

Using a Novel Ontology to Inform the Discovery of Therapeutic Peptides from Animal Venoms

Joseph D. Romano¹ and Nicholas P. Tatonetti, PhD¹

¹Departments of Biomedical Informatics, Systems Biology, and Medicine,
Columbia University Medical Center, New York, NY

Abstract

Venoms and venom-derived compounds constitute a rich and largely unexplored source of potentially therapeutic compounds. To facilitate biomedical research, it is necessary to design a robust informatics infrastructure that will allow semantic computation of venom concepts in a standardized, consistent manner. We have designed an ontology of venom-related concepts – named Venom Ontology – that reuses an existing public data source: UniProt's Tox-Prot database. In addition to describing the ontology and its construction, we have performed three separate case studies demonstrating its utility: (1) An exploration of venom peptide similarity networks within specific genera; (2) A broad overview of the distribution of available data among common taxonomic groups spanning the known tree of life; and (3) An analysis of the distribution of venom complexity across those same taxonomic groups. Venom Ontology is publicly available on BioPortal at <http://biportal.bioontology.org/ontologies/CU-VO>.

1. Introduction

Venoms are complex mixtures of biological macromolecules and cofactors used for either offensive or defensive purposes. In contrast to poisons (which are harmful only when swallowed), venoms can be introduced to their targets via injection, surface abrasion, or topical application. Most known venoms exert their effect using specialized proteins that have a strong and specific effect on target cells[1]. A vast array of animal species is known to be venomous, and many more are predicted to be venomous. Additionally, individual venoms usually consist of a diverse array of many components. As an example, many species of cone snail (genus *Conus*) are known to contain upwards of 100 individual peptide components in their venoms.

Although venoms are intended to cause harm to a prey or a potential threat, humans have used specific venoms for therapeutic purposes for millennia. Since the advent of modern medicine, researchers have validated many of these uses and have begun to uncover the fundamental mechanisms by which they treat diseases and aid in diagnoses. Some examples of successful venom-derived pharmaceuticals are: Ziconotide (from the cone snail *Conus magus* – treats severe chronic pain)[2], exenatide (from the Gila monster, *Heloderma suspectum* – treats both noninsulin-dependent diabetes mellitus and obesity)[3], and bombesin (from the European fire-bellied toad *Bombina bombina* – treats many gastrointestinal illnesses and, interestingly, sudden deafness)[4].

Venoms and their therapeutic uses represent a major unexplored frontier in computational and systems biology. While some studies have successfully characterized and assessed the biodiversity of venom components in a highly limited context (e.g., the diversity of ω -conotoxins in species of genus *Conus* [5,6], or the evolution of certain protein classes among closely related snake species [7-9]), these studies are limited. There is a severe lack of studies that analyze and assess the properties between – and the relationships among – venoms and their components on a large scale (e.g., across the entire tree of life). Better methods that allow researchers to extract generalizable knowledge about venoms from existing data would enable researchers to discover novel therapeutic uses for them in a systematic and data-driven way.

The most fundamental issue standing in the way of such studies is the almost complete lack of an informatics infrastructure uniting our existing knowledge on venoms. In this study, we present a novel ontology of venoms and related concepts that addresses this problem systematically. Biomedical ontologies allow for consistent and unambiguous naming of entities (in this cases, venoms, venom components, and the species from which they are sourced) and how they are interconnected. We also present a number of initial investigations regarding venom biodiversity across the tree of life, and explore how they can inform the discovery and refinement of novel therapeutic uses for venom compounds.

2. Methods

2.1 Building “Venom Ontology”

We used Protégé (ver. 4.2) [10] to create the class structure of the Venom Ontology using domain knowledge: By our definition, every venomous species has exactly one venom, and every venom has one or more molecular components that can be classified by the class of molecule they are (e.g., peptide, carbohydrate, inorganic cofactor). Recent reports suggest *Conus geographicus* modifies its venom based on whether it is used defensively or offensively [11], but for the purposes of this ontology they can be grouped together as a single venom. If a venom component is a peptide, it has a canonical amino acid sequence. Each of the entities may have one or more other pieces of metadata, including links to other ontologies and structured terminologies. After defining the class structure of the ontology, we populated the ontology with individuals (specific instances of the ontology’s classes) sourced from UniProtKB/Swiss-Prot’s Tox-Prot database [12]. This database is a manually curated list of venom peptides containing numerous annotation tags including species of origin, amino acid sequences, full taxonomic lineage, and automated cross-mappings to other online resources. However, the structure of Tox-Prot does not support semantic reasoning. Due to the large number of individual records in the Tox-Prot database (6,092 at the time of creating the ontology), we added the contained information programmatically by first exporting the ontology from Protégé to an RDF-formatted XML file [13], and then using Apache’s Jena framework [14] to parse the venom records and insert relevant data into the appropriate spot within the ontology’s class hierarchy.

