

SPECIAL REPORT



Integration of transcriptomic and proteomic approaches for snake venom profiling

Cassandra M. Modahl^a, Anthony J. Saviola^b and Stephen P. Mackessy^c

^aLiverpool School of Tropical Medicine, Liverpool, UK; ^bDepartment of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, Aurora, USA; ^cSchool of Biological Sciences, University of Northern Colorado, Greeley, USA

ABSTRACT

Introduction: Snake venoms contain many protein and peptide isoforms with high levels of sequence variation, even within a single species.

Areas covered: In this review, we highlight several examples, from both published and unpublished work in our lab, demonstrating how a combined venom gland transcriptome and proteome methodology allows for comprehensive characterization of venoms, including those from understudied rear-fanged snake species, and we provide recommendations for using these approaches.

Expert Opinion: When characterizing venoms, peptide mass fingerprinting using databases built predominately from protein sequences originating from model organisms can be disadvantageous, especially when the intention is to document protein diversity. Therefore, the use of species-specific venom gland transcriptomes corrects for the absence of these unique peptide sequences in databases. The integration of transcriptomics and proteomics improves the accuracy of either approach alone for venom profiling.

ARTICLE HISTORY

Received 31 August 2021

Accepted 15 October 2021

KEYWORDS

Mass spectrometry; non-model organisms; protein; snake; toxin

1. Introduction

Snake venoms are complex oral secretions, composed of many proteins and peptides that function individually and/or synergistically to target multiple physiological systems in prey or predators [1]. Through the process of gene duplication and neofunctionalization, a single toxin family can have dozens of isoforms with different activities, all dependent on amino acid residue substitutions within a conserved family scaffold [2,3]. Profiling and characterizing the blend of toxins present in a venom and determining their sequences can be a daunting task, especially considering the documented variation in toxin diversity between individual snakes at both the transcriptome [4] and the proteome levels [5–8]. However, advances in sensitivity and high-throughput technology in the fields of genomics, transcriptomics, and proteomics have provided powerful approaches to address these challenges [9]. The integration of these -omic fields in venom research has been termed ‘venomics’ [10–12].

Genomes provide locus-level resolution of variation of toxin genes and insight into the evolutionary processes generating sequence diversity, such as gene duplication, deletion, and single nucleotide polymorphisms [13–17]. Venom gland transcriptomes are useful to identify which genes are expressed, as variation in the expression of toxin gene isoforms has been well-documented [18,19] and multiple gene duplications can create nonfunctional pseudogenes [15,20]. Venom gland transcriptomes can also be used to obtain the full amino acid sequences of expressed venom proteins. Determining these sequences from proteins directly, such as by N-terminal

sequencing and *de novo* sequencing of tandem mass spectrometry (MS/MS) spectra, can be performed, but these methods are labor-intensive, expensive, and complete protein sequences are not guaranteed, as some proteins are N-terminally blocked, do not exhibit sequence for efficient protease digestion, or do not ionize well for MS/MS. Additionally, a complementary proteomic approach is necessary to identify which toxins contribute to the expressed venom phenotype [21].

Although top-down proteomics strategies have been used for venom profiling [8,22], a bottom-up MS/MS or shotgun proteomic approach is more popular. A review of 89 Viperinae venom proteomes identified only six species where top-down methods were used [23]. For bottom-up approaches, proteins are first digested with proteases such as trypsin (most commonly used), chymotrypsin, or Glu-C, and the spectra generated from the proteolytic peptides are used to infer what proteins are present in the sample. The high-throughput identification of peptides from MS/MS is typically achieved using one of the several database search engines, such as Mascot or SEQUEST [24,25]. For most model organisms, standard reference databases obtained from publicly available repositories for example, UniProt [26], RefSeq [27], and Ensembl [28], can be used for this analysis. In contrast, the majority of non-model organisms, such as venomous snakes, have few or poorly annotated genomes and insufficient protein sequence databases represent a major limitation to MS-based proteomics [29,30]. As a consequence, an increasing number of studies integrate transcriptomic-proteomic

- Use of a species-specific venom gland transcriptome allows for more accurate proteomic quantification of venom components
- Different databases can bias proteomic results, and smaller databases increase detection sensitivity
- Species-specific databases better detect unique peptide sequences

2. Identification of venom proteoforms: a case study in *Boiga irregularis*

Venom from the rear-fanged Brown Treesnake (*Boiga irregularis*) is predominately composed of 3FTxs. The number of 3FTx isoforms is quite large, with 65 3FTx transcripts identified

A top-down MS analysis of adult *B. irregularis* venom from a snake from Guam identified a total of 30 3FTxs proteoform masses [8]. A different individual, also originating from Guam, was used for top-down MS than for the venom gland transcriptome, and remarkably, 3FTx proteoforms were detected that matched to translated transcripts from the Indonesian *B. irregularis* venom gland transcriptome and to masses in the range of 3FTxs that were unassigned to any specific 3FTx in either venom gland transcriptome [8]. This demonstrates that venom gland transcriptomes should be completed for multiple individuals to document full toxin transcript diversity for a given species [4], and if these sequences are to be used as proteomic databases, ideally also for individuals from different geographic regions.

