Intraspecific venom variation in the medically significant Southern Pacific Rattlesnake (Crotalus oreganus helleri): Biodiscovery, clinical and evolutionary implications

Kartik Sunagar\textsuperscript{a,b,1}, Eivind A.B. Undheim\textsuperscript{c,d,1}, Holger Scheib\textsuperscript{d,1}, Eric C.K. Gren\textsuperscript{e,1}, Chip Cochran\textsuperscript{e,1}, Carl E. Person\textsuperscript{e,1}, Ivan Koludarov\textsuperscript{e}, Wayne Kelln\textsuperscript{e}, William K. Hayes\textsuperscript{e}, Glenn F. King\textsuperscript{d}, Agosthino Antunes\textsuperscript{a,b}, Bryan Grieg Fry\textsuperscript{c,d,*}

\textsuperscript{a}Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal
\textsuperscript{b}CIIMAR/CIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Rua dos Bragas 289, P 4050-123 Porto, Portugal
\textsuperscript{c}Venom Evolution Lab, School of Biological Sciences, University of Queensland, St. Lucia, Queensland, Australia
\textsuperscript{d}Institute for Molecular Bioscience, University of Queensland, St. Lucia, Queensland 4072, Australia
\textsuperscript{e}Department of Earth and Biological Sciences, Loma Linda University, Loma Linda, CA 92350, USA

\textbf{ARTICLE INFO}

\textbf{Article history:}
Received 6 September 2013
Accepted 13 January 2014

\textbf{Keywords:}
Venom
Evolution
Molecule
Toxin
Rattlesnake
Crotalus

\textbf{ABSTRACT}

Due to the extreme variation of venom, which consequently results in drastically variable degrees of neutralization by CroFab antivenom, the management and treatment of envenoming by \textit{Crotalus oreganus helleri} (the Southern Pacific Rattlesnake), one of the most medically significant snake species in all of North America, has been a clinician’s nightmare. This snake has also been the subject of sensational news stories regarding supposed rapid (within the last few decades) evolution of its venom. This research demonstrates for the first time that variable evolutionary selection pressures sculpt the intraspecific molecular diversity of venom components in \textit{C. o. helleri}. We show that myotoxic \(\beta\)-defensin peptides (aka: crotamines/small basic myotoxic peptides) are secreted in large amounts by all populations. However, the mature toxin-encoding nucleotide regions evolve under the constraints of negative selection, likely as a result of their non-specific mode of action which doesn’t enforce them to follow the regime of the classic predator–prey chemical arms race. The hemorrhagic and tissue destroying snake venom metalloproteinases (SVMPs) were secreted in larger amounts by the Catalina Island and Phelan rattlesnake populations, in moderate amounts in the Loma Linda population and in only trace levels by the Idyllwild population. Only the Idyllwild population in the San Jacinto Mountains contained potent presynaptic neurotoxic phospholipase \(A_2\) complex characteristic of Mohave Rattlesnake (\textit{Crotalus scutiatus}) and Neotropical Rattlesnake (\textit{Crotalus durissus terrificus}). The derived heterodimeric lectin toxins characteristic of viper venom, which exhibit a diversity of biological activities, including anticoagulation, agonism/antagonism of platelet activation, or procoagulation, appear to have evolved under extremely variable selection pressures. While most lectin \(\alpha\)- and \(\beta\)-chains evolved rapidly under the influence of positive Darwinian selection, the \(\beta\)-chain lectin of the...
Catalina Island population appears to have evolved under the constraint of negative selection. Both lectin chains were conspicuously absent in both the proteomics and transcriptomics of the Idyllwild population. Thus, we not only highlight the tremendous biochemical diversity in C. o. helleri’s venom-arsenal, but we also show that they experience remarkably variable strengths of evolutionary selection pressures, within each toxin class among populations and among toxin classes within each population. The mapping of geographical venom variation not only provides additional information regarding venom evolution, but also has direct medical implications by allowing prediction of the clinical effects of rattlesnake bites from different regions. Such information, however, also points to these highly variable venoms as being a rich source of novel toxins which may ultimately prove to be useful in drug design and development.

Biological significance

- These results have direct implications for the treatment of envenomed patients.
- The variable venom profile of Crotalus oreganus helleri underscores the biodiscovery potential of novel snake venoms.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Knowledge of venom composition has increased dramatically with improvements in technology and the advent of new techniques, in particular the use of mass spectrometry in venom proteomics [1–15] and venom gland transcriptome analysis [16–27]. Snake venoms are complex secretions composed of numerous enzymes, toxins, peptides, small organic molecules, and inorganic components that have diverse modes of action on both prey and human victims [28–32]. Snake venom serves both predatory and defensive purposes [28–30,33–38]. Variation in venom profiles has been shown between species within the same genus [5,11,12,15,27,39–49] and between individuals within the same species, with the intraspecific differences found among geographic locales [2,11,12,45–52], between sexes [46,47,53] and between juveniles and adults [9,46,47,54,55]. Venom variation has also been reported between venom glands of a single individual [56]. Some authors have argued that venom diversity is the product of neutral evolutionary processes and not subject to natural selection [57,58], whereas others have argued that strong natural selection has driven adaptation to particular prey species [12,30,31,40,46,47,59–63].

