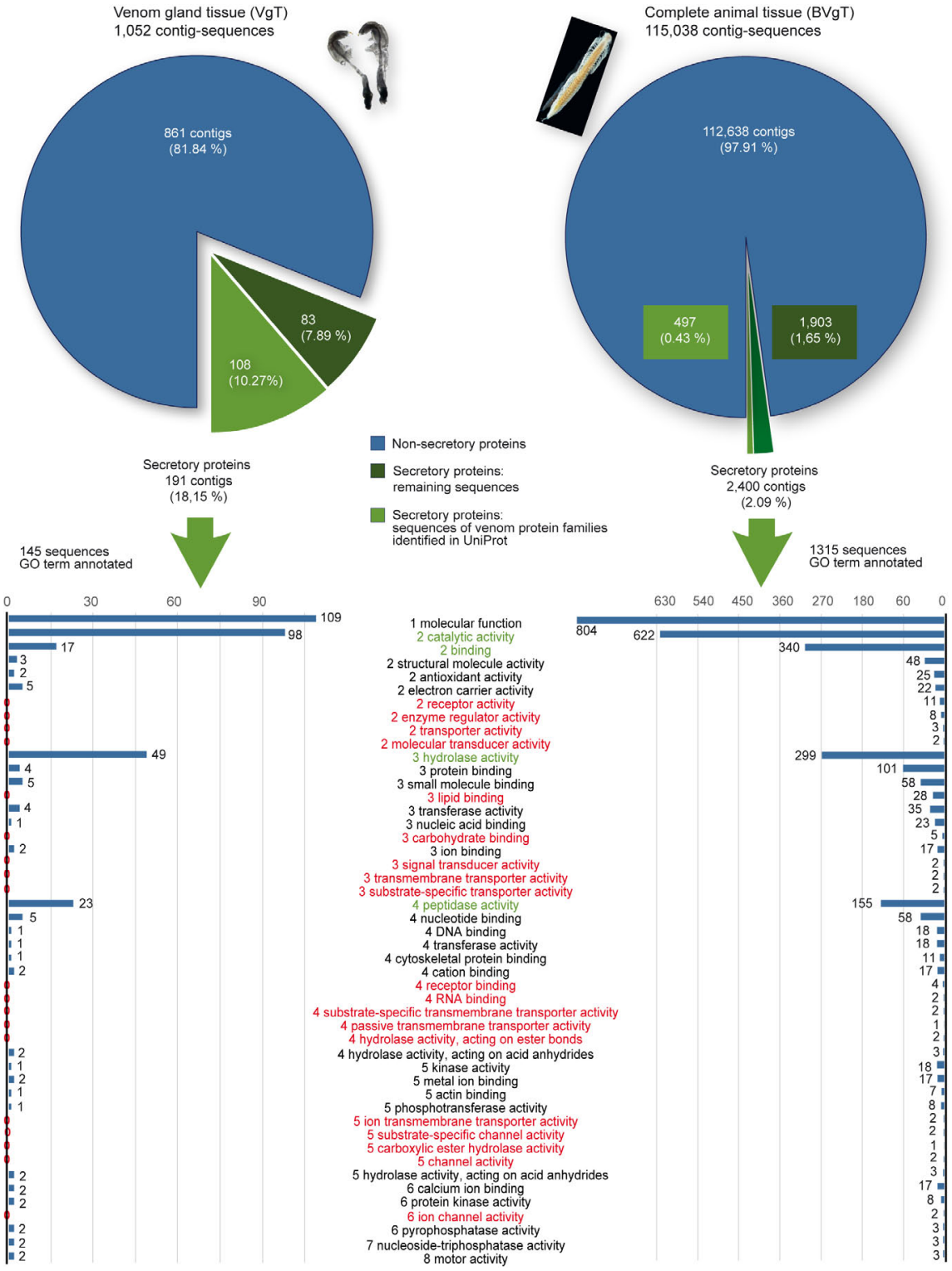


C.) Reconstruction of protein trees



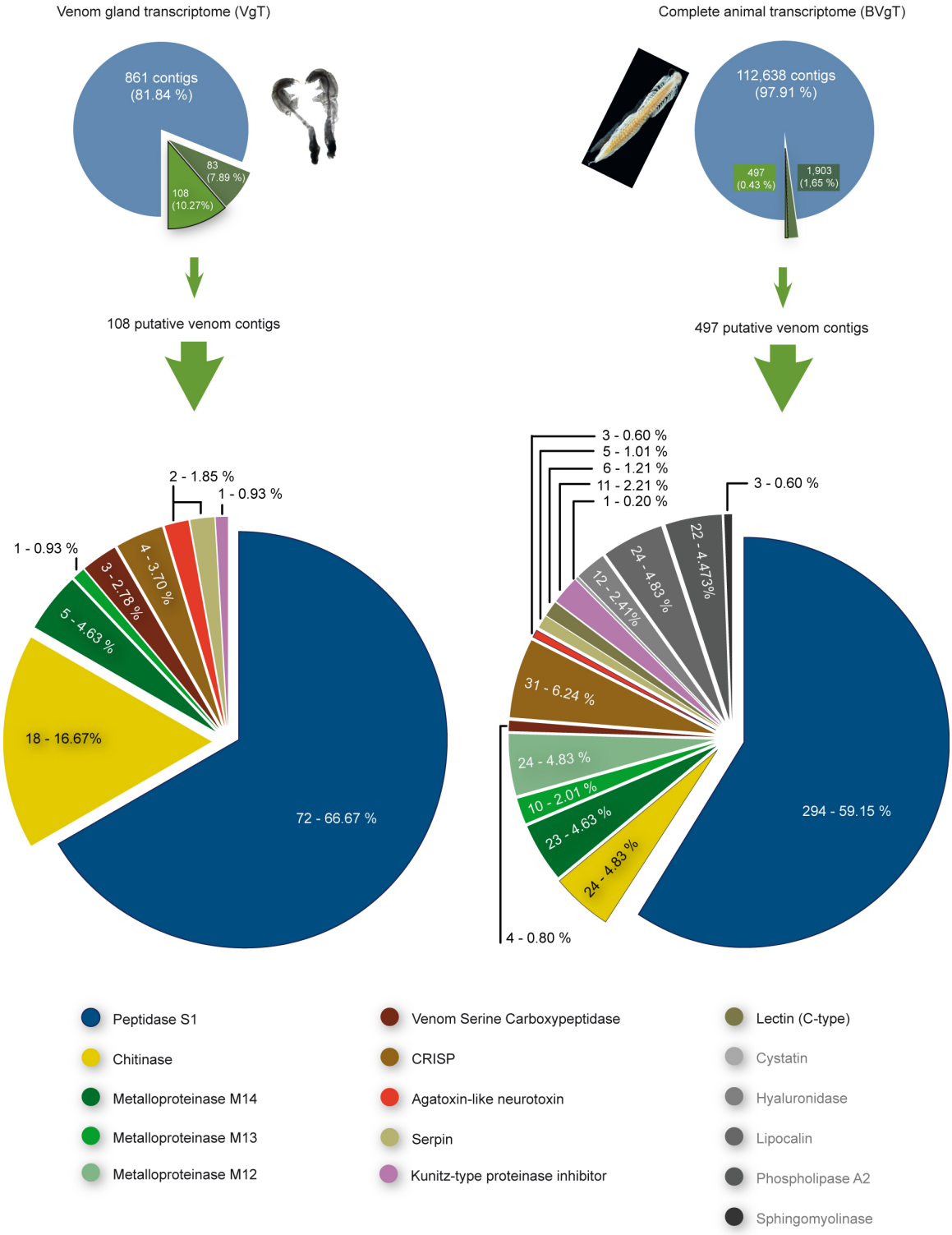
Supplementary Figure 1.

Methodological workflow used in this study. **A**.) Steps of the work done, from collecting material to the final assembly of cDNA transcripts. **B**.) Details of the processing pipeline developed to identify putative venom proteins. Scripts and software that were used are displayed at top of the arrows. **C**.) Reconstruction of phylogenetic trees to reveal the evolution of venom proteins.