2.2 Exploratory analysis of Venom Ontology data records

To demonstrate some potential applications of Venom Ontology, we performed three exploratory analyses of its contained data. The first of these involved assessing the similarity of amino acid sequences for venom peptides produced by species of the same genera. To accomplish this, we grouped species (stored as “Organisms” in the ontology) by genus, along with their derived peptide compounds. We then selected 2 genera that are well represented in the data set, and built “sequence similarity networks” for each of them. In selecting these genera, we looked for ones that are prolific enough within the ontology to generate informative (non-trivial) networks, yet not so prolific as to be unwieldy in terms of visualization or computation. In practice, we looked for two genera with approximately 20 species in the ontology. For each genus selected, we used BLASTp [15] to align all pairs of peptides within the genus. We constructed the networks using peptide sequences as nodes, and the alignments between them as edges. We transformed the BLAST scores (which represent the percent coverage of the alignments; denoted S) for alignments using the following equation:

$$S' = \frac{1}{e^S}$$

which allows us to define a “distance” between two peptide sequences (i.e., smaller values of S' indicate higher similarity), used as edge weights in the final networks. S' is a value in the interval $[0,1)$, and is generally very small (e.g., $< 1 \cdot 10^{-15}$). Finally, we filtered edges by setting a maximum expect value (“e-value” – a normalized p -value defining confidence that the alignment is non-random) threshold of $1 \cdot 10^{-50}$. Alignments that fell below this maximum cutoff almost certainly signify evolutionarily related sequences, and are therefore informative for the purposes of constructing these networks. For visualization purposes, we rendered the networks in Cytoscape [16] using the Prefuse force-directed layout [17], and colored nodes (individual peptides) by the species from which those peptides were sourced.

Our second analysis was a basic exploration of the distribution of both species and individual peptides in the ontology across the tree of life. We defined common groupings of animals (cnidarians, molluscs, insects, arachnids, fish, amphibians, reptiles, birds, and mammals) that may contain venomous species. From these large classes, we used NCBI’s Taxonomy database [18] to determine the highest-level taxa common to all members of those groups (grouping multiple taxa for paraphyletic groups, such as “fish”). For each of these taxa, we searched for their frequency of occurrence in the set of all species present in the database. We also enumerated the number of total sequences in the database for the groups listed.

The third and final analysis consisted of observing the complexity of venoms within the ontology. In this context, we simplistically define complexity as the number of distinct peptide components within the venom (e.g., a venom containing 20 peptide components is *more complex* than a venom containing only 10). We investigated the distributions of venom complexity for each of the taxonomic groups mentioned in the previous paragraph, making note of features such as mean number of peptide components per venom, standard deviation, and skewness (i.e., lack of symmetry, computed as the estimated third standardised moment $E[x^3]$). It should be noted that the results of these

analyses are subject to systematic biases depending on how well the data in Tox-Prot is representative of the totality of venoms that exist in nature (refer to §4.5 for further discussion).

3. Results

All code and data files used in this study are available for public use on GitHub at (http://github.com/JDRomano2/venom_ontology_code). The ontology is available online, hosted both on BioPortal (<http://bioportal.bioontology.org/ontologies/CU-VO>) and on the project's homepage, at <http://venomkb.tatonettilab.org/ontology>. A visualization of the ontology's class hierarchy and object property associations is shown in **Figure 1**.