Integrated transcriptomics and proteomics are ideal for not only more accurate identification of venom proteins but also

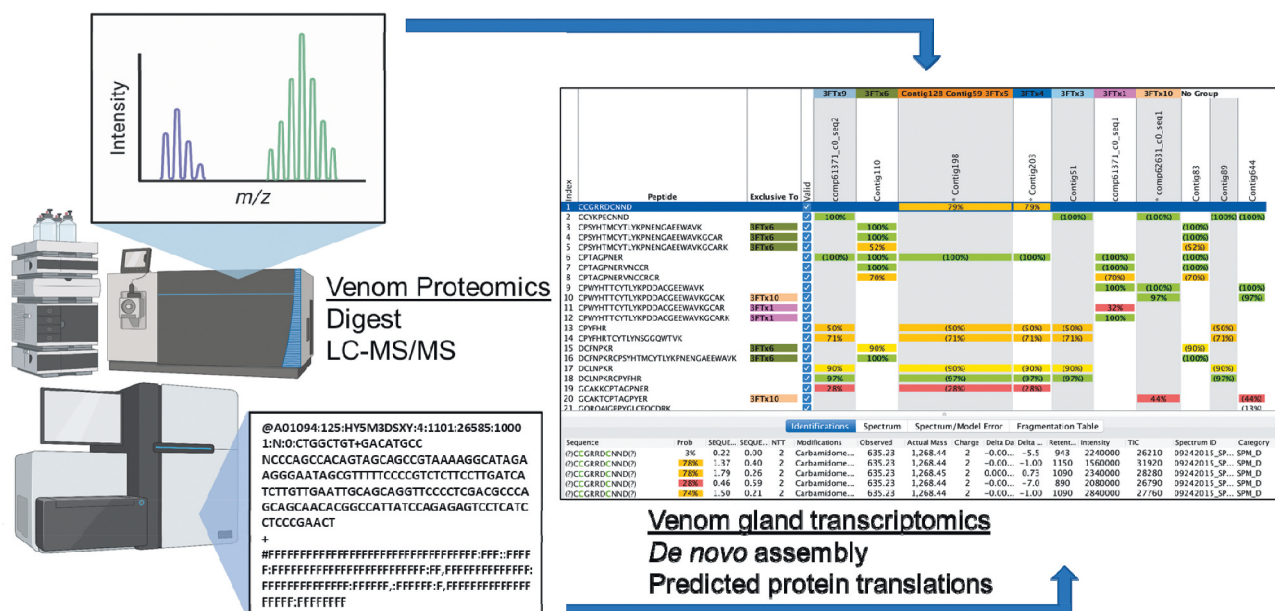


Figure 1. Integration of venom gland transcriptomics and venom proteomics work flows. Image created in bioRender (biorender.com).

better protein quantification [41], as it allows for better detection of all peptide fragments generated by MS/MS. There are several label-free methods of MS/MS quantification, such as Normalized Spectral Abundance Factors (NSAF) [8], which normalizes for protein length, or the use of an internal standard of known concentration that is then used to determine unknown concentrations of proteins based upon peptide intensities [41,42], similarly used for iBAQ (Intensity Based Absolute Quantification) [43,44]. Some proteomic analysis programs generate their own quantification numbers, such as the emPAI (exponentially modified Protein Abundance Index) number [45] from ProteinPilot and Mascot. To improve the accuracy of these methods, the use of a species-specific transcriptome-generated database can aid in the quantification of venom components and can provide exact protein sizes for NSAF calculations.

To illustrate this point, the venom proteome of the Amazon Puffing Snake (*Spilotes sulphureus*) was evaluated in several ways: via comparison with public databases (Figure 2(a)), with a transcriptome-guided reference database (Figure 2(b)), and via integration of identified peaks following size-exclusion HPLC (SE-HPLC) fractionation of the venom (Figure 2(c), absorbance measured at 220 nm). When protein

family relative abundances were evaluated using a venom gland transcriptome database with NSAF calculations or the HPLC-derived proteome, values were quite comparable. The use of public databases tended to overestimate the abundance of certain proteins, such as L-amino acid oxidase (LAAO), resulting in a misleading characterization of the venom proteome with the same MS/MS settings (Figure 2(a)). LAAO was unlikely to be this abundant because LAAO enzyme activity had not been detectable in this venom [46]. The reason for this inflated estimate is likely because some venom proteins, such as LAAO, have high sequence conservation between different venomous snake species and therefore these shared peptide fragments are easily identified in public databases where the same sequence would be present. However, some toxin families, such as 3FTxs, exhibit low sequence conservation, and these unique peptide fragments are absent from databases. For example, *S. sulphureus* 3FTx isoform identities ranged from 21–99% in just one individual (Supplemental Figure 1(a)), although there is conservation of the cysteine network creating the ‘three-finger’ toxin fold (Supplemental Figure 1B). This exclusion of novel 3FTx peptide fragments from public databases resulted in the overestimation of low abundance, conserved toxins (like