Venom gland tissue (VgT)
1,052 contig-sequences
861 contigs
(81.84 %)
83
(7.89 %)
108
(10.27%)
Complete animal tissue (BVgT)
115,038 contig-sequences
112,638 contigs
(97.91 %)
497
(0.43 %)
1,903
(1,65 %)
Non-secretory proteins
Secretory proteins:
remaining sequences
Secretory proteins:
sequences of venom protein families
identified in UniProt
Secretory proteins
191 contigs
(18,15 %)
Secretory proteins
2,400 contigs
(2.09 %)
145 sequences
GO term annotated
1315 sequences
GO term annotated
0 30 60 90
630 540 450 360 270 180 60 0
109
98
17
3
2
5
49
4
5
4
1
2
23
5
1
1
1
2
2
1
2
1
1
2
2
2
2
2
2
2
1 molecular function
2 catalytic activity
2 binding
2 structural molecule activity
2 antioxidant activity
2 electron carrier activity
2 receptor activity
2 enzyme regulator activity
2 transporter activity
2 molecular transducer activity
3 hydrolase activity
3 protein binding
3 small molecule binding
3 lipid binding
3 transferase activity
3 nucleic acid binding
3 carbohydrate binding
3 ion binding
3 signal transducer activity
3 transmembrane transporter activity
3 substrate-specific transporter activity
4 peptidase activity
4 nucleotide binding
4 DNA binding
4 transferase activity
4 cytoskeletal protein binding
4 cation binding
4 receptor binding
4 RNA binding
4 substrate-specific transmembrane transporter activity
4 passive transmembrane transporter activity
4 hydrolase activity, acting on ester bonds
4 hydrolase activity, acting on acid anhydrides
5 kinase activity
5 metal ion binding
5 actin binding
5 phosphotransferase activity
5 ion transmembrane transporter activity
5 substrate-specific channel activity
5 carboxylic ester hydrolase activity
5 channel activity
5 hydrolase activity, acting on acid anhydrides
6 calcium ion binding
6 protein kinase activity
6 ion channel activity
6 pyrophosphatase activity
7 nucleoside-triphosphatase activity
8 motor activity
804
622
340
48
25
22
11
8
3
2
299
101
58
28
35
23
5
17
2
2
2
155
58
18
18
11
17
4
2
2
1
2
3
18
17
7
8
2
2
1
2
3
17
8
2
3
3
3

Supplementary Figure 2a

Comparison between venom gland and complete animal tissue gene expression for secreted proteins. On the left the composition of the transcripts derived from the venom gland tissue (VgT) is displayed in contrast to the complete animal tissue transcripts (BVgT) on the right. The molecular function for each sequence of the secreted protein fraction of both libraries is identified by GO annotation using BLAST2GO. Red fonts denotes transcripts absent for the particular GO-terms, black font denotes presence in both libraries and green highlights the most highly expressed transcripts. See Supplementary Table 1 and 2 for a complete description of each sequence of identified secretory proteins.


Venom gland transcriptome (VgT)
861 contigs
(81.84 %)
83
(7.89 %)
108
(10.27%)
108 putative venom contigs
2 - 1.85 %
1 - 0.93 %
1 - 0.93 %
3 - 2.78 %
4 - 3.70 %
5 - 4.63 %
18 - 16.67%
72 - 66.67 %
Complete animal transcriptome (BVgT)
112,638 contigs
(97.91 %)
497
(0.43 %)
1,903
(1,65 %)
497 putative venom contigs
3 - 0.60 %
5 - 1.01 %
6 - 1.21 %
11 - 2.21 %
1 - 0.20 %
3 - 0.60 %
22 - 4.473%
24 - 4.83 %
12 - 2.41%
31 - 6.24 %
24 - 4.83 %
10 - 2.01 %
23 - 4.63 %
24 - 4.83 %
4 - 0.80 %
294 - 59.15 %
Peptidase S1
Chitinase
Metalloproteinase M14
Metalloproteinase M13
Metalloproteinase M12
Venom Serine Carboxypeptidase
CRISP
Agatoxin-like neurotoxin
Serpin
Kunitz-type proteinase inhibitor
Lectin (C-type)
Cystatin
Hyaluronidase
Lipocalin
Phospholipase A2
Sphingomyolinase

Supplementary Figure 2b

Comparison and composition of identified venom proteins between venom gland and complete animal libraries. On the left the composition of the venom proteins derived from the venom gland tissue (VgT) is displayed in contrast to the complete animal tissue transcripts (BVgT) on the right. The contig numbers are given followed by the percentage for each of the venom protein family. The legend for colour-coded venom protein families is included.

References

Almeida DD, Scortecci KC, Kobashi LS, Agnez-Lima LF, Medeiros SRB, Silva-Junior AA, Junqueira-De-Azevedo I, Fernandes-Pedrosa M. 2012. Profiling the resting venom gland of the scorpion *Tityus stigmurus* through a transcriptomic survey. BMC Genomics 13:362.