Venom in reptiles originated from a single early recruitment event approximately 180 million years ago (mya) during the early Jurassic period and is a pliesiotypic trait of the Toxicofera clade [10,12,18,20–23,30,31,40,64]. New World pit vipers are thought to have descended from a single ancestral Asian pit viper species that colonized the New World via the Bering land bridge [65,66], with rattlesnakes having a mid-Cenozoic origin in the Mexican highlands [67–69]. The venom arsenals of Crotaline snakes are characterized by a great diversity of venom components; generalized venom “types” have been proposed, depending upon metalloprotease activity and toxicity [70]. Type I venoms possess high levels of metalloprotease activity and lower toxicity (>1.0 μg/g mouse body weight), whereas type II venoms have low metalloprotease activity and higher toxicity (<1.0 μg/g mouse body weight). The presence of these two venom types in a diversity of well-defined species clades suggests that it is not dependent upon phylogeny [49,52,70–72].

Crotalus oreganus helleri is a medium-sized rattlesnake inhabiting Baja California northward through southern California, and the Pacific islands of Santa Catalina (Los Angeles County, California) and Coronado Del Sur (Tijuana, Mexico) [67]. Pronounced tectonic activity in the region has produced considerable variation in available habitat [73]. The species utilizes habitat ranging from sea level to >3000 m and prey encountered are highly varied. Significant regional variation in venom composition exists [51,74], with both type I and type II venom identified in local populations [49]; however this dichotomy of venom types fails to characterize the full extent of venom variability in the species. C. o. helleri is the most medically relevant species of the region and is responsible for the majority of severe envenomations in southern California [29,75]. Therefore, determining intraspecific variation of C. o. helleri venom components and the factors influencing their molecular evolution can yield important implications for clinical treatment of envenomation. Venom variation also offers substantial potential for bioprospecting and pharmaceutical discovery [8,18–23,30,40,76]. These variations have been the subject of many popular press reports that grossly misattribute them to unparalleled recent diversification of the venom [77] and thus display a fundamental lack of understanding on how venom evolves.

In this study, we investigated the diversity of toxins present in C. o. helleri, across its geographic range, using a combined proteomics–transcriptomics approach to investigate the relative molecular evolution and diversification within a given toxin type, and the relative expression levels of particular toxin types.

2. Materials and methods

2.1. Sampling

We sampled four southern California populations of C. o. helleri from areas with pronounced geological, elevational, and floristic differences. Human envenomations from snakes in these
different regions have exhibited distinct symptoms ranging from hemorrhage to muscle fasciculations to paralysis. The four populations chosen (Fig. 1) were: (1) Catalina Island, which is dominated by coastal sage scrub, interspersed with chaparral and oak woodland, has never been connected to the mainland [73] and has supported an isolated population since at least the Pleistocene; (2) Idyllwild in the San Jacinto Mountains has high altitude pine and cedar montane forests (elevation ~1600 m); (3) Loma Linda consists of low rolling hills covered with grasses and, on north facing slopes, Salvia miltiﬂora and other shrubs; and (4) Phelan comprises a transition zone between High Desert (Mohave) and coastal mountain scrub. We sampled one snake from each region for transcriptome sequencing. We used the same snake for proteome analysis of the Phelan and Loma Linda populations, and a separate individual of same sex and size from the exact same locality for the other two locations in addition to two more specimens for each location other than Loma Linda, for which only one more specimen was obtained due to the rarity of C. o. helleri in this location. We used only adult specimens for venom analysis due to potential ontogenetic shifts in venom composition [9,70].

2.2. Transcriptome sequencing, phylogenetics, selection analyses, and structural analyses

2.2.1. Transcriptome sequencing

Total RNA was extracted from venom glands using the standard TRIzol Plus method (Invitrogen). Extracts were enriched for mRNA using standard RNeasy mRNA mini kit (Qiagen) protocol. mRNA was reverse transcribed, fragmented and ligated to a unique 10-base multiplex identifier (MID) tag prepared using standard protocols and applied to one PicoTitrePlate (PTP) for simultaneous amplification and sequencing on a Roche 454 GS FLX + Titanium platform (Australian Genome Research Facility). An average of 50,000 sequences were read for each library. Automated grouping and analysis of sample-specific MID reads informatically separated sequences from the other transcriptomes on the plates, which were then post-processed to remove low quality sequences before de novo assembly into contiguous sequences (contigs) using v3.4.0.1 of the MIRA software program. Assembly details for the transcriptomes are shown in Supplementary Table 1. All raw reads have been deposited in the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) with the accession numbers of: SRR871501 C. o. helleri (Catalina Island), SRR871502 C. o. helleri (Idyllwild), SRR871503 C. o. helleri (Loma Linda), and SRR871504 C. o. helleri (Phelan). Assembled contigs were processed using CLC Main Work Bench (CLC-Bio) and Blast2GO bioinformatic suite to provide Gene Ontology, BLAST and domain/Interpro annotation. The above analyses assisted in the rationalization of the large numbers of assembled contigs into phylogenetic ‘groups’ for detailed phylogenetic analyses outlined below.

