MBE Advance Access published February 1, 2012

Research Article

Evolution of CRISPs associated with toxicoferan-reptilian venom and mammalian reproduction

Kartik Sunagar[1, 2], Warren E. Johnson[3], Stephen J. O'Brien[3, 4], Vítor Vasconcelos[1, 2] and Agostinho Antunes [1, 2, 3, §]

1. CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal

2. Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007, Porto, Portugal

3. Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD 21702-1201, USA.

4. Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg University, St. Petersburg, Russia.

[§]Corresponding author: aantunes@ciimar.up.pt

Key words: CRISP, positive selection, toxicoferan-reptilian venom evolution, adaptive evolution

Abstract

Cysteine-rich Secretory proteins (CRISPs) are glycoproteins found exclusively in vertebrates and have broad diversified functions. They are hypothesized to play important roles in mammalian reproduction and in reptilian venom, where they disrupt homeostasis of the prey through several mechanisms, including among others, blockage of cyclic nucleotide-gated and voltage-gated ion channels and inhibition of smooth muscle contraction.

We evaluated the molecular evolution of CRISPs in toxicoferan reptiles at both nucleotide and protein levels relative to their non-venomous mammalian homologues. We show that the evolution of CRISP gene in these reptiles is significantly influenced by positive selection, and in snakes (ω=3.84) more than in lizards (ω=2.33), while mammalian CRISPs were under strong negative selection (CRISP1=0.55, CRISP2=0.40 and CRISP3=0.68). The use of ancestral sequence reconstruction, mapping of mutations on the three-dimensional structure and detailed evaluation of selection pressures, suggests that the toxicoferan CRISPs underwent accelerated evolution aided by strong positive selection and directional mutagenesis; while their mammalian homologues are constrained by negative selection. Gene and protein-level selection analyses identified 41 positively selected sites in snakes and 14 sites in lizards. Most of these sites are located on the molecular surface (nearly 76% in snakes and 79% in lizards) while the backbone of the protein retains a highly conserved structural scaffold. Nearly 46% of the positively selected sites occur in the cysteine-rich domain of the protein. This directional mutagenesis, where the hotspots of mutations are found on the molecular surface and functional domains of the protein, acts as a diversifying mechanism for the exquisite biological targeting of CRISPs in toxicoferan reptiles. Finally, our analyses suggest that the evolution of toxicoferan-CRISP venoms might have been influenced by the specific predatory mechanism employed by the

organism. CRISPs in Elapidae, which mostly employ neurotoxins have experienced less positive selection pressure (ω=2.86) compared with the 'non-venomous' colubrids (ω=4.10) that rely on grip and constriction to capture the prey, and the Viperidae, a lineage that mostly employs haemotoxins (ω=4.19). Relatively lower omega estimates in Anguimorph lizards (ω=2.33) than snakes (ω=3.84) suggests that lizards probably depend more on pace and powerful jaws for predation than venom.

Introduction

Each year, thousands of people die worldwide from envenomation by various species; despite the fact that many of these animals have sophisticated warning behavior to avoid close encounters and that the resultant bites are mostly accidental or defensive in nature. Frequent bites from these animals in both developed and developing countries accentuates the importance of venom research. Intraspecific variation in venom composition, is likely influenced by multiple factors, including their phylogenetic history, diet, predator pressure, etc. (Sasa 1999; Barlow et al. 2009), can be large (Daltry, Wuster, Thorpe 1996; Tsai et al. 2007), which complicates the production of antivenom. Therefore, the assessment of diversifying mechanisms and selection pressures influencing the evolution of venom-encoding genes can potentially provide valuable information for structure-based drug design and the production of life-saving antidotes of greater specificity.

Venomous predators often employ a concoction of polypeptides and other molecules with diverse biological activities in their venom to attack multiple homeostatic systems within the prey in very specific and targeted manners. Venom components can target numerous physiological pathways, tissues and cell types that are accessible via blood and lymphatic systems or by direct injection into the musculature. Evolution of the venom arsenal in snakes has been studied considerably, with research recently focusing largely on venom-coding genes including phospholipase A2, snake venom disintegrins, snake venom metaloproteases (SVMP) and lectins. (Kini, Chan 1999; Calvete et al. 2003; Fry 2005; Morita 2005; Lynch 2007; Soto et al. 2007; Peichoto et al. 2010). However, there is a lack of understanding of the evolution of cysteine-rich secretory proteins (CRISPs) in toxicoferan reptiles, a hypothetical venomous clade comprising all the venomous and related non-venomous reptile species of the suborder *Serpentes* and *Iguania* (Snakes and

Lizards: *Anguidae*, *Varanidae* and *Helodermatidae*). CRISPs belong to a large family of proteins found extensively in vertebrates and they participate in diverse biological processes (Kitajima, Sato 1999; Udby et al. 2002; Gibbs et al. 2007; Gibbs, O'Bryan 2007). CRISPs are particularly enriched in the pancreatic tissues, salivary glands, reproductive tracts (Kierszenbaum et al. 1981; Haendler et al. 1993; Schambony et al. 1998) and reptilian venom ducts (Hill, Mackessy 2000; Yamazaki, Hyodo, Morita 2003; Fry et al. 2006). They have diverse biological activities including inhibition of various ion-channels and the inducement of proteolysis and paralysis of prey. There are several venom CRISPs with no known acute toxic effects or functions (Chang, Mao, Guo 1997). Scope and details of mammalian and reptilian CRISPs in reproduction, immune system and venom are poorly understood and will require increased evolutionary and functional analyses.