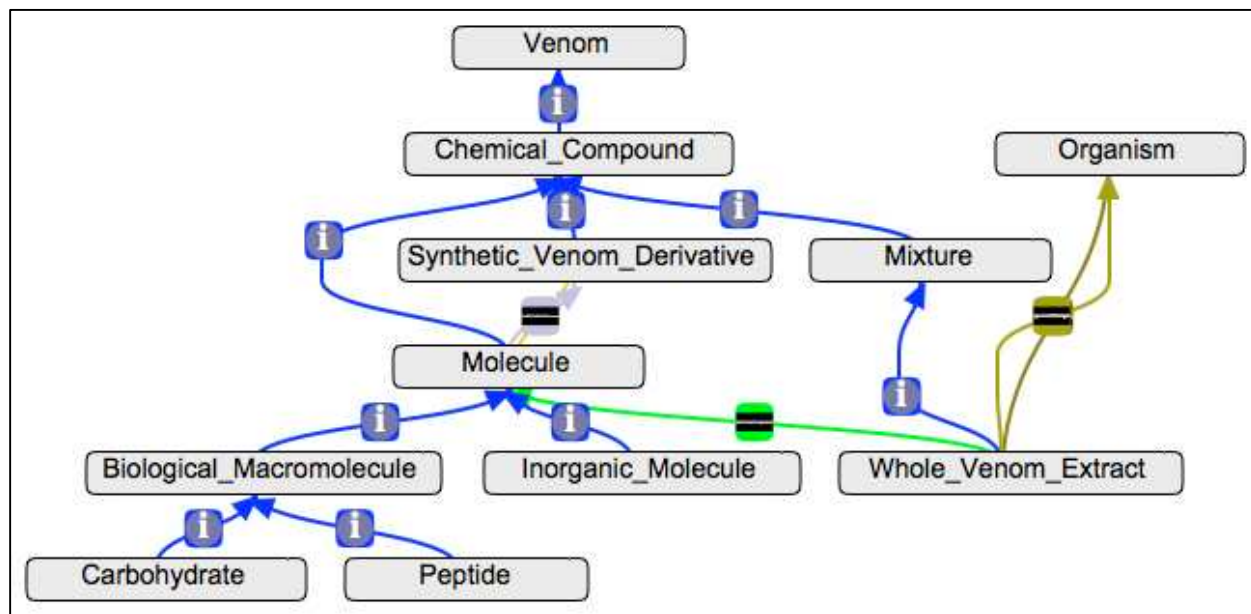


Figure 1. A schematic diagram of classes and class relationships in the Venom Ontology. Blue arrows denote “is_a” relationships, while other colors denote object property relationships. Diagram automatically generated by the Protégé plugin “OBO Graph View”.

3.1 Venom Ontology

Venom Ontology presently contains 614 known venomous species, and 6,092 curated peptides, each of which has a known amino acid sequence. There are correspondingly 614 “whole venom extract” entities, arising from the following axiom:

$$\text{Organism} \exists \geq 1 \text{ hasVenom.Whole_Venom_Extract} \sqcap \leq 1 \text{ hasVenom.Whole_Venom_Extract}$$

which states that every organism has exactly one whole venom extract. Due to our data source being peptide-centric, each whole venom extract (and, correspondingly, each organism) currently included in the ontology has at least one peptide, although this is not defined as necessary (i.e., the ontology allows for whole venom extracts to contain zero or more peptides). We added a small number of synthetic venom compounds (all clinically approved drugs) to the ontology by manually entering them as individuals for the “Synthetic_Venom_Derivative” class. This is a tractable approach presently, but as venom-derived therapeutic agents continue to be discovered and are coerced into a structured format, an automated means for adding them will become necessary – this point is elaborated on below, in §4.5. Venom Ontology was validated using the FaCT++ reasoning engine [19].

3.2 Analysis of the ontology's contained data

Our analysis of venom peptide sequence similarity for a number of well-represented genera highlights some noteworthy features of venoms that have significant implications for drug discovery. In **Figure 2**, we show two sequence similarity networks – one for genus *Loxosceles* (widow spiders) and one for *Bungarus* (kraits – a genus of venomous snakes) – yet our methods could be applied to any other taxonomic group that is present in the ontology. Since we only kept alignments with strong statistical support (low e-value – see §2.2 for details), the graphs are not

fully connected. Small connected components (e.g., the “islands” seen around the periphery of the networks) as well as clusters within larger connected components can be interpreted as groups of peptides that are likely to be closely related on a structural level. Although we originally expected sequences from a given species to segregate together, there are clusters in each of the networks that contain a diverse mixture of sequences from numerous species (denoted in the **Figure 2** by red arrows). The smaller connected components tend to be more homogeneous in terms of their species composition (e.g., they have higher cluster purity). Subjectively, it is also noteworthy that the networks do not display the properties of “scale-free” networks (characterized primarily by few nodes of very high degree, and many nodes of very low degree), which are arguably the most prevalent family of networks that arise from biological phenomena [20]. While speculation as to why this occurs is beyond the scope of this exploratory analysis, it would be an interesting topic to pursue in a follow-up study.

The distribution of species and sequences by higher taxonomic groupings is shown in **Table 1**. Both “fish” and “reptiles” are common names that consist of multiple clades (i.e., they are paraphyletic). It should be noted that 5 species, containing a total of 1,348 sequences, are not classified within any of these groups. While this only makes up 0.81% of the total number of species, it contains 22.13% of the total number of sequences found in the ontology. This seems to be the result of numerous sequences that have poorly formed or absent “taxonomic lineage” annotations in Tox-Prot (meaning that some of the ‘orphaned’/unclassified sequences likely come from already classified species that are included in the larger taxonomic groups).

After looking at properties of venoms exposed by the ontology at the genus level, we investigated the distribution more generally across the tree of life. Distributions of venom complexity are shown in **Table 2**. In this portion of the data analysis, we only show the common taxonomic groups from **Table 1** that have at least 1 venom and 1 peptide. The final row of the table shows the distribution across all species present in the ontology. Additionally, **Figure 3** shows a graphical representation of these distributions, drawn as violin plots with a logarithmic scale on the vertical axis.