2.2.2. Selection analyses

Translated nucleotide sequences were aligned using MUSCLE 3.8 [78] and the alignments were manually inspected to rectify errors. All nucleotide sequences and multiple sequence alignments used for selection analyses are available as Supplementary file 2 and Supplementary Figs. 1–4, respectively. In order to reconstruct gene phylogenies for selection assessments, maximum-likelihood method implemented in PhyML [79] was employed on the nucleotide datasets and node support was evaluated with 1000 bootstrapping replicates. All the maximum-likelihood trees are provided as Supplementary Figs. 5–7, with the results of branch-site REL test mapped onto them. In order to detect the nature of selection and its influence on various venom-encoding genes of C. o. helleri, we utilized maximum-likelihood models implemented in Codeml of the PAML [80]. We employed site-specific models that estimate positive selection statistically as a non-synonymous-to-synonymous nucleotide-substitution rate ratio (ω) significantly greater than 1. For technical details regarding models/methods see [20,81]. FUBAR [82] implemented in HyPhy [83] was employed to provide additional support to the aforementioned analyses and to detect sites evolving under the influence of pervasive diversifying and purifying selection pressures. Mixed Effects Model Evolution (MEME) [82] was also employed to efficiently detect episodically diversifying sites. To clearly depict the proportion of sites under different regimes of selection, an evolutionary fingerprint analysis was carried out using the evolutionary selection distance (ESD) algorithm implemented in Datamonkey [84]. We further utilized the branch-site Random Effects Likelihood (REL) test [85] to identify lineages evolving under the influence of episodic diversifying selection pressures.

2.2.3. Structural analyses

To depict the natural selection pressures influencing the evolution of various C. o. helleri venom-components (only those with sufficient numbers of full-length sequences were analyzed in this regard; β-defensin, kallikrein and lectin), we mapped the sites under positive selection on the homology models created using Phyre 2 web server [86]. PyMOL 1.3 [87] was used to visualize and generate the images of homology models. ConSurf web server [88] was used for mapping the evolutionary selection pressures on the three-dimensional homology models.

Homology models of the presynaptic PLA2 complex from C. o. helleri (Coh) (GenBank: GAKR01000015 [acid subunit] and GenBank: GAKR01000016 [basic subunit]) and the homologue from Crotalus scutulatus scutulatus (Css) (UniProt: P18998 [acid subunit] and UniProt: P62023 [basic subunit]) were built using the crystal structure of crotroxin from Crotalus durissus terrificus (Cdt).
Proteins were then digested with proteomics-grade porcine pancreatic trypsin (Sigma-Aldrich, St. Louis, MO, USA). We desalted samples using C₁₈ ZipTips (EMD Millipore, Billerica, MA, USA) according to the manufacturer’s protocol. The desalted tryptic peptides were resuspended in mobile phase A (2% acetonitrile, 0.1% formic acid in water). Liquid chromatography was conducted on a ThermoFinnigan LCQ Deca XP spectrometer (ThermoFinnigan, Waltham, MA, USA) equipped with a PicoView 500 nanospray apparatus using Xcalibur software (ver. 1.3; ThermoFinnigan, Waltham, MA, USA) for instrument control and data acquisition. Separation was performed on a 10-cm × 75-μm-i.d. C₁₈ BioBasic bead column (New Objective, Woburn, MA, USA) by injecting 20-μL samples. Mobile phase B consisted of 98% acetonitrile, 2% water, and 0.1% formic acid. The gradient program was: 0% B at 0.18 mL/min for 7.5 min; 0% B at 0.35 mL/min for 0.5 min; linear gradient to 20% B at 15 min at 0.35 mL/min; linear gradient to 75% B at 55 min at 0.3 mL/min (flow rate constant for remainder of the program); linear gradient to 90% B at 60 min; hold at 90 B until 85 min; linear gradient to 0% B at 90 min; hold at 0% B until 120 min. Spectra were acquired in positive ion mode with a scan range of 300–1500 m/z. We converted MS/MS data into peak list files using ExtractMSn implemented in BioWorks (version 3.1; ThermoFinnigan) with the following parameters: peptide molecular weight range of 300–3500, threshold of 100,000, precursor mass tolerance of 1.4, and minimum ion count of 35. We conducted MS/MS database searches using Mascot (licensed, version 2.2, Matrix Science, Boston, MA, USA) against the National Center for Biotechnology Information non-redundant (NCBI) database in the taxon Metazoa with a parent tolerance of 1.20 Da, fragment tolerance of 0.60 Da, and two missed trypsin cleavages allowed. We
specified carbamidomethylation of cysteine and oxidation of methionine in Mascot as fixed and variable modifications, respectively.