CRISPs are single polypeptide proteins with molecular weights of ~20-30 kDa (Yamazaki, Morita 2004; Yamazaki, Morita 2007) and are known to have a high degree of amino acid sequence similarity, and a highly conserved specific pattern of 16 cysteine residues. Ten of these cysteine residues form an integral part of the highly conserved cysteine-rich domain (CRD) at the c-terminus. The CRD is composed of two domains, a hinge region and an ion-channel regulator domain (ICR). A few reptilian venom CRISPs have been shown to interact with ryanodine receptors (RyRs), perhaps through the ICR, to inhibit the release of Ca2+ ions, thus inhibiting smooth muscle fiber functions (Nobile et al. 1994; Morrissette et al. 1995). 'Pseudechetoxin' isolated from the venom of Australian king brown snake *(Pseudechis australis)* has been shown to inhibit cyclic nucleotide-gated (CNG) ion channels. Many other targets such as Potassium and Calcium channels have also been suggested (Wang et al. 2005; Wang et al. 2006) and there could be many more targets yet to be discovered.

There are three types of CRISPs in most mammals [**CRISP1**: AEG or Acidic Epididymal Glycoprotein (AEG), **CRISP2**: Testis-specific protein 1 and **CRISP3**: specific granule protein)]. A fourth type has been described only in mice (CRISP 4). It is hypothesized that the ancestral salivary gland CRISP was modified as a venom-component in the toxicoferan reptiles rather than the usual gene-recruitment events that have led to the multiple copies of other venom proteins like PLA2, disintegrins, metaloproteases defensing etc. (Fry 2005).

In this study, we assessed for the first time the evolutionary history and selection pressures influencing the toxicoferan CRISPs using both gene- and protein-level approaches. We tested the hypothesis of whether reptilian CRISP variation has accumulated under the regime of positive selection and if their mammalian homologs were constrained because of their importance in other functions. Finally, we investigate the implications of observed mutations in the three-dimensional structure of the snake and lizard CRISP proteins to obtain further insight into the evolution of reptilian toxicoferan venom.

Materials and Methods

Sequence Retrieval and Alignment

To assess the molecular evolution of cysteine-rich secretory proteins, we compiled a dataset of 119 nucleotide sequences (46 Snakes + 21 Lizards + 52 Mammals). Nucleotide and protein sequences were downloaded from National Center for Biotechnology Information (NCBI: http://www.ncbi.nlm.nih.gov/) and UniProt databases (http://www.uniprot.org/) respectively. The sequences (identified by their GenBank™ accession numbers) were retrieved by using BLAST. Complete cDNA sequences from each lineage of toxicoferan-reptiles (Elapidae*:* AY299475.1; Viperidae*:* AY181983.1; Colubridae*:* DQ139891.1; Anguimorph lizards: EU790958.1*)* were used to retrieve the reptilian CRISPs while one sequence from each type of mammalian CRISPs (CRISP1: GU985267.1; CRISP2: BT030687.1; CRISP3: BC102058.1) was used for the retrieval of the mammalian sequences.

The translated nucleotide sequences were aligned using MUSCLE 3.8 (Edgar 2004). The alignments were manually inspected and edited by eye. We used the 16 universally conserved cysteine residues as anchors to refine the alignment. Gblocks (Talavera, Castresana 2007) was used to remove regions that had gaps in more than 50% of the sequences in the alignment.

Phylogenetic Analyses

The best-fit model of nucleotide substitution for our dataset was determined as TIM3+G+I by jModeltest (Posada 2008), according to Akaike's information criterion (AIC). Model averaged parameter estimates of gamma shape parameter (alpha) and the proportion of invariant sites (pinvar) were used for phylogenetic reconstruction. The phylogenetic

relationships among the toxicoferan-reptilian CRISPs were determined using Bayesian and maximum-likelihood approaches. MrBayes version 3.1 (Huelsenbeck, Ronquist 2001; Ronquist, Huelsenbeck 2003) was used for Bayesian inference. Tree searches were run using four Markov chains for a minimum of 10 million generations, sampling every 100th tree. The log likelihood score of each saved tree was plotted against the number of generations to establish the point at which the log-likelihood scores of the analyses reached their asymptote. 25% of the total trees sampled were discarded as burnin. The posterior probabilities for clades were established by constructing a majority rule consensus tree for all trees generated after the completion of the burnin. The analyses were repeated three times to make sure that the trees generated were not clustered around local optima. An optimal maximum likelihood phylogenetic tree was obtained using PhyML 3.0 (Guindon et al. 2010) and node support was evaluated with 1,000 bootstrapping replicates. Phylogenetic trees were rooted using the mouse (NM_009639.2) and the human (X94323.1) CRISP3 sequences. Dendroscope (Huson et al. 2007) was used to prepare the phylogenetic tree.

Sequence divergence (F84 genetic distance) was plotted against the transition (s) and transversion (v) rates and a test of nucleotide substitution was carried out using DAMBE (Xia, Xie 2001; Xia et al. 2003) to evaluate the influence of saturation of nucleotide substitutions on the evolutionary inferences. Recombination analyses were done using Single Breakpoint Recombination and Genetic Algorithms for Recombination Detection (GARD) implemented in the Datamonkey server (Pond, Frost, Muse 2005; Kosakovsky Pond et al. 2006; Delport et al. 2010).

Selection Analyses

Maximum likelihood models of coding-sequence evolution implemented in CODEML in the PAML (Yang 2007) package of programs version 4 were used to test the hypothesis that functional diversification of snake and lizard venom CRISP genes is driven by positive Darwinian selection. PAML compares the maximum likelihood estimates of dN and dS across an alignment to a predefined distribution and uses empirical Bayes methods to identify individual positively selected site (Nielsen, Yang 1998; Yang, Bielawski 2000).