Table 1. Distribution of species and sequences in Venom Ontology across common taxonomic groups. Some groups with no species or sequences are included for completeness.

Common Name	Taxonomic group(s)	# species in ontology	% total species	# sequences in ontology	% total sequences
Cnidarians	Cnidaria	0	0.00%	0	0.00%
Molluscs	Mollusca	97	15.80%	1089	17.88%
Insects	Insecta	79	12.87%	245	4.02%
Arachnids	Arachnida	183	29.80%	1089	17.88%
Fish	Actinopterygii	4	0.65%	12	0.20%
	Coelacanthimorpha	0	0.00%	0	0.00%
	Chondrichthyes	1	0.16%	2	0.03%
	Cyclostomata	0	0.00%	0	0.00%
Amphibians	Dipnoi	0	0.00%	0	0.00%
	Amphibia	2	0.33%	3	0.05%
Reptiles	Archelosauria	0	0.00%	0	0.00%
	Squamata	242	39.41%	2298	37.72%
Birds	Aves	0	0.00%	0	0.00%
Mammals	Mammalia	1	0.16%	6	0.10%
Other/unclassified		5	0.81%	1348	22.13%
Total		614		6092	

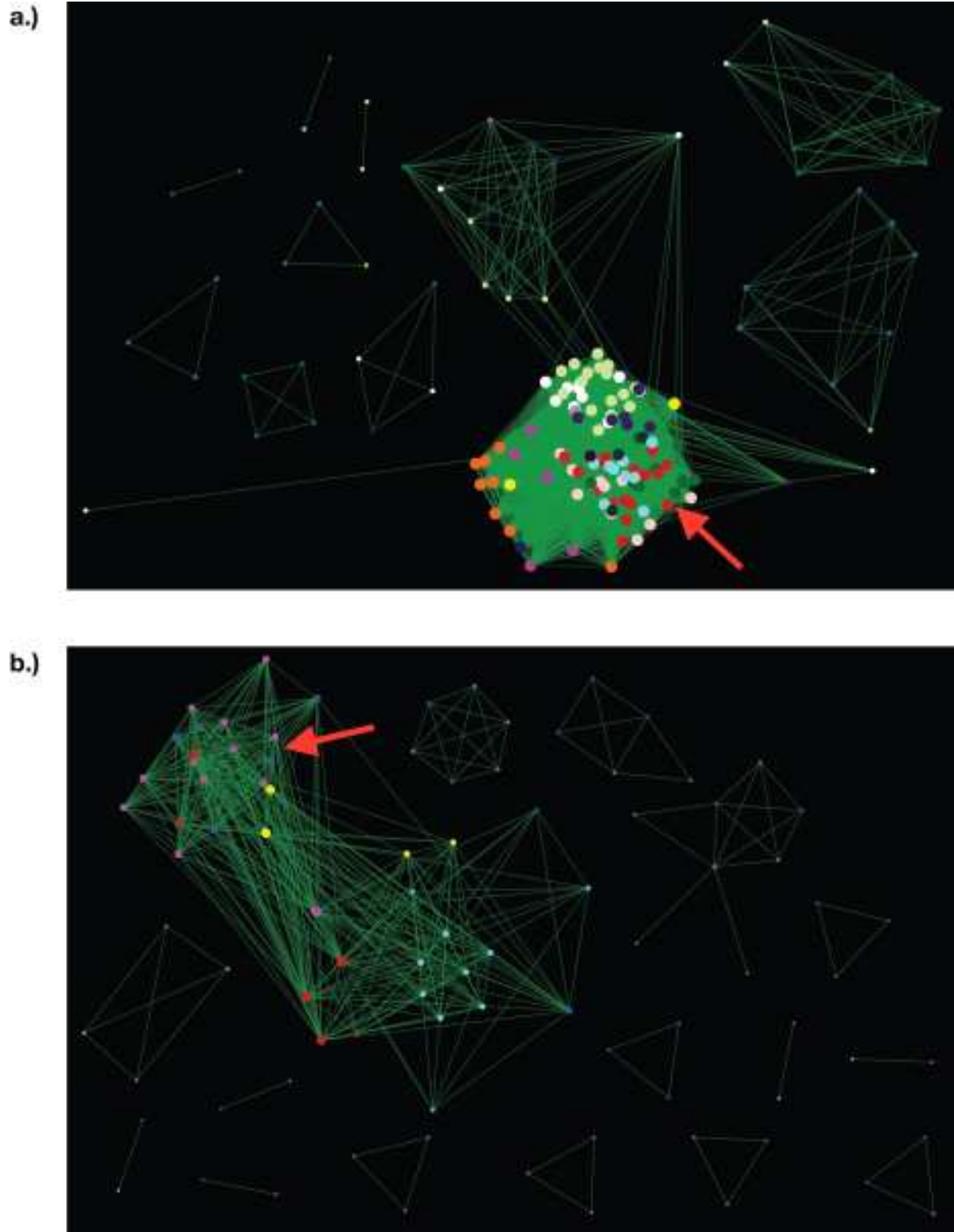