2.3.3. MALDI ToF MS and MALDI ToF/ToF MS/MS
RP-HPLC fractions were submitted to the Institute for Integrated Research in Materials, Environments and Society at California State University, Long Beach, to determine whole protein molecular masses and protein identification/similarity. For MALDI ToF/ToF MS/MS analysis, tryptic peptides were mixed with α-cyano-4-hydroxy cinnamic acid (CHCA) matrix and directly spotted onto MALDI plates. MS spectra were collected using 1000 laser shots/spectrum, and MS/MS spectra from 3000 shots/spectrum. Peptides with signal-to-noise ratio above 15 in MS mode were selected for MS/MS analysis, with a maximum of 15 MS/MS spectra allowed per spot. Internal calibration was achieved using ToF/ToF Calibration Mixture (AB SCIEX). We searched MS/MS data against the NCBInr database within Metazoa using GPS Explorer, running Mascot (version 2.1) search engine with a peptide tolerance of 300 ppm, MS/MS tolerance of 0.8 Da, and one missed cleavage allowed. We specified carbamidomethylation of cysteine as a fixed modification, and the following as variable modifications: carbamyl, Gln/pyro-Glu (N-term Q), and Glu/pyro-Glu (N-term E). Mass spectrometry data for the peaks in Supplementary File 1 is presented in Supplementary Spreadsheet 1.

2.3.4. Statistical analyses
To confirm that population differences existed among the 11 snakes with the quantitative RP-HPLC data presented in Supplementary Spreadsheet 2, we subjected the percent protein present in each of the 11 toxin families (area under the peaks) to a 4 × 11 (population × toxin family) analysis of variance (ANOVA [96]), treating population as a between-subjects factor and toxin family as a within-subjects factor. We rank-transformed the data to avoid analysis of percentage data that summed to 100 for
each individual. Although our samples were small and data were somewhat non-normal and heteroscedastic, general linear models generally handle data well that fail to meet parametric assumptions and the results were extremely robust. We also ran a non-parametric Kruskal–Wallis ANOVA for each toxin family to compare the populations, which allowed us to confirm the results from the parametric ANOVA; this latter test requires no assumptions about data distribution [96]. We computed effect

3. Results and discussion

Random sequencing recovered sequences for 13 different venom protein encoding gene families (Table 1), with all but Kunitz and Hyaluronidase recovered by both proteomics and transcriptomics. The inability of our combined approach to detect these two venom-components in both result sets may be due to a number of factors, such as, i) differential transcription/translation: not all toxins being replenished at equal stoichiometric rates or simultaneously; ii) technical limitation: the relative separation ability of the HPLC column utilized; iii) co-elution of toxins: one toxin type dominating another and thus obscuring the signal of a toxin present in significantly lower amounts; iv) transcriptomics: the non-exhaustive random sampling procedure utilized which would statistically be likely to recover the most abundant toxin types, with lower-level expressed toxins not recovered; and/or v) microRNA silencing: whereby toxin coding regions undergo transcription but not translation [100]. Lectin toxins, however, were conspicuously absent in both the proteomics and transcriptomics of the Idyllwild population. Sequences analyzed in this study have the GenBank accession numbers of: C. o. helleri (Catalina Island) GAKQ01000001–GAKQ01000026; C. o. helleri (Idyllwild) GAKR01000001–GAKR01000018; C. o. helleri (Loma Linda) GALC01000001–GALC01000026; and C. o. helleri (Phelan) GAKS01000001–GAKS01000031. It must be noted that in accordance with the new GenBank deposition rules to exclude fragments of less than 200 base pairs, only the full length sequences were deposited. Thus 27 β-defensins were not deposited, even though their processed and secreted toxin regions were sequenced (only regions of the signal peptide were incomplete). Thus, while these sequences could not be deposited into GenBank, they were utilized in the analyses and are included in the Supplementary material.