We evaluated the evidence for positive selection on CRISP genes in the toxicoferan reptiles and their homologous mammalian counterparts by employing branch models. The most simple one-ratio model assumes the same dN/dS for all the branches in the phylogenetic tree (Goldman, Yang 1994). The assumption of constant evolutionary selection pressure on all the lineages in the phylogenetic tree over millions of years of evolutionary time sounds quite unrealistic. Hence, the free-ratio model assumes separate dN/dS ratios for all branches in the tree. However, this model is parameter rich and is prone to inaccurate estimations (Yang et al. 2000). Both the one-ratio and the free-ratio models detect positive selection when the average of omega values over the entire length of protein is greater than one – i.e. when the majority of the amino acids are under the influence of selection. To assess selection pressures acting upon individual lineages, we employed the two-ratio model as well as the optimized branch-site test (Yang, Nielsen 2002; Zhang, Nielsen, Yang 2005). A likelihood-ratio test was conducted by comparing the two-ratio model that allows omega to be greater than 1 in the foreground branch, with the null model that does not. The branch-site model by comparison, allows omega to vary both across sites of the protein and across branches in the tree and has reasonable power and accuracy to detect short bursts of episodic adaptations (Zhang, Nielsen, Yang 2005).

Unlike the lineage specific branch and branch-site models, the GA-Branch Test implemented in the HyPhy (Pond, Frost 2005) package does not require the foreground and background branches to be defined *a priori.* The algorithm works on the principle that there could be many models that better fit the data than a single *a priori* hypothesis and uses a robust multi-model inference to collate results from all models examined and provides confidence intervals on dN/dS for each branch.

Another draw-back of most lineage specific models is that they assume a single omega value for the entire length of the sequence. They lack the ability to identify sites in proteins that might be under the influence of positive selection more than others, such as the surface of venom molecules. Thus lineage specific models can underestimate the degree of positive selection acting on biological sequences. To account for rate variation among sites we used the site-specific models (Nielsen, Yang 1998; Yang 2000), which are powerful tools for detecting diversifying selection. Positive selection is detected statistically as a non-synonymous-to-synonymous nucleotide-substitution rate ratio **(ω)** significantly greater than 1. Because no a priori expectation exists for the distribution of ω values, we compared likelihood values for three pairs of models with different assumed ω distributions: M0 (constant ω rates across all sites) versus M3 (allows the ω to vary across sites within 'n' discrete categories, n≥3); M1a (a model of neutral evolution) where all sites are assumed to be either under negative (ω <1) or neutral selection (ω = 1) versus M2a (a model of positive selection) which in addition to the site classes mentioned for M1a, assumes a third category of sites; sites with ω >1 (positive selection) and M7 (Beta) versus M8 (Beta and ω), and models that mirror the evolutionary constraints of M1 and M2 but assume that ω values are drawn from a beta distribution (Nielsen, Yang 1998). Only if the alternative models (M3,

M2a and M8: allow sites with ω >1) show a better fit in Likelihood Ratio Test (LRT) relative to their null models (M0, M1a and M8: do not show allow sites ω >1), are their results considered significant. LRT is estimated as twice the difference in maximum likelihood values between nested models and compared with the χ^2 distribution with the appropriate degree of freedom - the difference in the number of parameters between the two models. The Bayes empirical Bayes (BEB) approach (Yang, Wong, Nielsen 2005) was used to identify amino acids under positive selection by calculating the posterior probabilities that a particular amino acid belongs to a given selection class (neutral, conserved or highly variable). Sites with greater posterior probability (PP ≥ 95%) of belonging to the 'ω > 1 class' were inferred to be positively selected.

Some studies have suggested that the maximum-likelihood method of evaluating positive selection produces false positive results even when no positively selected sites exist (Suzuki, Nei 2004) or when positively selected sites and negatively selected sites are mixed (Anisimova, Bielawski, Yang 2002). Further support for the PAML results was obtained using a complementary protein level approach implemented in TreeSAAP (Woolley et al. 2003). Models of reptilian and mammalian cysteine-rich proteins depicting the overall conservation of amino-acids were built using the Consurf webserver (Ashkenazy et al. 2010).

Detection of positive Darwinian selection across lineages using branch and branch-site models requires the foreground branches (lineages tested to be under positive selection) and background branches (rest of the lineages) to be defined *a priori*. When a predefined biological hypothesis is unavailable or the functions of genes are not well understood, then it becomes difficult to define foreground branches. A possible approach then would be to

treat each branch in the phylogeny alternately as the foreground lineage and test multiple hypotheses. A likelihood-ratio test can then conducted by comparing a model that allows omega to be greater than 1 in the foreground branch, with the same model where foreground branches are constrained to have an omega equivalent to 1. It is suggested that when using the branch-site model to test multiple branches in the phylogenetic tree for positive selection, it is necessary to control the family-wise error rate (FWER or Type I error) (Zhang, Nielsen, Yang 2005). Bonferroni's correction is the easiest method to achieve this, which uses α/n as the significance level to test each hypothesis; where 'α' is the significance level and 'n' being the number of independent true null hypotheses.
If the lineages have different proportion of sites under selection, then comparing their omega estimates to assess the strength of selection could be misleading. This can be overcome by estimating the omega simultaneously for the datasets being compared. We employed the clade model analyses (Bielawski, Yang 2004) to facilitate an effective comparison between different lineages of reptiles. These models are based on the branch-site models allowing variations in the omega value among the sites with a proportion of sites evolving under divergent selection between the clades.

To clearly depict the proportion of sites under selection, an evolutionary fingerprint analysis was carried out using the ESD algorithm implemented in datamonkey (Pond et al. 2010).