Figure 2. Two sequence similarity networks for venom peptides within the same genus. a.) shows peptides from species in genus *Loxosceles*, and b.) shows peptides from species in genus *Bungarus*. Relative node size is based on the degree of the node, and length of the edges is based on the inverse BLASTp score (see eq. (1)). Nodes of the same color are peptides from the same species of animal. Red arrows indicate “clusters” with high species diversity (i.e., similar peptides found in a number of closely related species).

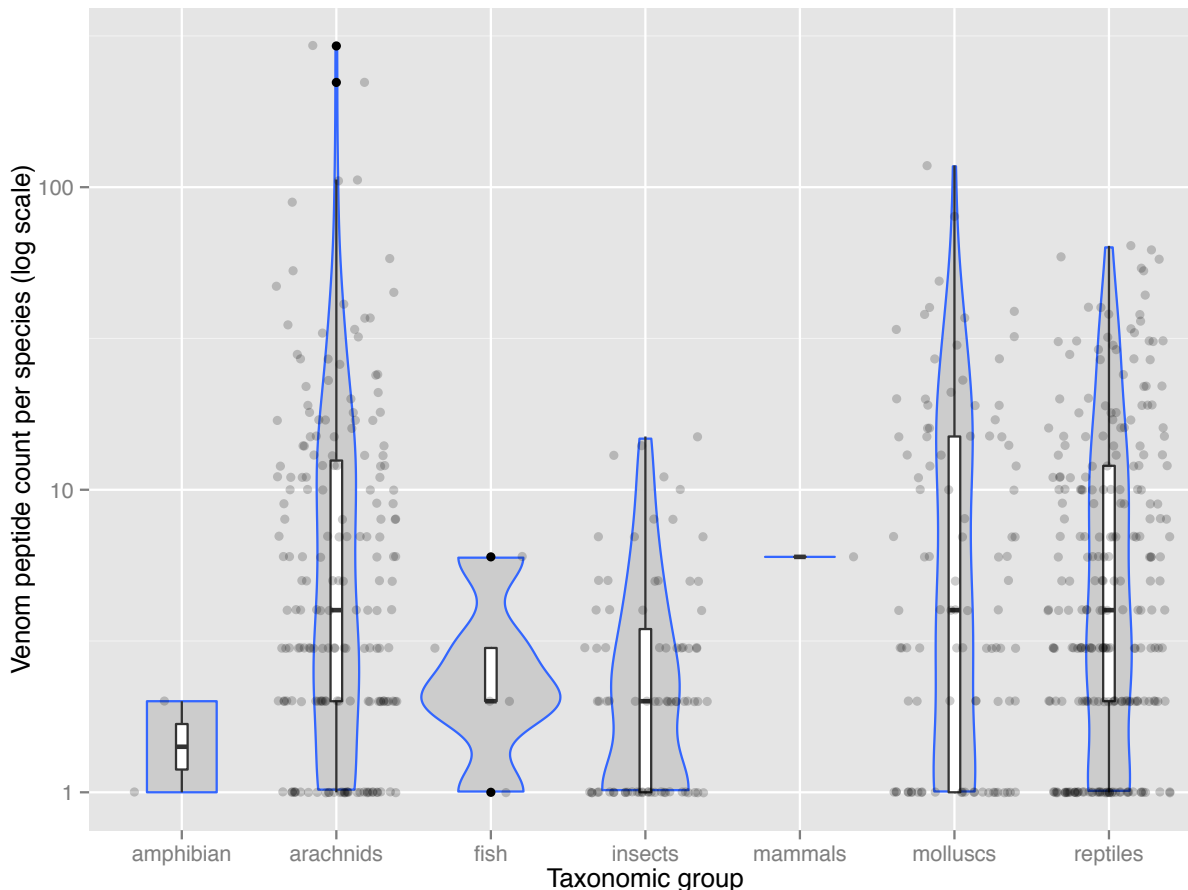


Figure 3. Violin plots showing distributions of venom complexity in 7 common taxonomic groups. Numeric summary statistics are listed in Table 2 for each of the groups shown. Complexity is measured as the number of venom peptides in Venom Ontology for a single species – the vertical axis is the complexity measure for a given species, and the widths of individual plots correspond to the density of the distribution at that complexity measure. Individual species are shown as transparent dots – they are spread horizontally (“jittered”) to better visualize dense groups of data points.