Our proteomics analyses revealed significant differences in the venoms of the four populations (Fig. 2), with venom RP-HPLC profiles within a population largely congruent among individuals (Supplementary Fig. 8; note: only two Loma Linda specimens were able to be analyzed due to the rarity of C. o. helleri in this locality). The parametric ANOVA yielded a highly significant interaction between population and toxin family (F_{3,8,22.9} = 13.15, P < 0.001, adjusted partial η^2 = 0.31; Greenhouse–Geisser adjustment of degrees-of-freedom applied), indicating that the distribution of toxins among the toxin families differed significantly among the populations. The Kruskal–Wallis ANOVAs confirmed that toxin quantity varied significantly among populations for some (nerve growth factor, cysteine-rich secretory protein [CRiSP], lectin; all P = 0.21–0.35, η^2 = 0.86–0.97) but not all toxin families. Five additional toxins (BPP, β-defensin, kallikrein, PLA2, SVMP) approached significance (P < 0.10) with exceptional effect sizes (η^2 > 0.63). Thus, the ANOVAs confirmed population differences despite the small sample sizes.

Some toxin types were notable for being either highly conserved in their coding sequences (β-defensin, natriuretic), whereas others were extremely variable (kallikrein, lectin, PLA2, SVMP). While the β-defensins and bradykinin potentiating peptides (BFPs) were of low complexity, our proteomics analyses of the relative expression levels revealed that they are expressed...
in very high amounts in all populations, with \( \beta \)-defensin in particular invariantly expressed in large quantities (Fig. 2). The multi-product natriuretic/BPP precursor was invariant within and between populations in both the plesiotypic natriuretic peptide domain and the apotypic (derived) BPP domains located within the propeptide region. In contrast, the lectin sequences were highly variable, including the apotyposis of novel cysteines which may facilitate novel structural folding or unique subunit formation with lectins or other toxin types (Fig. 3). Consistent with the proteomic results of this study and a previously published study of San Jacinto Mountain specimens [49] as well as observed notable clinical effects, only the Idyllwild population contained both the acidic and basic subunits of the neurotoxic PLA\(_2\) complex type, with both chains virtually identical to the well-characterized potent presynaptic neurotoxins from \( C. d. terrificus \) and \( C. s. scutulatus \) (Fig. 4). It was also notable that the Idyllwild population secreted the lowest amount of SVMPs (Fig. 2), with only a single isoform obtained in the transcriptome and only detectable in trace levels in the proteome. In contrast, the other populations secreted SVMPs in large amounts, with the Phelan having the greatest complexity while the Catalina Island population had less complexity but a much higher relative expression level. This is consistent with the pattern observed for \( C. s. scutulatus \), that there is an inverse relationship between the relative amount of neurotoxic PLA\(_2\) and hemorrhagic SVMP [37,101,102]. Thus, it is quite evident how a biochemical arsenal with such variability in neurotoxic, hemotoxic and myotoxic venom-components can complicate clinical treatment of bite victims, not only through the production of highly variable clinical effects, but also as a consequence the reciprocal variability in the efficacy of anti-venom binding. It should be noted that the venom proteomics of

Fig. 5 – Molecular evolution of \( C. o. helleri \) \( \beta \)-defensins. Three-dimensional homology models (built using the PDB template 1Z99) of \( \beta \)-defensins with evolutionary conservation of amino acids mapped onto them, depicting the locations of positively selected sites (in red) detected by site-model 8 (PP \( \geq \) 0.95, BEB). Schematic representation of the models, which not only depicts the locations of positively selected sites (red sticks) but also highlights disulfide bonds (orange sticks), \( \alpha \) helices (purple) and \( \beta \) sheets (green), are also presented.
multiple animals (n = 3; except Loma Linda population, where these animals are extremely rare) from the same region were fairly similar. Hence, it can be safely assumed that the venom-gland transcriptomics of randomly chosen animals represents the overall venomics (genetic makeup of the venom gland) of the representative population.

Understanding the nature and strength of natural selection pressures, which sculpt genetic diversity, is the central theme of molecular evolutionary studies. Since non-synonymous mutations are more likely to influence the structure and function of a protein and hence in turn influence the fitness of the organism, evaluating the rate of accumulation of non-synonymous mutations (dN) in genes, relative to synonymous mutations (dS), as a ratio known as $\omega$ (or dN/dS ratio), is essential. We assessed the role of evolutionary selection pressures in shaping various venom proteins in different populations of C. o. helleri using various state-of-art selection assessment methodologies. We detected a significant influence of positive Darwinian selection on the evolution of most venom protein encoding genes in these snakes (Figs. 5–7; Tables 2–4; Supplementary Tables 2–5; Supplementary Figs. 1–7 and 9–11).