Ancestral Sequence Analyses

To understand the evolutionary pathway that shaped the extant cysteine-rich secretory proteins over millions of years, we reconstructed ancestral sequences using the ancestral sequence reconstruction (ASR) algorithm implemented on the Datamonkey server (Delport

et al. 2010) and used a maximum-likelihood based joint reconstruction approach (Pupko et al. 2000). We evaluated selection pressures acting on the ancestral toxicoferan-reptilian CRISPs and mapped the sites under positive selection on their three-dimensional crystal structure.

Functional Divergence

We used the maximum-likelihood method implemented in Diverge 2.0 (Gu 1999) to test if there was any significant change in evolutionary rates at amino-acid sites of the cysteine-rich secretory proteins after the reptilian lineages diverged. The method estimates the expected substitutions at each amino-acid site and then calculates the coefficient of functional divergence (ɵ) which is the probability that evolutionary rate at a site is statistically independent between the two gene-clusters. Such sites are referred to as Type I sites if they are conserved in one subfamily but vary greatly in another, implying that they have been subjected to different functional constraints and hence might have different functions. The co-efficient of functional divergence (ɵ) and the posterior probability for functional divergence were calculated for each position in the alignment.

Structural analyses

To depict the underlying effects of the selection pressures on reptilian and mammalian CRISPs, we mapped the sites under positive selection on their crystal structures. We used the Swiss-model server (Arnold et al. 2006) to search for the homologous sequence with the highest identity, whose three-dimensional structure was already deduced empirically by X-ray crystallography or NMR. The crystal structures of 2gizA (Natrin: *Naja atra*) and 2ddbB (Pseudecin: *Pseudechis porphyriacus*) were obtained from the Swiss-model server as the best-fit template for the target CRISP sequences from *Ophopphagus hannah*: AY299475.1 (~86% identity) and *Varanus tristis:* GU441468.1 (~50% identity)

respectively while 1wvrA (Triflin: *Trimeresurus flavoviridis*) was determined as the best fit template for the mammalian CRISPs (~86% identity). Sequences of all these structures were used in the analyses. Pymol (DeLano 2002) was used to produce the images of the three-dimensional models of CRISPs. The program GETAREA (Fraczkiewicz, Braun 1998) was used to calculate the Accessible Surface Area (ASA) / solvent exposure of amino-acid side chains. It uses the atom co-ordinates of the PDB file and indicates if a residue is buried or exposed to the surrounding medium by comparing the ratio between side-chain Accessible Surface Area (ASA) and the "random coil" values per residue. An amino-acid is considered to be buried if it has an ASA less than 20% and exposed if ASA is more than or equal to 50%.

Results and Discussion

Lizards are generally not associated with venom, not because they lack venom but because most lizards lack a specialized venom-delivery apparatus to inject venom into the prey for maximum effect. Until recently, only two species of lizards (genus *Heloderma)* with venom potent enough to kill humans were considered venomous. Despite possessing an arsenal of deadly venom components, *most* colubrids cannot efficiently deliver venom in sufficient quantities, and thus do not pose a serious medical threat to humans. However, it is now recognized that even the Anguimorph lizards and all colubrids possess venom-encoding genes (Fry et al. 2006).

Many venom-components are hypothesized to evolve via the birth and death model of evolution, where new genes are created by repeated duplication events, and are subsequently either maintained in the genome or deleted or become non-functional by pseudogenization (Fry et al. 2003a; Lynch 2007). Directional selection reinforces the functionally important toxin types through adaptive evolution, creating a venom-gland-specific multigene family (Fry et al. 2008). The likelihood of neofunctionalisation is increased through random mutation, gene conversion and unequal crossing-over (Fry et al. 2003a). The inevitable arms race between the predator and prey leads to precise biological targeting as a single amino acid change could alter the specificity or potency of the toxin. Toxicoferan-reptilian CRISPs are unique in that they are derived through the evolutionary modification of an ancestral CRISP present in the salivary gland rather than gene-recruitment events (Fry 2005). Mammals, in contrast, have three types of CRISPs (and a fourth type in mice) that derive from mammalian lineage specific gene duplication events.

To assess how these venom proteins evolved in the reptilian lineage in comparison with their non-venomous mammalian homologues, we employed both nucleotide and amino-

acid-level selection analyses. We conducted the site, branch, branch-site and clade model tests to identify sites under positive selection and to assess its influence on the three-dimensional structure and function of the CRISP protein.

CRISP Phylogeny

Bayesian and maximum-likelihood analyses of toxicoferan-reptilian CRISPs retrieved phylogenetic trees with similar topologies (Figure 1). The overall topology of these phylogenetic trees is not in concordance with the generally accepted reptilian phylogeny (Kelly, Barker, Villet 2003). Viperidae is the most basal clade of Colubroids or the advanced snakes followed, by Elapidae and Colubridae. However, the phylogenetic trees incorrectly depict Elapidae as the most basal clade, which reflects an often-observed differences between the gene and species trees (Rosenberg 2002). We note that three elapid sequences form a clade within Colubridae while other sequences from the same species group with other elapids as expected. The topology of the phylogenetic trees could be the result of ancestral polymorphisms given that CRISPs were recruited in reptiles before the divergence of the toxicoferan lineages.

Figure 1.

When studying sequences that are separated by millions of years of evolutionary time, saturation of nucleotide substitution can bias evolutionary inferences. Substitution saturation can result in an underestimation of dS and an inflation of dN/dS. We plotted the sequence divergence (F84 genetic distance) against the transitions (s) and transversions (v) rates and found no evidence for nucleotide saturation at the 3rd position or any other position. The Index Score (ISS) remained significantly lower than the critical score (ISS.C). We also evaluated the effects of recombination and found no evidence for recombination in the dataset.

Selection Analyses

To assess the consequence of positive Darwinian selection on CRISP proteins, we used likelihood models of coding-sequence evolution (Goldman, Yang 1994; Yang 1998) implemented in CODEML of the PAML (Yang 2007) package (Table 1.1 and Table 1.2).