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

Balasubramanian PG, Beckmann A, Warnken U, Schnolzer M, Schuler A, Bornberg-Bauer E, Holstein TW, Ozbek S. 2012. Proteome of *Hydra* nematocyst. J Biol Chem 287:9672-9681.

Brust A, Sunagar K, Undheim EAB, et al. (15 co-authors). 2013. Differential evolution and neofunctionalization of snake venom metalloprotease domains. Mol Cell Proteomics 12:651-663.

Cardoso KC, Da Silva MJ, Costa GGL, Torres TT, Del Bem LEV, Vidal RO, Menossi M, Hyslop S. 2010. A transcriptomic analysis of gene expression in the venom gland of the snake *Bothrops alternatus* (urutu). BMC Genomics 11:605.

Casewell NR, Harrison RA, Wüster W, Wagstaff SC. 2009. Comparative venom gland transcriptome surveys of the saw-scaled vipers (Viperidae: *Echis*) reveal substantial intra-family gene diversity and novel venom transcripts. BMC Genomics 10:564.

Colinet D, Deleury E, Anselme C, Cazes D, Poulain J, Azema-Dossat C, Belghazi M, Gatti JL, Poiré M. 2013. Extensive inter- and intraspecific venom variation in closely related parasites targeting the same host: the case of *Leptopilina* parasitoids of *Drosophila*. Insect Biochem Mol Biol 43:601-611.

Fernandes-Pedrosa M, Junqueira-De-Azevedo I, Gonçalves-De-Andrade R, Kobashi L, Almeida D, Ho P, Tambourgi D. 2008. Transcriptome analysis of *Loxosceles laeta* (Araneae, Sicariidae) spider venomous gland using expressed sequence tags. BMC Genomics 9:279.

Fry B, Roelants K, Champagne DE, et al. (12 co-authors). 2009. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms Annu Rev Genomics Hum Genet 10:483-511.

Jiang L, Zhang D, Zhang Y, Peng L, Chen J, Liang S. 2010. Venomics of the spider *Ornithoctonus huwena* based on transcriptomic versus proteomic analysis. Comp Biochem Physiol D Genomics and Proteomics 5:81-88.

Liu Z-C, Zhang R, Zhao F, et al. (12 co-authors). 2012. Venomic and transcriptomic analysis of centipede *Scolopendra subspinipes dehaani*. J Proteome Res 11:6197-6212.

Min G-S, Sarkar IN, Siddall ME. 2010. Salivary transcriptome of the North American medicinal leech, *Macrobdella decora*. J Parasitol 96:1211-1221.

Morgenstern D, Rohde BH, King GF, Tal T, Sher D, Zlotkin E. 2011. The tale of a resting gland: transcriptome of a replete venom gland from the scorpion *Hottentotta judaicus*. Toxicon:1-36.

Rendón-Anaya M, Delaye L, Possani LD, Herrera-Estrella A. 2012. Global transcriptome analysis of the scorpion *Centruroides noxius*: new toxin families and evolutionary insights from an ancestral scorpion species. PLoS ONE 7:e43331.

Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, Caudle SB. 2011. A high-throughput venom-gland transcriptome for the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes. Toxicon 57:657-671.

Ruder T, Sunagar K, Undheim EAB, Ali SA, Wai T-C, Low DHW, Jackson TNW, King GF, Antunes A, Fry BG. 2013. Molecular phylogeny and evolution of the proteins encoded by coleoid (cuttlefish, octopus, and squid) posterior venom glands. J Mol Evol 76:192-204.

Ruiming Z, Yibao M, Yawen H, Zhiyong D, Yingliang W, Zhijian C, Wenxin L. 2010. Comparative venom gland transcriptome analysis of the scorpion *Lychas mucronatus* reveals intraspecific toxic gene diversity and new venomous components. BMC Genomics 11:452.

Terrat Y, Biass D, Dutertre S, Favreau P, Remm M, Stöcklin R, Piquemal D, Ducancel F. 2012. High-resolution picture of a venom gland transcriptome: case study with the marine snail *Conus consors*. Toxicon 59:34-46.

Undheim EAB, King GF. 2011. On the venom system of centipedes (Chilopoda), a neglected group of venomous animals. Toxicon 57:512-524.

Whittington CM, Papenfuss AT, Locke DP, et al. (12 co-authors). 2010. Novel venom gene discovery in the platypus. Genome Biol 11:R95.