Site-specific selection assessments indicated that $\beta$-defensins, which were expressed in relatively large amounts by all C. o. helleri populations examined, followed a regime of weak positive selection: Catalina Island: $\omega = 1.33$ and 3 positively selected (PS); Idyllwild: $\omega = 2.07$ and 3 PS; Loma Linda: $\omega = 1.14$ and 2 PS; Phelan: $\omega = 1.31$ and 5 PS; All: $\omega = 1.18$ and 11 PS (Fig. 5; Table 2). However, the mapping of mutations onto sequence alignments indicated that most hypermutable sites detected by site-specific methods in $\beta$-defensins were concentrated in the non-secreted regions of the toxin that are not likely to contribute in the envenoming process. It was also evident

Fig. 6 – Molecular evolution of C. o. helleri kallikreins. Three-dimensional homology models (Loma Linda population modeled using the PDB template 1OP0; all others using 2AIQ) of kallikreins with evolutionary conservation of amino acids mapped onto them, depicting the locations of positively selected sites (in red) detected by site-model 8 (PP $\geq 0.95$, BEB) are presented. Schematic representation of the models, which not only depicts the locations of positively selected sites (red sticks) but also highlights disulfide bonds (orange sticks), $\alpha$ helices (purple) and $\beta$ sheets (green), are also presented.
that the entire stretch of nucleotides encoding the secreted region of β-defensins evolved under the extreme influence of negative selection, with 76% of residues being extremely well conserved (percent identity ≥ 90%; Supplementary Fig. 1). This was also supported by the results of MEME, an extremely accurate method of detecting episodic bursts of adaptation, which detected fewer episodically diversifying sites in β-defensins (Table 2). Mapping of variable sites on the structure of the β-defensin ‘crotamine’ from C. d. terrificus (PDB code: 1Z99 [103]), which is homologous and thus structurally very similar
to β-defensins, revealed that the N-terminal positions 23 (Y in sequence 1; Supplementary Fig. 1) and 25 (R in sequence 1; Supplementary Fig. 1) were also solvent exposed, they were located inside the disulfide bridge-stabilized protein core and not resolved). This can be explained by the fact that the highly conserved positions, namely, at position 61 (K in sequence 1; Supplementary Fig. 1), 62 (S in Supplementary Fig. 1) as well as the C-terminal residues in sequence 1; Supplementary Fig. 1) and 25 (R in sequence 1; Supplementary Fig. 1) were extremely well conserved. Not-surprisingly, 29% of the residues in C. o. helleri were cationic (K, R and H) and were extremely well conserved (percent identity ≥ 80%; Supplementary Fig. 1). Hence, it is expected that the evolutionary constraints favor the preservation of cationic residues required for toxicity. The branch-site REL (BSR) test, which significantly identifies lineages that follow the regime of episodic diversification, clearly highlighted the differences in strengths of evolutionary selection pressures acting upon β-defensins in C. o. helleri populations (Supplementary Fig. 5). In the Phelan population this test detected as many as four episodically diversifying branches in β-defensin gene lineage, while detecting only one branch each in Catalina Island and Loma Linda populations, and two branches in the Idyllwild population (Supplementary Fig. 5).

| Table 2: C. o. helleri intraspecific venom dynamics: β-defensins. |

<table>
<thead>
<tr>
<th>Population</th>
<th>FUBAR a</th>
<th>MEME sites b</th>
<th>BSR c</th>
<th>PAML d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M8</td>
<td>M2a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CI</td>
<td>≥ 1; 0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(2 + 1)</td>
<td>(2 + 1)</td>
<td>1.33</td>
<td>1.33</td>
</tr>
<tr>
<td>ID</td>
<td>≥ 1; 0</td>
<td>1</td>
<td>2</td>
<td>2.07</td>
</tr>
<tr>
<td></td>
<td>(0 + 2)</td>
<td></td>
<td>1.14</td>
<td>1.14</td>
</tr>
<tr>
<td>LL</td>
<td>≥ 1; 0</td>
<td>1</td>
<td>2</td>
<td>2.07</td>
</tr>
<tr>
<td></td>
<td>(0 + 2)</td>
<td></td>
<td>1.14</td>
<td>1.14</td>
</tr>
<tr>
<td>PH</td>
<td>≥ 1; 0</td>
<td>4</td>
<td>6</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td>(5 + 6)</td>
<td></td>
<td>1.18</td>
<td>1.16</td>
</tr>
<tr>
<td>Combined</td>
<td>≥ 1; 0</td>
<td>4</td>
<td>6</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td>(5 + 6)</td>
<td></td>
<td>1.18</td>
<td>1.16</td>
</tr>
</tbody>
</table>

a: mean dN/dS.

Populations: CI = Catalina Island; ID = Idyllwild; LL = Loma Linda; PH = Phelan.

b: Sites detected as experiencing episodic diversifying selection (0.05 significance) by the Mixed Effects Model Evolution (MEME).

c: Number of branches detected by the branch-site REL (random effects likelihood) test as episodically diversifying.

Table 3: C. o. helleri intraspecific venom dynamics: Kallikrein.