The one-ratio model is the simplest of the branch-specific models and estimates a single ω value for all branches in the phylogeny. The estimate of global ω for the snake CRISP gene under this model is 1.14. This is an average over all codons and lineages and thus highlights the dominant role of positive selection in shaping snake venom CRISPs. Global ω estimates for lizard CRISPs was 0.78 (Table 1.2). In contrast, mammalian CRISPs exhibit significant negative selection (CRISPI: 0.42, CRISPII: 0.32 and CRISPIII: 0.50) [supplementary material].

To evaluate the selection pressures on various reptilian lineages, we employed the lineage specific two-ratio model (Table 1.1). Similar to the one-ratio model, the lineage specific two-ratio model tends to be very conservative as it can only detect positive selection if omega ratio averaged over all the sites along the lineage is significantly greater than one. Estimates of omega for snake and lizard lineages under this model are 1.21 and 0.64 respectively which also accentuates the influence of positive selection on snake CRISPs. Both these comparisons were significant compared with the strict-branch model, which is essentially the same model but with omega constrained at 1. This again proves the dominance of positive selection on snake CRISPs. We further employed the model to evaluate the selection pressures along individual snake lineages. The omega estimate for the Elapidae lineage under this model was 1.28, while for the Colubridae and Viperidae lineages the omega estimations were 1.16 and 1.05 respectively (Table 1.1). These omega-estimates were not significant when compared to the strict-branch model estimates.

Because the branch-model estimates omega for each lineage by averaging over all the branches and the site-models by averaging over all the sites, they often fail to identify episodic adaptations that affect only few amino-acids and/or lineages. Hence, we employed the branch-site model (Table 1.1) that allows omega to vary both across the lineages in the phylogenetic tree and across the sites in the genes and hence have a reasonable power and accuracy to detect short bursts of episodic adaptation targeting fewer amino-acid residues (Zhang, Nielsen, Yang 2005). The branch-site model-A identified ~25% of sites (ω=3.76) in snakes and only ~18% of the sites (ω=3.21) in lizards as under positive selection. The omega-estimates for individual snake lineages, revealed greater evidence of positive selection in Viperidae (~15% of sites; ω=5.66) than Elapidae (~16% of sites; ω=5.52) and Colubridae (~8% of sites; ω=6.73) considering both the number of positively selected sites

and the strength of positive selection (Table 1.1). We employed multiple-test corrections using Bonferroni's test to keep the family-wise error rate (FWER) less than alpha (at 0.1% significance).

The estimated ω in snakes under the clade models is 3.84 in comparison with 2.33 in lizards (LRT=295.3, DF=3, P << 0.001), again highlighting the dominant role of positive Darwinian selection on the snakes. The analyses of individual snake lineages revealed that the influence of positive selection on Viperidae (ω=4.19) is more than that of Elapidae (2.86) and Colubridae lineages (ω=4.10) [LRT=231.39; D.F=3; p<<0.001].

Table 1.1

We further employed the site specific models that account for rate variation across the sites (Table 1.2). Estimates using the Bayes Empirical Bayes (B.E.B) approach implemented in M8 suggested that up to ~26% of the residues in snakes and 16% of the residues in the lizard CRISPs are under positive selection.

In snake CRISPs, site model M2a and M8 detected 37 and 38 positively selected amino-acid residues respectively under the B.E.B approach. Model M2a detected 24 sites with PP ≥ 0.99 and 13 sites with PP ≥ 0.95 while site model M8 detected 26 sites with PP ≥ 0.99 and 12 sites with PP ≥ 0.95. In lizards however, model 2 identified only 10 amino-acid sites (2 sites with PP ≥ 0.99 and 8 sites with PP ≥ 0.95) under selection while model 8 identified 14 such sites (6 sites with PP ≥ 0.99 + 8 sites with PP ≥ 0.95). Clearly, Darwinian selection seems to be much more influential in snake than lizard CRISPs both in terms of number of amino-acids under selection and the strength of selection (Table 1.1 and Table 1.2).

Table 1.2

The omega estimation under M8 for the mammalian CRISPs was 0.55, 0.40 and 0.68 for CRISP1, CRISP2 and CRISP3 respectively. The likelihood ratio of model 8 was not significant in comparison with model 7 (Table 1.3) and none of the models of positive selection were significant in comparison to their null models (Supplementary material).

To evaluate if mammalian CRISPs evolve under the influence of negative selection we employed the Fixed-effects likelihood and Random-effects likelihood models that identify amino-acids evolving under positive and negative selection. The results suggest that there is a strong influence of negative selection on the mammalian CRISPs (Table 1.3), which may be attributed to functional constraints on these proteins, perhaps because of their hypothesized role in reproductive pathways and any genetic variation might easily affect homeostasis. Notwithstanding, a few positively selected amino-acid sites were detected in CRISP3 with codeml model 8, FEL and REL.

Table 1.3

Selection analyses at the nucleotide level alone cannot distinguish events that may contribute to the diversification of the toxin molecules. The maximum likelihood method of evaluating positive selection may sometimes retrieve false positive results (Anisimova, Bielawski, Yang 2002; Ashkenazy et al. 2010).

Hence, to provide additional support for the amino-acid sites detected to be under positive selection by PAML, we employed a complementary protein-level approach implemented

in TreeSAAP (Woolley et al. 2003) (Table 2). TreeSAAP measures the selective influences on 31 structural and biochemical amino acid properties during cladogenesis, and performs goodness-of-fit and categorical statistical tests based on ancestral sequence reconstruction. The number of radical changes in the amino-acid properties was used as a proxy for determining the strength of positive selection at a particular amino-acid position (more radical changes in amino-acid properties might indicate adaptive evolution). An empirical value of 6 amino-acid property changes was set as a threshold. Thus, the residues that had lesser than 6 amino-acid property changes were categorized as type I sites while those that had more than 6 were categorized as type II sites.