4. Discussion

4.1 Some ontology classes possess no individuals, yet are still informative

The Venom Ontology contains a number of terminal classes that do not currently have any members (“individuals”), including the venom component subclasses “Biological_Macromolecule/Carbohydrate” and “Inorganic_Molecule”. The main rationale for their inclusion is threefold: (1) The ontology is meant to convey computable semantic knowledge of venoms, and with the current structure ontology reasoning software is able to understand that venoms may contain a number of different components, of which only some may be peptides. (2) Since future revisions to the ontology may incorporate new data sources, we hope to be able to populate these classes with informative instances in a future release. (3) We hope to be able to generate members for these classes using machine learning methods that don’t require a curated dataset of venom components (such as “ontology learning from text” [21]). Another class – “Synthetic Venom Derivative” – seems to be specific enough to allow for manual population using domain knowledge of existing synthetic versions of venoms used as pharmaceuticals. However, existing synthetic venom derivatives are more numerous than it would initially seem. For example, a number of conantokins (a specific sub-class of conotoxins – sourced from snails in the genus *Conus*) have been modified and produced synthetically, yet none have received approval for clinical use [22,23]. For this reason, a potential follow-up to this study would be a comprehensive survey of synthetic derivatives of venom peptides.

4.2 Grouping venom peptides by genus reveals clusters of similar venoms across species

As briefly alluded to in §3.2, the networks in **Figure 2** show clusters of venom peptides that contain members from a number of closely related species. This suggests a novel approach for discovering libraries of therapeutic venom-derived peptides with a similar therapeutic effect. During drug development, having a large number of drug candidates available improves the likelihood of finding a molecule that simultaneously has the greatest therapeutic effect while minimizing toxic effects (a notoriously challenging obstacle in repurposing venoms for clinical use). This proposed approach provides a data-driven framework for discovering venom-derived therapeutic agents, which is an improvement over traditional methods that are almost entirely based on serendipitous discovery or borrowed from ancient traditional medicine [24].

4.3 Non-reptile venomous species are underrepresented in existing data

Recent analyses of venom biodiversity reveal surprising patterns, including that the prevalence of venomous fish is far higher than in any other major taxonomic group, including reptiles [25]. Table 2, however, shows a strong bias towards venomous reptiles in available data (fish peptides make up only 0.23% of venom sequences in the Tox-Prot dataset, while reptilian peptides make up 37.72%). Other discrepancies are also apparent: for example, only one venomous mammal is included in the database: *Ornithorhynchus anatinus* (duck-billed platypus). While it is uncommon for mammals to be venomous, reviews on the subject have identified numerous others aside from *O. anatinus*, including multiple shrews, bats, and certain species of loris (taxonomic family Lorinae).

By knowing about these discrepancies, we can prioritize future venom research to include presently underrepresented categories of animals, which should in-turn increase the likelihood of discovering novel compounds that have diverse therapeutic effects.

Table 2. Distribution of venom complexity across the tree of life, by common taxonomic groups. A venom's complexity is defined as the number of distinct peptide components it contains.

Common Name	Minimum	Median	Mean	Maximum	Skewness*
Molluscs	1	4	11.230	118	3.638
Insects	1	2	3.101	15	2.211
Arachnids	1	4	13.020	293	6.576
Fish	1	2	2.800	6	1.517
Amphibians	1	1.5	1.500	2	n/a
Reptiles	1	4	9.496	64	2.271
Mammals	6	6	6.000	6	n/a
All Species	1	4	9.922	293	7.987

* Skewness is the estimated third standardised moment of the empirical distribution. Higher skewness indicates greater lack of symmetry about the mean

4.4 Apparent complexity of venoms varies across the tree of life

Venoms usually consist of a complex mixture of organic and inorganic molecules, each of which has a particular effect. If we define “complexity” as the number of distinct peptide components in a venom, our results show that venom complexity is highly variable across the tree of life. In **Table 2** we list summary statistics for venom complexity distribution across 7 common taxonomic groupings. These data are additionally visualized in **Figure 3** as a violin plot. The plot, shown with number of peptides per venom on a logarithmic scale, highlights that there are many outliers in the dataset – species with extremely complex venoms compared to the mean of 9.922 peptides per venom. Furthermore, each of the taxonomic groups has its own unique distribution. Although the sizes of some groups in the ontology are too small to result in viable statistical inferences (e.g., mammals and amphibians), variable distributions of venom complexity suggest that complexity is regulated in some manner that is conserved by evolution – otherwise, all of the distributions would converge. In particular, insects seem to have venoms that are relatively simple compared to arachnids, molluscs, and reptiles. Interestingly, arachnids have the largest number of outlier species that have extremely complex venoms. Reptiles, by far the most well-represented group in the dataset, have notably fewer highly complex outliers than either molluscs or arachnids. As an example of a quantitative approach to comparing these distributions, **Table 3** shows the *p*-values of the Mann-Whitney *U* test applied pairwise to all of the distributions shown in **Figure 3**. These observations may be an artifact of data completeness (see §4.6), but if not, they can help to guide research towards more rich libraries of venoms that may include important therapeutic compounds.

