Journal of Proteomics 135 (2016) 62–72

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

# Ecological venomics: How genomics, transcriptomics and proteomics can shed new light on the ecology and evolution of venom

Kartik Sunagar [a], David Morgenstern [b,*], Adam M. Reitzel [c], Yehu Moran [a,*]

[a] Department of Ecology, Evolution and Behavior, Alexander Silberman Institute of Life Sciences, Hebrew University of Jerusalem, Jerusalem 91904, Israel
[b] Proteomics Resource Center, Langone Medical Center, New York University, New York, USA
[c] Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

ARTICLE INFO

*Article history:*
Received 15 July 2015
Received in revised form 2 September 2015
Accepted 9 September 2015
Available online 15 September 2015

*Keywords:*
Venom
Toxin
Genomics
Transcriptomics
Proteomics
Venomics

ABSTRACT

Animal venom is a complex cocktail of bioactive chemicals that traditionally drew interest mostly from biochemists and pharmacologists. However, in recent years the evolutionary and ecological importance of venom is realized as this trait has direct and strong influence on interactions between species. Moreover, venom content can be modulated by environmental factors. Like many other fields of biology, venom research has been revolutionized in recent years by the introduction of systems biology approaches, i.e., genomics, transcriptomics and proteomics. The employment of these methods in venom research is known as 'venomics'. In this review we describe the history and recent advancements of venomics and discuss how they are employed in studying venom in general and in particular in the context of evolutionary ecology. We also discuss the pitfalls and challenges of venomics and what the future may hold for this emerging scientific field.



## 1. Introduction

Venom is defined as a secretion, produced in a specialized gland or cell of one animal that is actively delivered to the target animal through the infliction of a wound. It is a complex mixture of toxin peptides, proteins, salts and other chemicals employed mostly for prey capture and/or for defense from predators and aggressors [1,2]. Further, venoms are also employed by some animals such as leeches, ticks and vampire bats to facilitate their specialized blood feeding habits [3,4]. As venoms are characterized by an unusual diversity of components, and may contain hundreds of toxin peptides they generate considerable interest among evolutionary biologists and biochemists alike. The structural variability of animal toxins is remarkable [5] and they exhibit a large array of biological activities [1]. Genes coding these peptides are hypothesized to evolve under positive Darwinian selection due to their participation in an evolutionary "arms race", where the evolution of venom resistance in prey and the invention of potent venom components in the secreting animal exert reciprocal selection pressures [6–8]. On the other hand, it is clear now that several protein families with non-venomous functions are recurrently and independently recruited into the venoms of different animal lineages [3], and that some venoms evolve mostly under purifying selection with only episodic positive selection [9]. The study of venoms is of growing interest for the pharmacological and biotechnological communities as they are increasingly recognized as a rich source for lead compounds that can drive forward the development of insecticides and pharmaceutical drugs [10–12]. However, venom can also serve as a model system where the relationship between genetic variability, protein biochemistry and interspecific interactions is more amenable for elucidation and definition than in most other systems. After all, an increase in the potency of venom can directly increase the fitness of the predator secreting the venom and decrease the fitness of the prey or vice versa, depending on which species is venomous. Thus, venom potency can directly affect the strength of antagonistic interaction between the prey and the predator. Hence, evolutionary ecology can greatly benefit from the study of venom. Unfortunately, in many cases where much is known about venom composition we lack knowledge regarding the ecology of the venomous animal and in many cases where the ecology is well-understood little is known about venom composition. There are indications from snakes that the diet of a venomous species might be closely-tied to its venom composition [13–15]. Moreover, it seems that cone snails employ different venom arsenals for prey capture and defense [16] and that scorpions can control the composition of the mixture they inject via their sting [17]. It is plausible that such tight relationships between interspecific interactions and venom composition might be due to the high metabolic cost of venom production, which makes potent venom advantageous, as less venom is required for neutralizing prey or foe [18]. However, to better understand such selective pressures we need a much better picture of the

* Corresponding authors.
*E-mail addresses:* venomologist@gmail.com (D. Morgenstern), yehu.moran@mail.huji.ac.il (Y. Moran).

http://dx.doi.org/10.1016/j.jprot.2015.09.015
1874-3919/

complete venom composition, its temporal dynamics and venom variation between individuals and populations as well as spatiotemporal variations in diet and interspecific interactions. Filling these gaps with the help of "omic" tools could help us understand the ecological factor and evolutionary pressures that shaped the venom and also what role the venom plays in the interspecific interactions of the venomous animal and its ecological niche.

The term "venomics" was first used as a description for proteomic study of snake venom composition [19,20], but in 2006 as a description of an ambitious project aimed to provide a full picture of venom-related biology by sequencing the full genomes, transcriptomes of venom glands and venom protein contents of several venomous animal species [21]. Since then it has also been used for describing studies of much smaller scales that use several systems biology approaches for the comprehensive investigation of venom components, e.g., studies combining shotgun mass spectrometry analysis of venom components with transcriptome sequencing of a venom gland for charting an extensive list of venom contents and getting a more complete picture on the proteomic landscape of venoms and their biology [22,23]. The field of venomics is growing fast but even with the current availability of new high-throughput methods, the de novo sequencing of genomes, transcriptomes and proteomes is far from an easy feat. In this perspective paper we will discuss the current state of knowledge, the pitfalls and challenges awaiting for studies in this field and what are the future directions that could help it grow in the context of evolutionary ecology.