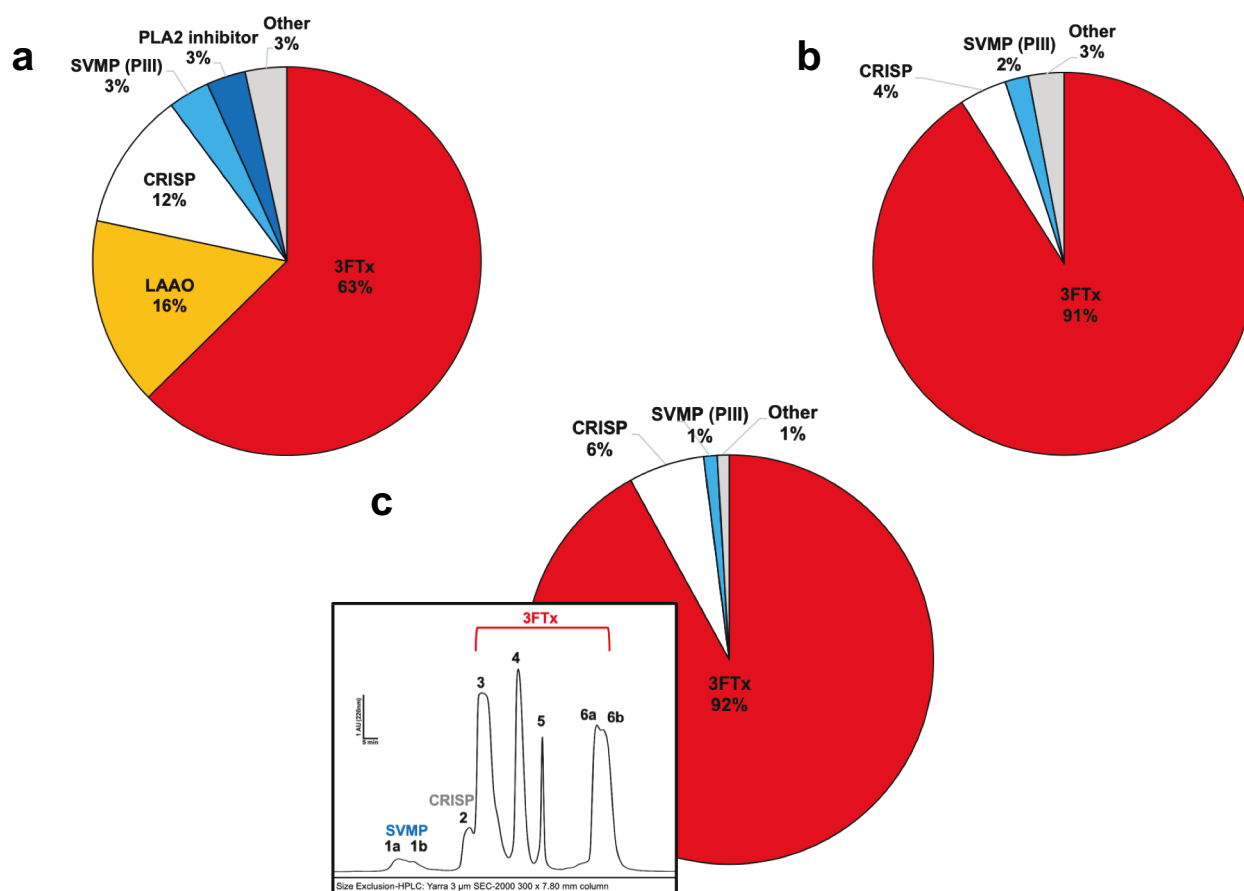


Figure 2. Amazon Puffing Snake (*Spilotes sulphureus*) venom profiles using shotgun proteomic spectra searches against (a) public databases and (b) a species-specific venom gland transcriptome database or (c) based on size-exclusion high-performance liquid chromatography (SE-HPLC) fractionation and analysis. For shotgun proteomic quantifications, normalized spectral abundance factors (NSAF) were used and protein family percentages are relative to the sum of all toxins. For SE-HPLC, venom proteins present in each peak were determined by tandem mass spectrometry, N-terminal sequencing, and/or enzyme assays [45], and percentages were calculated from peak areas and are relative to the total area. Abbreviations: 3FTx – three-finger toxin; CRISP – cysteine-rich secretory protein; LAAO – L-amino acid oxidase; PLA2 inhibitor – phospholipase A₂ inhibitor; SVMP – snake venom metalloproteinase (P-III class).

LAO) in the *S. sulphureus* venom profile. The use of a species-specific transcriptome to complement the venom proteome ameliorated this problem (Figure 2(b)), producing abundance estimations by MS/MS that matched what was observed using SE-HPLC fractionation and analysis (Figure 2(c)). It is also important to note that although the translated transcriptome is ideal as a species-specific database for MS/MS, there is not always a quantitative correspondence between the transcriptome and proteome [33,46]; however, it is usually seen that transcripts with the highest expression levels are also abundant in the venom [43].

Finally, the venoms field has seen an increase in studies using shotgun proteomic label-free quantification (LFQ), such as intensity or spectral counting, to compare the abundance of different proteins in the same sample. This strategy, however, was developed to compare abundance changes of the same protein in different samples [47,48] and is based on the assumption that the same protein will produce the same proteolytic peptides that ideally exhibit identical behaviors on the mass spectrometer. On the other hand, the different physicochemical properties of peptides derived from different parent proteins can, and often do, result in drastically different behaviors on the mass spectrometer. Therefore, comparing intensity or spectral counts generated from different peptides will result in incorrect and unreliable estimations of protein abundances, and as a consequence, we caution against the inappropriate use of LFQ (see [30]). There is also the general drawback to venom proteomics quantification with MS-based methodologies in that it is difficult to compare across studies where different instrumentation and settings are used.