Table 3. Mann-Whitney U test results for all pairs of venom complexity distributions. A p -value of less than 0.05 signifies that two distributions are statistically different.

	Arachnids	Fish	Insects	Mammals	Molluscs	Reptiles
Amphibians	0.1165	0.417	0.4389	0.6667	0.1669	0.1262
Arachnids		0.1554	1.85E-07	0.8424	0.9086	0.7252
Fish			0.732	0.3657	0.2162	0.1701
Insects				0.1935	2.74E-05	2.20E-07
Mammals					0.8584	0.8128
Molluscs						0.8778

4.5 Using Venom Ontology in conjunction with VenomKB to support drug discovery

In a previous study, we describe VenomKB – a knowledge base cataloguing putative therapeutic uses of venoms and venom-derived compounds, constructed via manual literature review and automated knowledge discovery techniques applied to MEDLINE[26]. Linking these two separate data resources may optimize the process of computational drug discovery by implying a polyhierarchical structure on many of VenomKB’s data records (specifically, ones that map to instances in Venom Ontology). For example, if a record in VenomKB describes the therapeutic effect of a compound produced by species “X”, we may be able to find highly similar (and possibly more efficacious) molecules by using Venom Ontology to identify venom peptides from species that are in the same genus as species “X”. In the future, we intend to add a component to the ontology that resolves venom names with their synonyms, which could allow us to identify venoms with multiple therapeutic effects, as well as increase confidence in therapies when multiple studies corroborate the same effect. We plan to fully integrate these two resources, so that VenomKB can be browsed by navigating the hierarchical structure of Venom Ontology, and vice versa.

4.6 Limitations – Structured data on venoms are largely incomplete

One of the most important features of this set of investigations is that it is not based on a complete survey of venoms across the tree of life. Since we only have discovered a small handful of the vast number of venomous species believed to exist (and have actually studied even a smaller number), we treat the Tox-Prot dataset as a “best approximation” of venom diversity based on available data. This obviously introduces various sources of systematic bias into the inferences that are made from the ontology’s contained data.

In §2.2, we mention this limitation in regards to our definition of venom complexity (the relative number of peptide components contained within a venom). We analyze our data under the assumption that the Tox-Prot data set does not prioritize certain species for “completeness” – in other words, that the ratio of the actual number of peptides to the number that are in the data set remains consistent for all species. However, this may not be the case. The available data for some species may be substantially more complete than for others. Also, it may be more challenging to run proteomic analyses on some species than others. Each of these factors would affect the consistency of completeness across the dataset. A future goal that could help eliminate these potential sources of bias would be only to populate the ontology with complete proteomic surveys of species’ venoms.

We intend for the Venom Ontology to be one of the first major steps towards systematically and consistently coercing newly discovered venoms and venom components into a standardized format. The ontology’s structure suggests numerous ways to define a consistent vocabulary for these semantic concepts.

5. Conclusion

Although the study of venoms traditionally falls into the fields of biological systematics, toxicology, biodiversity, and ecology, they constitute a largely untapped library that could be useful for drug discovery purposes. Recently, venoms have piqued the interest of researchers from diverse fields, including both translational bioinformatics and systems pharmacology. Additionally, venom research requires a substantial amount of informatics infrastructure to support rigorous and informative computational and data-driven analysis. In this study, we provide a novel ontology of venom-related concepts that facilitates many of these analyses. In demonstrating possible uses for the ontology, we have discerned several important features of venoms that suggest promising areas to study in the future, including studying venomous species that are underrepresented in existing data sets, and using peptide sequence characteristics to build libraries of related venomous compounds that may improve on existing drugs derived from venoms. As the ontology grows in future releases, we expect to discover many more ways the ontology can facilitate data-driven research of

venoms, greatly advancing knowledge and available methodology in a diverse range of medical and biologically based fields.

6. Acknowledgements

We thank Mary Regina Boland for valuable comments on a previous version of the manuscript. JDR is supported by NIH T15 LM00707. JDR and NPT are supported by R01 GM107145. The authors report no conflicts of interest.