<table>
<thead>
<tr>
<th>Population</th>
<th>FUBAR a</th>
<th>MEME sites b</th>
<th>BSR c</th>
<th>PAML d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M8</td>
<td>M2a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CI</td>
<td>≥ 1; 0</td>
<td>6</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>(6 + 1)</td>
<td></td>
<td>1.35</td>
<td>1.31</td>
</tr>
<tr>
<td>ID</td>
<td>≥ 1; 0</td>
<td>6</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>(2 + 9)</td>
<td></td>
<td>1.63</td>
<td>1.63</td>
</tr>
<tr>
<td>LL</td>
<td>≥ 1; 0</td>
<td>9</td>
<td>9</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>(8 + 16)</td>
<td></td>
<td>1.60</td>
<td>1.57</td>
</tr>
<tr>
<td>PH</td>
<td>≥ 1; 0</td>
<td>15</td>
<td>5</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>(10 + 14)</td>
<td></td>
<td>1.38</td>
<td>1.41</td>
</tr>
<tr>
<td>Combined</td>
<td>≥ 1; 0</td>
<td>45</td>
<td>12</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>(18 + 18)</td>
<td></td>
<td>1.36</td>
<td>1.38</td>
</tr>
</tbody>
</table>

a: mean dN/dS.

Populations: CI = Catalina Island; ID = Idyllwild; LL = Loma Linda; PH = Phelan.

b: Sites detected as experiencing episodic diversifying selection (0.05 significance) by the Mixed Effects Model Evolution (MEME).

c: Number of branches detected by the branch-site REL (random effects likelihood) test as episodically diversifying.

d: Number of sites under pervasive purifying selection at the posterior probability > 0.9 (FUBAR).

While the kallikreins found in each of the C. o. helleri populations examined were found to be rapidly evolving under the influence of positive selection [Catalina Island: ω = 1.35 and 7 PS; Idyllwild: ω = 1.63 and 11 PS; Loma Linda: ω = 1.60 and 24 PS; Phelan: ω = 1.38 and 24 PS; All: ω = 1.36 and 36 PS] (Fig. 6; Supplementary Fig. 2), the number of positively selected sites detected by M8’s Bayes empirical Bayes (REB) approach, varied from 7 to 24, highlighting the differential rate of evolution of kallikreins in various C. o. helleri populations (Table 3). The number of branches detected by the BSR test as episodically diversifying in kallikrein encoding genes varied from 4 to 10 in various populations (Supplementary Fig. 6), again highlighting the differential role of selection in shaping these venom protein encoding genes.
Table 4 — C. o. helleri intraspecific venom dynamics: Lectins.

<table>
<thead>
<tr>
<th>Population</th>
<th>FUBAR a</th>
<th>MEME sites b</th>
<th>BSR c</th>
<th>PAML d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α chain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CI</td>
<td>ω &gt; 1 : 12</td>
<td>2</td>
<td>4</td>
<td>M8</td>
</tr>
<tr>
<td></td>
<td>ω &lt; 1 : 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LL</td>
<td>ω &gt; 1 : 15</td>
<td>1</td>
<td>3</td>
<td>M2a</td>
</tr>
<tr>
<td></td>
<td>ω &lt; 1 : 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PH</td>
<td>ω &gt; 1 : 15</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ω &lt; 1 : 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>ω &gt; 1 : 26</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ω &lt; 1 : 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β chain</td>
<td>ω &gt; 1 : 0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>CI</td>
<td>ω &lt; 1 : 1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>LL</td>
<td>ω &gt; 1 : 2</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ω &lt; 1 : 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PH</td>
<td>ω &gt; 1 : 16</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ω &lt; 1 : 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>ω &gt; 1 : 22</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ω &lt; 1 : 2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ω: mean dNdS.
Populations: CI = Catalina Island; ID = Idyllwild; LL = Loma Linda; PH = Phelan.

a Fast Unconstrained Bayesian AppRoximation.

b Sites detected as experiencing episodic diversifying selection (0.05 significance) by the Mixed Effects Model Evolution (MEME).

c Number of branches detected by the branch-site REL (random effects likelihood) test as episodically diversifying.

d Positively selected sites detected by the Bayes Empirical Bayes approach implemented in M8 and M2a. Sites detected at 0.99 and 0.95 significance are indicated in the parenthesis.

Table 4 — C. o. helleri intraspecific venom dynamics: Lectins.

Influence of positive selection (Table 4; Supplementary Fig. 3). However, β-chain of the Catalina Island population was remarkably revealed to have evolved under the influence of negative selection (ω = 0.46, 0 PS; Table 4; Supplementary Fig. 4). Other than at position 56, amino acid residues in all other positions were invariant (Supplementary Fig. 4). The rapid rate of molecular evolution observed in lectins is consistent with the great diversity of novel sequences recovered, including the apyoposis or the derivation of novel cysteine residues (Fig. 3). The rapid accumulation of hypermutable sites under the influence of positive selection in β-chain lectins from all C. o. helleri populations except those from Catalina Island, where the toxin-encoding gene has evolved under strong negative selection, is intriguing and warrants further experimental evaluations to understand the stark differences in the magnitude of selection pressures. While the BSR test detected a few lineages as episodically diversifying in the α-chain lectins of various populations, the results of this test in the β-chain lectins were particularly interesting (Supplementary Fig. 7). This test failed to detect any branch in the Catalina Island population, while detecting a single branch in Loma Linda population as episodically diversifying (Supplementary Fig. 7). In contrast, as many as 6 branches were detected as following the regime of episodic adaptation in β-chain lectins of the Phelan population (Supplementary Fig. 7). Similar to the results of all state-of-art selection assessment methods outlined above, the evolutionary fingerprints of venom-encoding genes in C. o. helleri clearly depicted the differential influence of natural selection on their evolution (Supplementary Figs. 9–11).