4. Database considerations for venom proteins: a case study in *Thamnophis elegans*

A potential source of confusion when analyzing venom-derived MS data is whether spectra should be searched against just the venom gland transcriptome-derived species-specific database or UniProt databases (restricted to Squamata) merged with the species-specific sequences. It has been shown that venom proteome results will vary with databases; for example, Patra et al. [49] saw differences in results when utilizing databases consisting of snake family, genus, or species. While ultimately, one must consider the goals of the experiment, databases derived from venom gland transcriptomes provide highly specific protein sequences that will greatly improve peptide identifications and overall sequence coverage [50,51]. On the other hand, larger databases contain a plethora of irrelevant sequences, which for most species are not representative of the protein composition of their venom. The increased search space of larger databases reduces the number of successful peptide identifications by increasing the frequency of false assignments [50–55]. Therefore, it has been suggested that databases should be as compact as possible and include protein sequences believed to be in the sample [50].

To explore these database questions, we profiled the venom from the Western Terrestrial Garter Snake (*Thamnophis elegans vagrans*), which is currently unstudied.

Venom was obtained from *T. e. vagrans* by manual extraction [56], solubilized in urea buffer, subjected to reduction, alkylation, and trypsin digestion, and analyzed using a shotgun proteomics approach on an Orbitrap Velos mass spectrometer. We also dissected the Duvernoy's venom gland from *T. e. vagrans*, performed RNA sequencing and *de novo* assembled a venom gland transcriptome for this species, using the same methodology as previously reported [32,46]. Predicted translated venom proteins were used as a species-specific database to identify and quantify the venom proteome.

To compare several differences associated with searching two different databases, we searched spectra acquired from *T. e. vagrans* venom against a database derived from the *de novo* assembled *T. e. vagrans* venom gland transcriptome (referred hereafter as *Thamnophis*-only) and against the SwissProt/Trembl database restricted to Serpentes, merged with the venom gland transcriptome database (referred hereafter as the combined database). The combined database contains 631,016 entries, including reverse decoys and contaminants, whereas the *Thamnophis*-only database contains 418 protein sequences, including reverse decoys and contaminants. Data were searched separately against the two databases mentioned above with ProLucid [57], and the false discovery rate (FDR) was 1%. After filtering, the combined database search resulted in 2,095 spectra matching to 744 peptide sequences resulting in 407 identified proteins, although the vast majority of these proteins were non-venom proteins (Supplemental Table S1). Search results against the *Thamnophis*-only database resulted in 2,007 spectra matching 509 peptides resulting in 32 proteins, of which 28 were also identified in the combined database. Two waprins-like proteins, a serine protease inhibitor, and a C-type lectin subunit, were not identified in the combined database search results (Figure 3). Of the 28 *Thamnophis*-specific proteins shared between the two database search results, 24 had greater sequence coverage for the *Thamnophis*-only database, and there was no difference in percent sequence coverage for the other four proteins (Figure 3(a)). There were also more spectral counts assigned to 25 of the 28 shared proteins for the *Thamnophis*-only database, and for 12 of the proteins there were >100 spectral count differences between the two databases (Figure 3(b)). This result is significant, as spectral counting has become an increasingly popular LFQ strategy for estimating protein abundance in venom-focused shotgun proteomics experiments (but see [30]).

5. Unique peptide identifications for venom proteins: a case study in *Crotalus oreganus helleri*

Postsearch filtering based on unique peptides (a peptide that matches with only one protein in the database) is often applied as a validation criterion for protein identifications [58]. The increased number of proteins in public databases can decrease the number of unique peptides, which, when used as a requirement (i.e. each protein must have *n* unique peptides to be considered positively identified), can actually result in a loss of valid protein identifications. This is especially apparent for venom proteins from front-fanged viperid and