## 2. Genomics and transcriptomics in venom research

### 2.1. Genomics of venomous animals

The DNA sequencing methods developed in parallel by Sanger and Gilbert brought a true revolution to the field of biology [24,25]. Roughly 25 years later in a remarkable project by a large consortium and an investment of more than 3 billion US dollars and 10 years of work, the full human genome was sequenced by applying the method developed by Sanger [26]. As this method is costly and relatively slow, using it to fully sequence a genome of several hundreds of millions of base pairs (bp) would cost several millions of US dollars and would take several years. However, a full-genome sequence of an organism is priceless for biologists as it can give the full repertoire of genes, not only those that are actively transcribed, but also the many additional genetic features such as introns, intergenic regions and cis- and trans-transcriptional regulatory elements that have pivotal roles in the control of gene expression and in the evolution and physiology of an organism [27]. Still, the astounding costs have put a severe limitation on the ability to sequence full animal genomes and very few genomes of venomous animals have been sequenced to date. As an organism of pivotal agricultural importance the European honeybee *Apis mellifera* was the first venomous animal to be sequenced [28] (Table 1). However, it has an extremely streamlined venom with only a handful of components [29]. The next animal to have its genome sequenced was the starlet anemone *Nematostella vectensis* that serves as an important model organism in evolutionary developmental biology studies [30] (Table 1). This species is a representative of the venomous phylum Cnidaria (sea anemones, corals, jellyfish and hydroids), but the contents of its venom were unknown before its genome was sequenced [31]. Thus, this species provided the first example for a genome-guided venom discovery [32,33]. Following *Nematostella*, the genomes of other cnidarians such as *Hydra magnipapillata* and the reef-building coral *Acropora digitifera* were sequenced as well [34,35] (Table 1). Another genome of a venomous animal to be sequenced by the Sanger method is that of platypus (*Ornithorhynchus anatinus*), which uses its venom for intraspecific aggression among males during the mating season [36]. Lastly, the full genome of the parasitic wasp *Nasonia vitripennis* was sequenced by the same method as well as partial genomic sequence of two other *Nasonia* species [37]. The females of these tiny wasp species inject venom not only for paralyzing host fly larvae, but also for manipulating the host's gene regulation, possibly for making it more suitable for the needs of their progeny [38]. The sequencing of the *Nasonia* genome provided an important tool in the study of their venom that can be obtained only in miniscule amounts, and hopefully provides an important step in the direction of understanding how these complex genetic manipulations are achieved [39].

The introduction of high throughput sequencing techniques, also known as Next Generation Sequencing (NGS), in the last decade revolutionized the field of genomics. Initially, pyrosequencing technology platforms, such as those by 454 and then reversible terminator technology platforms such as the ones from Illumina, made sequencing of millions of bps by a single run possible [40]. In recent years NGS outputs are booming, for example, at the time of writing this manuscript (July 2015), the Illumina HiSeq 4000 sequencing system offered the ability to sequence 1500 gigabases per run, i.e., to sequence about 12 human genomes at a ×30 coverage in one run (Table 2). This second genomic revolution has had a profound effect on all fields of biology as sequencing costs per bp are constantly decreasing. The NGS technologies enabled the sequencing of the genomes of the scorpion *Mesobuthus martensii*[41] and the king cobra (*Ophiophagus hannah*) [42] (Table 1). Sequencing of the full genomes of these two animals coupled to RNA-Seq (the sequencing of expressed protein-coding genes via complementary DNA) revealed the nearly-full landscape of venom-encoding genes in these animals. In addition, the genome sequences provided novel insights into several evolutionary and zoological aspects related to venom: from the ability of scorpions to resist their own venom to a possible answer to the enigmatic origin of venom glands in snakes [41,42].

NGS also enabled the sequencing of the first two spider genomes [43] (Table 1). The vast majority of known spiders are venomous, and

**Table 1**
Sequenced genomes of venomous animals.

| Organism | Approximate genome size (million base pairs) | Sequencing platform (see Table 2 for details) | Sequencing depth | Scaffold N50 (thousand base pairs) | References for genome and venom |
|---|---|---|---|---|---|
| Arthropods | | | | | |
| *Acanthoscurria geniculata* (tarantula) | 6500 | Illumina | 40× | 47 | [43] |
| *Apis mellifera* (honey bee) | 262 | Originally Sanger, later 454 and SOLiD | 6× (Sanger); 20× (SOLiD); 4× (454) | Originally 359, later 997 | [28,29,73] |
| *Mesobuthus martensii* (scorpion) | 1323 | Illumina | 248× | 223 | [41,171] |
| *Nasonia vitripennis* (parasitoid wasp) | 295 | Sanger | 6× | 709 | [37,39] |
| *Solenopsis invicta* (fire ant) | 352 | Illumina and 454 | ? | 720 | [46,47] |
| *Stegodyphus mimosarum* (spider) | 2550 | Illumina | 91× | 480 | [43] |
| Cnidaria | | | | | |
| *Hydra magnipapillata* | 1050 | Sanger | 8× | 92.5 | [34,172,173] |
| *Nematostella vectensis* (sea anemone) | 357 | Sanger | 6.5× | 470 | [30,32,174] |
| Vertebrates | | | | | |
| *Ophiophagus hannah* (king cobra) | 1590 | Illumina | ? | 226 | [42] |
| *Ornithorhynchusanatinus*(platypus) | 1840 | Sanger | 6× | 967 | [36,175] |