References

- 1 Kordiš D, Gubenšek F. Adaptive evolution of animal toxin multigene families. *Gene* 2000;**261**:43–52. doi:10.1016/S0378-1119(00)00490-X
- 2 Sanford M. Intrathecal ziconotide: a review of its use in patients with chronic pain refractory to other systemic or intrathecal analgesics. *CNS Drugs* 2013;**27**:989–1002. doi:10.1007/s40263-013-0107-5
- 3 Mafong DD, Henry RR. Exenatide as a treatment for diabetes and obesity: implications for cardiovascular risk reduction. *Curr Atheroscler Rep* 2008;**10**:55–60.
- 4 Kaneto A, Kaneko T, Nakaya S, *et al.* Effect of bombesin infused intrapancreatically on glucagon and insulin secretion. *Metab Clin Exp* 1978;**27**:549–53.
- 5 Olivera BM, Teichert RW. Diversity of the neurotoxic Conus peptides: a model for concerted pharmacological discovery. *Mol Interv* 2007;**7**:251–60. doi:10.1124/mi.7.5.7
- 6 Olivera BM, Rivier J, Clark C, *et al.* Diversity of Conus neuropeptides. *Science* 1990;**249**:257–63.
- 7 Gutiérrez JM, Lomonte B. Phospholipase A2 myotoxins from Bothrops snake venoms. *Toxicon* 1995;**33**:1405–24.
- 8 Yang CC. Chemistry and evolution of toxins in snake venoms. *Toxicon* 1974;**12**:1–43.
- 9 Rusmili MRA, Yee TT, Mustafa MR, *et al.* Proteomic characterization and comparison of Malaysian Bungarus candidus and Bungarus fasciatus venoms. *J Proteomics* 2014;**110**:129–44. doi:10.1016/j.jprot.2014.08.001
- 10 Noy NF, Crubezy M, Ferguson RW, *et al.* Protégé-2000: an open-source ontology-development and knowledge-acquisition environment. *AMIA Annu Symp Proc* 2003;**2003**:953.
- 11 Dutertre S, Jin A-H, Vetter I, *et al.* Evolution of separate predation- and defence-evoked venoms in carnivorous cone snails. *Nat Commun* 2014;**5**:3521. doi:10.1038/ncomms4521
- 12 Jungo F, Bougueleret L, Xenarios I, *et al.* The UniProtKB/Swiss-Prot Tox-Prot program: A central hub of integrated venom protein data. *Toxicon* 2012;**60**:551–7. doi:10.1016/j.toxicon.2012.03.010
- 13 Beckett D, McBride B. RDF/XML syntax specification (revised). *W3C recommendation* 2004.
- 14 McBride B. Jena: A semantic web toolkit. *IEEE Internet computing* 2002.
- 15 Camacho C, Coulouris G, Avagyan V, *et al.* BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;**10**:421. doi:10.1186/1471-2105-10-421
- 16 Shannon P, Markiel A, Ozier O, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**:2498–504. doi:10.1101/gr.1239303
- 17 Heer J, Card SK, Landay JA. Prefuse: a toolkit for interactive information visualization. 2005.

- 18 Federhen S. The NCBI Taxonomy database. *Nucleic Acids Res* 2012;**40**:D136–43. doi:10.1093/nar/gkr1178
- 19 Tsarkov D, Horrocks I. FaCT++ Description Logic Reasoner: System Description. In: *Automated Reasoning*. Berlin, Heidelberg: : Springer Berlin Heidelberg 2006. 292–7. doi:10.1007/11814771_26
- 20 Barabási A-L. Scale-free networks: a decade and beyond. *Science* 2009;**325**:412–3. doi:10.1126/science.1173299
- 21 Wong W, Liu W, Bennamoun M. Ontology learning from text: A look back and into the future. *ACM Computing Surveys (CSUR)* 2012.
- 22 Robinson SD, Norton RS. Conotoxin gene superfamilies. *Mar Drugs* 2014;**12**:6058–101. doi:10.3390/md12126058
- 23 Cnudde SE, Prorok M, Castellino FJ, *et al.* Metal ion determinants of conantokin dimerization as revealed in the X-ray crystallographic structure of the Cd(2+)/Mg (2+)-con-T[K7gamma] complex. *J Biol Inorg Chem* 2010;**15**:667–75. doi:10.1007/s00775-010-0633-2
- 24 Lewis RJ, Garcia ML. Therapeutic potential of venom peptides. *Nat Rev Drug Discov* 2003;**2**:790–802. doi:10.1038/nrd1197
- 25 Smith WL, Wheeler WC. Venom evolution widespread in fishes: a phylogenetic road map for the bioprospecting of piscine venoms. *J Hered* 2006;**97**:206–17. doi:10.1093/jhered/esj034
- 26 Romano JD, Tatonetti NP. VenomKB, a new knowledge base for facilitating the validation of putative venom therapies. *Sci Data* 2015;**2**:150065. doi:10.1038/sdata.2015.65