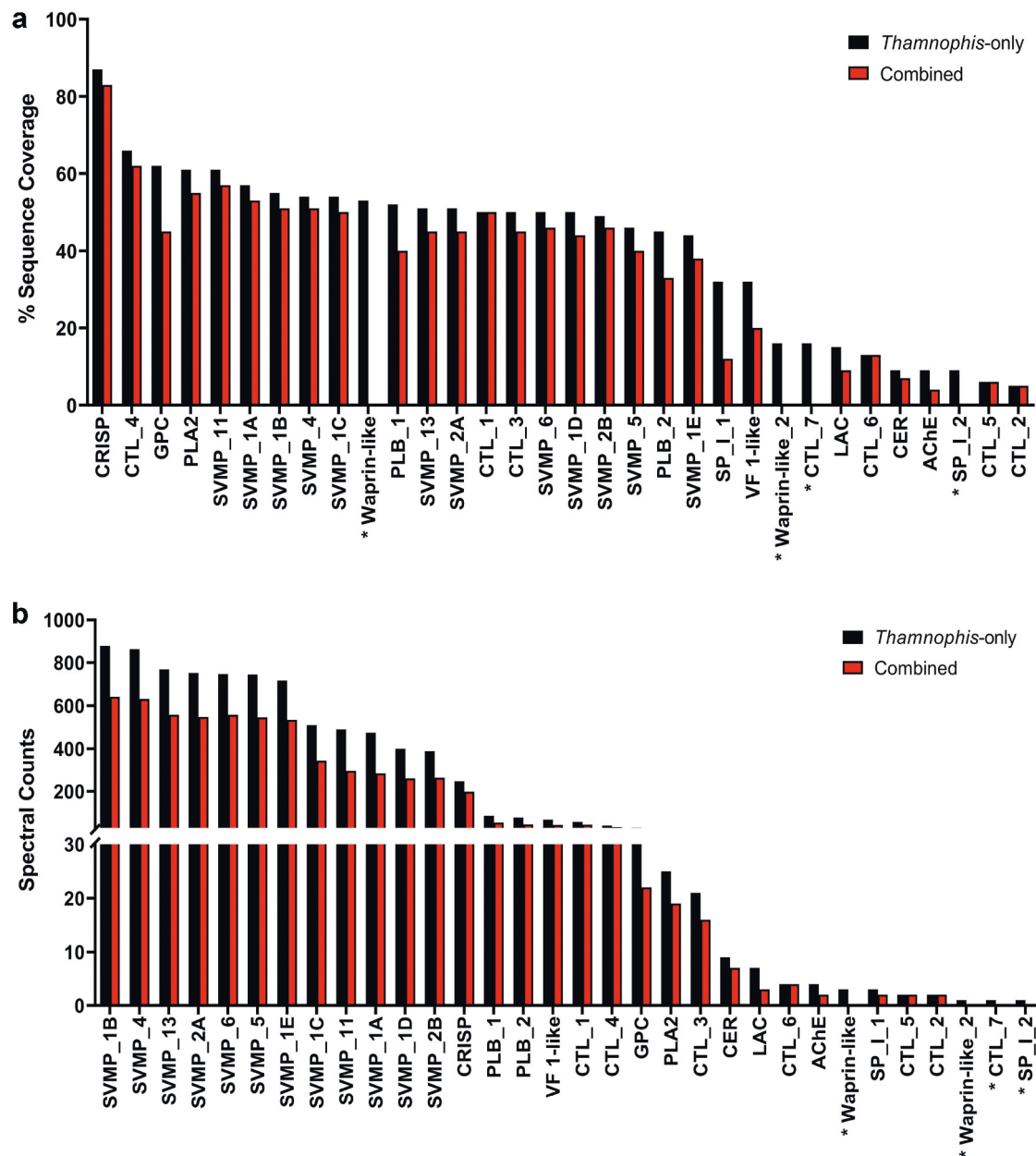


Figure 3. Comparison of the percent sequence coverage (a) and spectral counts (b) for *Thamnophis*-specific proteins identified between the two database search results. The *Thamnophis*-only database contains annotated protein sequences from *T. elegans vagrans* venom transcriptome, whereas the combined database contains all Squamata Swiss-Prot/TrEMBL protein sequences plus the sequences derived from the *T. e. vagrans* venom transcriptome. Both databases contained reverse decoy and contaminant sequences. Proteins labeled with * were not identified in the combined database search, despite these sequences being present. Abbreviations: AChE – acetylcholinesterase; CER – ceruloplasmin; CRISP – cysteine-rich secretory protein; CTL – C-type lectin; GPC – glutaminyl-peptide cyclotransferase; LAC – lactadherin; PLA2 – phospholipase A₂; PLB – phospholipase B; SP_1 – serine protease inhibitor; SVMP – snake venom metalloproteinase; VF 1-like – venom factor 1-like.

elapid species that have protein sequences that share a high degree of homology with other proteins in public databases. Again, as the venom gland transcriptome-derived databases contain highly specific protein sequences, many of the identified peptides will be unique to a specific protein.

An example of how this can result in the loss of protein identifications can be seen when searching spectra from Southern Pacific Rattlesnake (*Crotalus oreganus helleri*) venom against the species-specific venom gland-derived transcriptome [59] and looking at the kallikrein-like serine protease T1DE97. When the search was conducted against the venom gland

transcriptome database for *C. o. helleri*, 13 unique peptides were assigned to this protein (Table 1). However, when searching the same spectra against the Swiss-Prot/TrEMBL databases (restricted to Serpentes) merged with the *C. o. helleri* protein sequences and requiring a minimum of 1 unique peptide for protein validation, T1DE97 is lost in the filtering step (see also [60]). Every unique peptide identified in the venom gland transcriptome only search is shared with at least one other protein in the Serpentes Swiss-Prot/TrEMBL databases (Table 1). While unique peptides can be utilized for distinguishing between protein isoforms [61], whether to apply this filtering step should be

Table 1. A list of unique peptides assigned to the kallikrein-like serine protease T1DE97 when spectra from *C. o. helleri* venom are searched against a species-specific venom gland transcriptome database [59]. Each unique peptide identified from the species-specific database search matches with at least one other protein in the Swiss-Prot/Trembl databases. If a unique peptide requirement were applied to the search results with the latter database, T1DE97 would not be included in the final results. The list of proteins from Swiss-Prot/TrEMBL is not exhaustive, and some peptides match to >20 protein sequences.