There were 41 sites detected by both PAML (Model 2 and Model 8, BEB analysis, PP $\geq$ 0.95) and TreeSAAP ($p \leq 0.001$) as under positive selection in snake CRISPs (Table 3). Ten of these sites were Type I (more than or equal to six radical changes in amino-acid properties). However, in lizards, only 11 sites were detected by both PAML and TreeSAAP to be under positive selection and of these none were Type II. This further emphasizes that snake CRISPs not only have more number of positively selected sites, but that they also experience greater selection pressures at these sites in comparison with their lizard homologs.

Table 2

Structural analyses

To investigate selection patterns and to assess their influence on the structure and function of these molecules, we mapped the sites under selection on the crystal structure of CRISPs (Figure 2.1 and 2.2).

Figure 2.1

Figure 2.2

Previously-studied snake venom proteins like PLA2, three-finger toxins, etc. have evolved via mutations targeted around their functional domains while the structural residues are constrained and maintain a highly conserved scaffold (Mackessy 2002; Kini 2004; Fry et al. 2006; Yamazaki, Morita 2007; Doley et al. 2008; Fry et al. 2008). Similarly, both snake and lizard CRISPs had a highly conserved structural scaffold while most mutations were located in the functional domains. They have 16 cysteines that form eight disulphide bridges and are universally conserved in all CRISP proteins (Guo et al. 2005). Interestingly, 10 of these cysteine residues are located in the c-terminus as part of the CRISP domain. This Cysteine Rich Domain (CRD) has structurally flanking six cysteine repeats that exhibit a high degree of similarity with the K+ channel-blocking venom from anemones (Alessandri-Haber et al. 1999; Guo et al. 2005; Shikamoto et al. 2005; Lange et al. 2006; Suzuki et al. 2008). The CRD domain in one of the mammalian CRISP has been shown to be associated with the ion channel regulatory activity (Gibbs et al. 2006). Hence, this region could mediate the interaction of CRISPs with ion-gated channels in other toxicoferan-reptiles as well. In snake CRISPs, 44% (18) of the total amino-acids (41) under positive selection were located on the CRD domain. Remarkably, out of the 38 amino-acids defining the CRD, 18 were under positive selection. However, some findings suggest that CRD alone might not be enough to mediate such interactions and other domains, particularly PR-1 might also be essential (Suzuki et al. 2008). We found 22 positively-selected amino acids within the PR-1 domain. Thus both CRD and PR-1 domains act as molecular hotspots for mutations which could facilitate the interaction of CRISPs with

various ion-gated channels. Variations in PR-1 domain could also be due to the bi-functionality of these proteins with the CRD mediating ion-channel interactions while PR-1 performs yet-to-be identified functions.

Venom components like phospholipase A2 (PLA2: belongs to a family of diversely distributed protein phospholipases) in snakes are known to accumulate numerous mutations on the outer surface of the molecule (Kini, Chan 1999). Our analysis reveals that CRISPs exhibit the same pattern (Figure 3). Of the 41 amino-acid sites identified to be under positive selection by both PAML and TreeSAAP, 31 (75.6%) were located on the outer surface of the snake CRISP molecule with an Accessible Surface Area (ASA) ratio of at least 40%, while the remaining 10 are buried with an ASA of less than 20%. In the lizard CRISPs, of the 14 amino-acids under selection, 11 (78.5%) were exposed while only 2 were buried. This is likely a general characteristic of venoms, where the molecular surface is diversified while the structurally important residues are constrained. This preserves the venom function (especially if it is an enzyme) while simultaneously diversifying the range of target cells and tissues. This enables predators to more readily adapt to the new ecological niches by being able to rapidly exploit new prey species.

Figure 3

Ancestral Sequence Analyses

We reconstructed the ancestral toxicoferan-reptilian CRISPs to assess patterns and strength of selection pressures and to detect shifts in evolutionary pressures during the course of evolutionary time (Figure 4). We further mapped the sites under positive-selection on the crystal structure of CRISPs to depict their locations and to determine the effect of selection on the structure and function of these proteins. Our analysis reveals that the ancestral

cysteine-rich secretory proteins of both snakes and lizards were influenced less by positive selection than modern forms. Many residues in the extant Viperidae CRISPs appear to show less variation compared with the ancestral forms. Thus it is evident that these proteins undergo shifts in selection pressures across different time-scales.

Figure 4

Functional Divergence

Greater magnitude of negative selection pressure at an amino-acid position can imply a functional importance of that amino-acid (Kimura 1983). Hence, the site-specific shifts in evolutionary rates of paralogous/orthologous proteins could imply functional divergence. The Gu99 likelihood ratio test was used to detect if proteins had functionally diverged in different lineages of toxicoferan reptilian and mammalian CRISPs (Table 3). The theta-ML estimation indicates the level of functional divergence between proteins in different clusters of the tree. It also calculates the posterior probability to detect the amino-acids responsible for such a divergence. Significance of these comparisons is tested by a likelihood-ratio test between the alternate hypothesis that allows ɵ to be more than zero with the null model that does not. The theta-ML estimate for the comparison between Colubridae and Elapidae (ɵ = 0.408), Viperidae and Elapidae (ɵ = 0.5048), Viperidae and lizards (ɵ = 0.3336) and Colubridae and Viperidae (ɵ = 0.3112) were all significant. These estimations suggest that there could be drastic differences in the way the cysteine-rich secretory proteins function in these lineages.