The structure and surface chemistry of the presynaptic PLA2 complex from C. o. helleri is very well conserved when compared to the homologues from C. d. terrificus and C. s. scutulatus (Fig. 4). Both amino acid type distribution on the protein surface as well as studying surface charges and surface hydrophobicity of all three PLA2 complexes revealed only minor differences. While the positive and negative charged patches in globo were located in the same positions, minimal differences were observed in the size and charge of these surface regions. Since the PLA2s of C. d. terrificus and C. s. scutulatus are well-characterized to be potent neurotoxins (cf.[89,105]), we conclude that the described similarities of C. o. helleri PLA2 to the former ones are responsible for neurotoxic effects of PLA2s observed in the C. o. helleri population. The precise evolutionary regimes followed by genes encoding PLA2 and SVMPs in these snakes remain to be elucidated.

Thus, it is evident that C. o. helleri venom-encoding genes have experienced differential evolutionary selection pressures. Differential rate of molecular evolution or expression occurred not only between toxin types within the venom of a particular population, but also for the same toxin type between populations. These results demonstrate that the different populations of C. o. helleri follow distinct evolutionary trajectories, with the differential venom profiles likely driven by variation in predatory ecology. This is a reflection of the complex evolutionary history of this species, which ranges from sea level to high mountain peaks and occupies a diverse range of habitats. These habitats possess differing lizard and mammal prey assemblages [106–108], and evidence from other snakes suggests that strong natural selection has driven venom adaptation to particular prey species [12,20,21,30,31,40,46,47,59,61–63,109]. Although climate might be expected to influence venom composition, our data suggest otherwise concerning the dichotomy of type I (proteolytic or “tenderizer”) versus type II (more toxic) venoms [70]. It has been suggested that snakes at higher elevation with the greatest temperature fluctuations could be expected to possess a type I venom to facilitate digestion [70]. However, the population that faces the highest temperature fluctuations (Idyllwild) possesses a type II venom that lacks almost entirely the metalloproteases typical of type I venoms. These results also indicate significant differences in potential human envenomation profiles, consistent with the complex clinical picture previously observed, with some populations being hemorrhagic while others are neurotoxic. The exquisite diversity of venom components highlighted in this study and the variation in intensity and the nature of natural selection shaping the molecular toxin scaffolds may not only result in distinct envenoming profiles but may also induce variable responses to antivenom. Hence, understanding the true molecular diversity of venom and the evolutionary forces that shape them not only
aids in the prediction of clinical effects but also reveals that these highly variable venoms are a rich source of novel toxins, some of which may have significant potential for use as lead compounds in drug design and development. Thus, the results of this study not only contribute to the body of knowledge regarding venom evolution but also have applied outcomes both from a clinical perspective and also from drug design. These results will also be useful in science communication to demonstrate that there is indeed significant variation in the venom of this medically important species, but that such evolution has not occurred recently but rather the venom diversity seen today is reflective of the long evolutionary history, not of recent changes as popularly misunderstood. Thus this species is a model for the broader penetration of lay-person understanding of venom diversity and the clinical and economic importance of such variation.

Acknowledgments

BGF was funded by the Australian Research Council (ARC) and the University of Queensland. This study was also supported by the ARC Discovery Grant DP130103813 to GFK. EABU would like to acknowledge funding from the University of Queensland (International Postgraduate Research Scholarship, UQ Centennial Scholarship, and UQ Advantage Top-Up Scholarship) and the Norwegian State Education Loans Fund. KS was funded by a PhD grant (SFRH/BD/61959/2009) from F.C.T. (Fundação para a Ciência e a Tecnologia). AA was funded by the project PTDC/AACMB/121301/2010 (FCOMP-01-0124-FEDER-019490) from F.C.T. CC was supported by the National Science Foundation Graduate Research Fellowship under Grant No. 2012134810 and therefore must include the statement “Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.” We thank Joel Almquist, Erick and Erin Briggs, Aaron Corbit, Karin Greenwood, Heidi and Todd Hoggan, Maximus Kyung Hyun Lee, and Julie King for donating snakes or providing research assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2014.01.013.

REFERENCES