m/z	z	Unique peptide sequences from <i>C. o. helleri</i> venom gland transcriptome	Proteins in Swiss-Prot/TrEMBL databases that share peptide seq. (List is not exhaustive)
429.71	2	FFCLGSK	A0A2I7YS62 <i>C. molossus</i>
548.3	2	RLNPGFYTK	A0A2I7YSB6 <i>C. tigris</i>
590.80	2	IMGWGTISSTK	B0VXU0 <i>S. c. edwardsii</i>
667.34	2	LNNPVNSAHIAP	J3S3W4 <i>C. adamanteus</i>
674.34	2	AAYPWWPVTTTR	T1D6M5 <i>C. horridus</i>
754.38	2	VGGRPCNINEHR	
791.38	2	CANINILDYAVCR	
831.42	2	TNNEWKDIMLR	
1221.64	2	LNNPVNSAHIAPLSLPSSPPSVGS	
424.23	3	NFQIQLGVHVK	
1605.21	2	DSCQGDGGGLICNGQFQGIVSWGAQPCGR	A0A2I7YS86 <i>C. scutulatus</i>
1080.76	4	ILCAGILEGGKDSQGDGGGLICNGQFQGIVSWGAQPCGR	
957.83	3	LNNPVNSAHIAPLSLPSSPPSVGSLCR	A0A2I7YS62 <i>C. molossus</i> A0A2I7YSB6 <i>C. tigris</i> B0VXU0 <i>S. c. edwardsii</i> J3S3W4 <i>C. adamanteus</i>

considered within the goals of the experiment. Undoubtedly, parallel venom gland transcriptomic and proteomic workflows allow for a more complete understanding of venom systems at the gene and protein level. The results presented here should also apply to most other species for which venom gland transcriptomes are available (e.g Ref 60).

6. Expert opinion

6.1. Future directions

An area of venom proteomics that has been neglected is the use of MS approaches to evaluate venom protein glycosylations. N-linked glycan structures have been found to be present in many venoms, especially those of Viperidae [62], but venoms from most species have not been studied. Standard bottom-up workflows that are currently the more common venomomics approaches cannot be used to examine glycosylations because collisional activation does not usually cleave glycosylated peptides, making them harder to detect. In these instances, the initial removal of N-linked glycans by PNGase F could aid in both identification of these glycosylated sites and better protein coverage. The use of multienzyme digestions also offers a means to increase the number of different peptide fragments that can be identified, and this method would increase coverage and accuracy in determining venom proteoforms present [63]. Integration of transcriptome databases will aid both of these techniques, as glycosylation sites can be predicted from translated transcriptomes and provide proteoform reference sequences. These future directions in venomomics pipelines will provide a more comprehensive picture of snake venom proteins and post-translational modifications of these potent biomolecules.

6.2. Concluding remarks

Increases in the sensitivity of high-throughput transcriptomics and proteomics and the greater availability and affordability of these methods have benefited research using nonmodel

organisms. This is especially true for organisms where it was previously difficult to profile venoms, such as rear-fanged venomous snakes that can be problematic to extract venom and often have low yields [56]. These methods have accelerated the rate of characterizing toxins, as novel sequences can be quickly identified, and knowledge of the complete protein sequence can be useful for recombinant expression or peptide synthesis experiments. Our laboratory has had considerable success integrating multiple -omic approaches to profile snake venoms by building species-specific venom gland transcriptome databases to search MS spectra. Combined with functional assays, we have been able to identify unique sequence features linked to specific activities, such as three-finger toxins with prey-specific toxicity [46] and venoms with diverse metalloproteinase activities [32]. Furthermore, we have shown in this review that there are advantages of having species-specific transcriptomes to identify and quantify secreted venom proteins and that these smaller transcriptome databases typically perform better than larger combined databases. This approach can be utilized with analyses of any animal venoms and should result in more complete proteomic characterizations.

Declaration of interests

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Funding

This work was supported by a grant from the NSF (DEB Phylogenetic Systematics; DEB-1655571) to SPM.

References

1. Mackessy SP, ed. Handbook of venoms and toxins of reptiles. 2nd ed. Boca Raton (FL): CRC Press/Taylor & Francis Group; 2021. p. 652.