Table 3

The comparative theta-ML estimates of CRISP1 vs CRISP2 (0.50), CRISP1 vs CRISP3 (0.70) and CRISP2 vs CRISP3 (0.80) indicate shifts in function after the duplication event. CRISP1 and CRISP2 seem to be drastically different from CRISP3 proteins. CRISP1, 2 (and CRISP4 in mice) are mostly found in mammalian reproductive system while CRISP3 has a wide distribution in the body and is hypothesized to play a role in innate immune response.

Evolutionary Fingerprint analyses

Evolutionary fingerprint analysis which fits the general discrete bivariate model of site to site variation in selection and detects the number of selective classes, the dN/dS rates for each class, was conducted for reptilian and mammalian CRISPs. The intensities depicted in Figure 5A and 5B correspond to the posterior density while the width of the circle corresponds to the accuracy of estimations (compact circles represent most accurate estimations).

Figure 5

The plots reveal that the majority of amino-acids in snake CRISPs have evolved under positive selection. In contrast, in lizards most were under negative or neutral selection and only a handful were influenced by positive selection (Figure 5A). The lineage-specific plots indicate that in colubrids the majority of amino-acid sites were under positive selection while other sites are under negative selection. In Elapidae most sites were under either positive or neutral selection, while very few sites are evolving under negative selection. Although, many amino-acid sites in the Viperidae lineage seem to be under both positive

and negative selection, the algorithm could not confidently classify them into these classes (Figure 5A). Almost all the sites were under negative selection in mammalian CRISP1 and CRISP2 proteins while very few sites in CRISP3 were evolving under positive selection. A small proportion of sites in the CRISP2 protein also seemed to be under the influence of positive selection, but the results were inconclusive (Figure 5B).

Conclusions

Snake venoms appear to have evolved via a strategy of gene-recruitment (Fry, Wuster 2004), where, an existing gene is channelized to neofunctionalisation by altered gene expression, followed by the accumulation of mutations, gene duplications and the preservation of functional constraints (Todd, Orengo, Thornton 1999; Miyata, Suga 2001). This often results in a multigene family where most proteins preserve their ancestral molecular scaffold while accumulating mutations on the outside of the molecule to amplify their biological targeting.

Reptilian CRISPs became involved in the venom functional pathway through the modification of the ancestral salivary gland CRISP (Fry 2005). The detailed evaluation of selection pressures at both nucleotide and amino acid level by the employment of the site, branch, branch-site and clade models demonstrate that these modifications were significantly influenced by positive selection in all the lineages of toxicoferan-reptiles, and in snakes more than in lizards, both in the number of residues under selection and the magnitude/strength of the selection (Table 1.1 and 1.2). If the principle function of reptilian CRISPs is indeed subduing the prey by targeting various ion-channels, then an accelerated accumulation of mutations guided by positive selection would enable them to target new ion-channels and would be particularly useful for the reptiles venturing into new ecological niches. This should be distinguished from the arms race that exists between an adapting predator and prey.

Both snake and lizard CRISPs possess a highly conserved structural scaffold while most of the mutations occur on the surface and in the functional regions. Although the main function of CRISPs is poorly understood, they possibly participate in the envenomation

process by the blockage of various ion-channels such as cyclic nucleotide-gated (CNG) ion channels, Potassium and Calcium ion-channels and possible others yet to be discovered. The blockage of Ca2+ ion-channel leads to inhibition of smooth muscle contraction and hypothermia which could cause dizziness in the prey and thus further subdues the prey. This would be particularly helpful for snakes that ingest their prey whole. The paralyses of prey is vital especially for the front-fanged snakes (Elapidae, Viperidae and Atractaspididae), as they could easily injure themselves or break fangs if the prey struggles while being swallowed. These snakes possess compressor muscles that squeeze the venom out of the venom gland, into the grooves of the front fangs. Vipers are equipped with retractable fangs that can stab and retract in less than a second. Thus, the front-fanged snakes use a combination of venom-glands and front-fangs to deliver quick and deadly bites. They then wait for the victim to become completely paralyzed or die of envenomation before swallowing them whole. In contrast, rear-fanged colubrids have primitive venom-glands with no or only rudimentary compressor muscles. Hence, they have to chew on their soft skinned prey to deliver the deadly venom. Sustained compression during biting likely deforms and helps the release of venom from the gland in the absence of well-developed compressor muscles.

Accumulation of variations in CRISP proteins likely facilitates the targeting of new ion-channels in the prey, which further subdue the prey. Many elapids, rely on other powerful neurotoxins to paralyze the prey, which might explain why selection pressures on elapid CRISPs (ω=2.86) is not as high as in other lineages (Table 1.1). Colubrids rely only on grip and constriction and hence CRISPs in them, which exploit variations guided by positive selection (ω=4.10) may further assist these predators in subduing the prey. Most vipers possess a large proportion of haemotoxins as their principle venom component for the

destruction of muscles, lymphatic systems, etc. But haemotoxins are generally considered to be less potent than neurotoxins as the latter can completely paralyze the prey in minutes while haemotoxins are relatively slow-acting. Perhaps this is why Viperidae CRISPs (ω=4.19) exhibit a strong positive selection pressure in comparison to the Elapidae CRISPs (Table 1.1).

Anguimorph lizard CRISPs are the least selected genes amongst the toxicoferan reptiles (ω =2.33), perhaps because these species depend mainly on speed and jaw strength to capture the prey, rather than prey paralyses. Moreover, it has been shown that CRISPs form the major portion of colubrid and Anguimorph lizard venom (Fry et al. 2003b), further demonstrating the need for functional analyses of CRISPs in reptiles, as they likely have multiple-functions, similar to other snake venoms such as the three-finger toxins (Fry et al. 2003a).