2. Kini RM. Molecular moulds with multiple missions: functional sites in three-finger toxins. *Clin Exp Pharm Physiol.* **2002**;29(9):815–822.
3. Kini RM. Excitement ahead: structure, function and mechanism of snake venom phospholipase A₂ enzymes. *Toxicon.* **2003**;42(8):827–840.
4. Rokyta DR, Margres MJ, Ward MJ, et al. The genetics of venom ontogeny in the eastern diamondback rattlesnake (*Crotalus adamanteus*). *PeerJ.* **2017**;5:e3249.
5. Alape-Giron A, Sanz L, Escolano J, et al. Snake venomomics of the lancehead pit viper *Bothrops asper* geographic, individual, and ontogenetic variations. *J Proteome Res.* **2008**;7(8):3556–3571.
6. Meier J. Individual and age-dependent variations in the venom of the fer-de-lance (*Bothrops atrox*). *Toxicon.* **1986**;24(1):41–46.
7. Mackessy SP, Sixberry NM, Heyborne WH, et al. Venom of the brown tree snake, *Boiga irregularis*: ontogenetic shifts and taxa-specific toxicity. *Toxicon.* **2006**;47(5):537–548.
8. Pla D, Petras D, Saviola AJ, et al. Transcriptomics-guided bottom-up and top-down venomomics of neonate and adult specimens of the arboreal rear-fanged brown tree snake, *Boiga irregularis*, from Guam. *J Proteomics.* **2017**;174:71–84.
9. Modahl CM, Brahma RK, Koh CY, et al. Omics technologies for profiling toxin diversity and evolution in snake venom: impacts on the discovery of therapeutic and diagnostic agents. *Ann Rev Animal Biosci.* **2020**;8(1):91–116.
10. Calvete JJ. Next-generation snake venomomics: protein-locus resolution through venom proteome decomplexation. *Expert Rev Proteomics.* **2014**;11(3):315–329.
11. Calvete JJ, Juárez P, Sanz L. Snake venomomics. Strategy and applications. *J Mass Spectrometry.* **2007**;42(11):1405–1414.
12. Calvete JJ, Sanz L, Angulo Y, et al. Venoms, venomomics, antivenomics. *FEBS Lett.* **2009**;583(11):1736–1743.
13. Vonk FJ, Casewell NR, Henkel CV, et al. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc Nat Acad Sci USA.* **2013**;110(51):20651–20656.
14. Dowell NL, Giorgianni MW, Kassner VA, et al. The deep origin and recent loss of venom toxin genes in rattlesnakes. *Curr Biol.* **2016**;26(18):2434–2445.
15. Giorgianni MW, Dowell NL, Griffin S, et al. The origin and diversification of a novel protein family in venomous snakes. *Proc Nat Acad Sci USA.* **2020**;117(20):10911.
16. Shibata H, Chijiwa T, Oda-Ueda N, et al. The habu genome reveals accelerated evolution of venom protein genes. *Sci Rep.* **2018**;8(1):11300.
17. Kerkkamp HMI, Kini RM, Pospelov AS, et al. Snake genome sequencing: results and future prospects. *Toxins (Basel).* **2016**;8(12):15.
18. Margres M, Aronow K, Loyacano J, et al. The venom-gland transcriptome of the eastern coral snake (*Micrurus fulvius*) reveals high venom complexity in the intragenomic evolution of venoms. *BMC Genomics.* **2013**;14(1):1–18.
19. Casewell NR, Harrison RA, Wüster W, et al. Comparative venom gland transcriptome surveys of the saw-scaled vipers (Viperidae: *echis*) reveal substantial intra-family gene diversity and novel venom transcripts. *BMC Genomics.* **2009**;10(1):564.
20. Suryamohan K, Krishnankutty SP, Guillory J, et al. The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. *Nat Genet.* **2020**;52(1):106–117.
21. Aird SD, Da Silva NJ, Qiu L, et al. Coral snake venomomics: analyses of venom gland transcriptomes and proteomes of six Brazilian taxa. *Toxins (Basel).* **2017**;9(6):187.
22. Petras D, Heiss P, Harrison RA, et al. Top-down venomomics of the East African green mamba, *Dendroaspis angusticeps*, and the black mamba, *Dendroaspis polylepis*, highlight the complexity of their toxin arsenals. *J Proteomics.* **2016**;146:148–164.
23. Damm M, Hempel B-F, Süssmuth RD. Old world vipers—A review about snake venom proteomics of Viperinae and their variations. *Toxins (Basel).* **2021**;13(6):427.
24. Eng JK, Searle BC, Clauser KR, et al. A face in the crowd: recognizing peptides through database search. *Mol Cell Proteomics.* **2011**;10(11):R111.009522.
25. Verheggen K, Raeder H, Berven FS, et al. Anatomy and evolution of database search engines—a central component of mass spectrometry based proteomic workflows. *Mass Spec Rev.* **2020**;39(3):292–306.
26. Bateman A, Martin M-J, Orchard S. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**;49(D1):D480–d489.
27. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**;44(D1):D733–745.
28. Yates AD, Achuthan P, Akanni W, et al. Ensembl 2020. *Nucleic Acids Res.* **2020**;48(D1):D682–D688.
29. Calvete JJ. Proteomic tools against the neglected pathology of snake bite envenoming. *Expert Rev Proteomics.* **2011**;8(6):739–758.
30. Calvete JJ, Lomonte B, Saviola AJ, et al. Mutual enlightenment: a toolbox of concepts and methods for integrating evolutionary and clinical toxinology via snake venomomics and the contextual stance. *Toxicon X.* **2021**; 10070.
31. Durban J, Sanz L, Trevisan-Silva D, et al. Integrated venomomics and venom gland transcriptome analysis of juvenile and adult Mexican rattlesnakes *Crotalus simus*, *C. tzabcan*, and *C. culminatus* revealed miRNA-modulated ontogenetic shifts. *J Proteome Res.* **2017**;16(9):3370–3390.
32. Modahl CM, Frieze S, Mackessy SP. Transcriptome-facilitated proteomic characterization of rear-fanged snake venoms reveal abundant metalloproteinases with enhanced activity. *J Proteomics.* **2018**;187:223–234.
33. Tan CH, Tan KY, Fung SY, et al. Venom-gland transcriptome and venom proteome of the Malaysian king cobra (*Ophiophagus hannah*). *BMC Genomics.* **2015**;16(1):687.
34. Wiezel GA, Shibao PYT, Cologna CT, et al. In-depth venom of the Brazilian rattlesnake *Crotalus durissus terrificus*: an integrative approach combining its venom gland transcriptome and venom proteome. *J Proteome Res.* **2018**;17(11):3941–3958.
35. Walker AA, Robinson SD, Hamilton BF, et al. Deadly proteomes: a practical guide to proteotranscriptomics of animal venoms. *Proteomics.* **2020**;20(17–18):1900324.
36. McGivern JJ, Wray KP, Margres MJ, et al. RNA-seq and high-definition mass spectrometry reveal the complex and divergent venoms of two rear-fanged colubrid snakes. *BMC Genomics.* **2014**;15(1061). [10.1186/1471-2164-15-1061](https://doi.org/10.1186/1471-2164-15-1061)
37. Modahl CM, Mackessy SP. Venoms of rear-fanged snakes: new proteins and novel activities. *Front Ecol Evol.* **2019**;7(279). [10.3389/fevo.2019.00279](https://doi.org/10.3389/fevo.2019.00279)
38. Peichoto ME, Tavares FL, Santoro ML, et al. Venom proteomes of South and North American opisthoglyphous (Colubridae and Dipsadidae) snake species: a preliminary approach to understanding their biological roles. *Comp Biochem PhysiolPart D: Genomics and Proteomics.* **2012**;7(4):361–369.
39. Pawlak J, Mackessy SP, Sixberry NM, et al. Irditoxin, a novel covalently linked heterodimeric three-finger toxin with high taxon-specific neurotoxicity. *FASEB J.* **2009**;23(2):534–545.
40. Nei M, Gu X, Sitnikova T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Nat Acad Sci USA.* **1997**;94(15):7799–7806.
41. Neilson KA, Keighley T, Pascovici D, et al. Label-free quantitative shotgun proteomics using normalized spectral abundance factors. *Meth Mol Biol.* **2013**;1002:205–222.
42. Paoletti AC, Parmely TJ, Tomomori-Sato C, et al. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Nat Acad Sci USA.* **2006**;103(50):18928–18933.
43. Rokyta DR, Margres MJ, Calvin K. Post-transcriptional mechanisms contribute little to phenotypic variation in snake venoms. *G3.* **2015**;5(11):2375–2382.

44. Fabre B, Lambour T, Bouyssié D, *et al.* Comparison of label-free quantification methods for the determination of protein complexes subunits stoichiometry. *EuPA Open Proteom.* **2014**;4:82–86.
45. Ishihama Y, Oda Y, Tabata T, *et al.* Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics.* **2005**;4(9):1265–1272.
46. Modahl CM, Mrinalini FS, Mackessy SP. Adaptive evolution of distinct prey-specific toxin genes in rear-fanged snake venom. *Proc. Royal Soc. B – Biol. Sci.* **2018**;285(1884)
47. Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. **2010**; *J Biomed Biotech.* 2010:840518.
48. Megger DA, Bracht T, Meyer HE, *et al.* Label-free quantification in clinical proteomics. *Biochim Biophys Acta - Proteins Proteomics.* **2013**;1834(8):1581–1590.
49. Patra A, Kalita B, Chanda A, *et al.* Proteomics and antivenomics of *Echis carinatus carinatus* venom: correlation with pharmacological properties and pathophysiology of envenomation. *Sci Rep.* **2017**;7(1):17119.
50. Jeong K, Kim S, Bandeira N. False discovery rates in spectral identification. *BMC Bioinformatics.* **2012**;13(16):S2.
51. Noble WS. Mass spectrometrists should search only for peptides they care about. *Nature Meth.* **2015**;12(7):605–608
52. Cargile BJ, Bundy JL, Stephenson JL Jr. Potential for false positive identifications from large databases through tandem mass spectrometry. *J Proteome Res.* **2004**;3(5):1082–1085.
53. Resing KA, Ahn NG. Proteomics strategies for protein identification. *FEBS Lett.* **2005**;579(4):885–889.
54. Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J Proteomics.* **2010**;73(11):2092–2123.
55. Wang X, Slebos RJ, Wang D, *et al.* Protein identification using customized protein sequence databases derived from RNA-Seq data. *J Proteome Res.* **2012**;11(2):1009–1017.
56. Saviola AJ, Peichoto ME, Mackessy SP. Rear-fanged snake venoms: an untapped source of novel compounds and potential drug leads. *Toxin Rev.* **2014**;33(4):1–17.
57. Xu T, Park SK, Venable JD, *et al.* ProLuCID: an improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *J Proteomics.* **2015**;129:16–24.
58. Zhang Y, Wen Z, Washburn MP, *et al.* Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Analyt Chem.* **2010**;82(6):2272–2281
59. Sunagar K, Undheim EA, Scheib H, *et al.* Intraspecific venom variation in the medically significant Southern Pacific Rattlesnake (*Crotalus oreganus helleri*): biodiscovery, clinical and evolutionary implications. *J Proteomics.* **2014**;99:68–83.
60. Robinson KE, Holding ML, Whitford MD, *et al.* Phenotypic and functional variation in venom and venom resistance of two sympatric rattlesnakes and their prey. *J Evol Biol.* **2021**;34(9):1447–1465.
61. Stastna M, Van Eyk JE. Analysis of protein isoforms: can we do it better? *Proteomics.* **2012**;12(19–20):2937–2948.
62. Andrade-Silva D, Ashline D, Tran T, *et al.* Structures of N-Glycans of *Bothrops* venoms revealed as molecular signatures that contribute to venom phenotype in viperid snakes. *Mol Cell Proteomics.* **2018**;17(7):1261–1284.
63. Morsa D, Baiwir D, La Rocca R, *et al.* Multi-Enzymatic Limited digestion: the next-generation sequencing for proteomics? *J Proteome Res.* **2019**;18(6):2501–2513.