The trends observed in our sequence analyses should also be studied in the future at population level to determine the importance of intraspecific and individual diversity. These predators do not rely on a single venom-type for killing and subduing their prey. Instead, a cocktail of different venom types is employed where one component may successfully bring down a specific type of prey while the same could be completely ineffective against the other. Moreover, the type of venom employed is not always consistent along lineages. Many vipers like the South American rattle snake *(Crotalus durissus)* employ neurotoxins (in contrast to the typical haemotoxins employed by most vipers) while some elapid venoms like that of a red-bellied black snake *(Pseudechis porphyriacus)* are chiefly composed of haemotoxins (rather than the usual elapid

neurotoxins). Some colubrids like the brown tree-snake *(Boiga irregularis)* employ neurotoxins while others like the Boomslang *(Dispholidus typus),* use deadly haemotoxins.

The presently known forms of mammalian CRISPs resulted from mammalian lineage-specific duplication events. New copies of genes are relieved of any negative selection pressures after the duplication event and begin accumulating variations. In contrast, mammalian CRISP duplicates remained under negative selection pressures (CRISP1 = 0.55, CRISP2 = 0.40 and CRISP3 = 0.68) owing to the likely vital functional roles they play in the maintenance of homeostasis. None-the-less, the branch-site model A, clade-model analyses, FEL and REL tests and the evolutionary fingerprint analyses detect a small number of sites under selection in CRISP3 proteins (Table 1.3, Figure 5B and supplementary material). CRISP3 is hypothesized to be associated with the innate immune response (Pfisterer et al. 1996; Haendler et al. 1997) and hence is likely to be benefited by such variation.

Reptilian CRISPs might derive their ability to interact with various ion-channels by the virtue of their cysteine-rich domain (CRD), where almost half the amount of mutations (~46%) of the whole protein occurs. As suggested previously (Suzuki et al. 2008), the pathogenesis-related group 1 (PR-1) domain might supplement this ability. This domain accumulates the rest of the positively selected sites. However, the possibility of the multi-functionality of these proteins cannot be rejected, with CRD and the PR-1 domains performing different functions. Such directional mutagenesis of venom where molecular surface and functional regions act as hotspots for mutations should have diversified the ability of CRISPs to target numerous cell and/or tissue types in the prey. The amino-acid diversification influenced by a strong positive selection on these proteins suggests that they

are one of the most indispensible components of reptilian venom. In contrast, the significant negative selection pressures observed in mammalian CRISPs highlights the important roles they play in mammalian systems.

Supplementary Materials

Supplementary materials are available at Molecular Biology and Evolution online

(http://www.mbe.oxfordjournals.org/).

1. Sequence used in the study

2. Multiple Sequence Alignment

2. 1 Snake CRISP multiple sequence alignment

2. 2 Snake CRISP multiple sequence alignment

2. 3 Lizard CRISPs

3. Amino-acid variability of mammalian cysteine-rich secretory proteins

4. Maximum-likelihood estimations of mammalian CRISPs in detail

Acknowledgements

We are indebted to João Paulo Machado, Rute Fonseca, Debapriyo Chakroborty, Rui Borges and Emanuel Maldonado for support and discussions. We are also thankful to the anonymous reviewers for their valuable comments and suggestions which lead to the improvement of the manuscript. This research is supported by FCT (Fundação para a Ciência e a Tecnologia) grant SFRH/BD/61959/2009 conferred to KS and in part by the FCT project PTDC/BIA-BDE/69144/2006 (FCOMP-01-0124-FEDER-007065), PTDC/AAC-AMB/104983/2008 (FCOMP-01-0124-FEDER-008610) and PTDC/AAC-AMB/121301/2010.

Literature Cited

Alessandri-Haber N, Lecoq A, Gasparini S, et al. 1999. Mapping the functional anatomy of BgK on Kv1.1, Kv1.2, and Kv1.3. Clues to design analogs with enhanced selectivity. J Biol Chem 274:35653-35661.

Anisimova M, Bielawski J P, Yang Z. 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. Molecular Biology and Evolution 19:950-958.

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195-201.

Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res 38:W529-533.

Barlow A, Pook C E, Harrison R A, Wuster W. 2009. Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. Proc Biol Sci 276:2443-2449.

Bielawski J P, Yang Z. 2004. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. J Mol Evol 59:121-132.

Calvete J J, Moreno-Murciano M P, Theakston R D, Kisiel D G, Marcinkiewicz C. 2003. Snake venom disintegrins: novel dimeric disintegrins and structural diversification by disulphide bond engineering. Biochem J 372:725-734.
Chang T Y, Mao S H, Guo Y W. 1997. Cloning and expression of a cysteine-rich venom protein from Trimeresurus mucrosquamatus (Taiwan habu). Toxicon 35:879-888.
Daltry J C, Wuster W, Thorpe R S. 1996. Diet and snake venom evolution. Nature 379:537-540.
DeLano W L. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA.
Delport W, Poon A F, Frost S D, Kosakovsky Pond S L. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455-2457.
Doley R, Pahari S, Mackessy S P, Kini R M. 2008. Accelerated exchange of exon segments in Viperid three-finger toxin genes (Sistrurus catenatus edwardsii; Desert Massasauga). BMC Evol Biol 8:196.
Edgar R C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32:1792-1797.
Fraczkiewicz R, Braun W. 1998. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. Journal of Computational Chemistry 19:319-333.
Fry B G. 2005. From genome to "venome": Molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Research 15:403-420.
Fry B G, Scheib H, van der Weerd L, Young B, McNaughtan J, Ryan Ramjan S F, Vidal N, Poelmann R E, Norman J A. 2008. Evolution of an arsenal: Structural and functional diversification of the venom system in the advanced snakes (Caenophidia). Molecular and Cellular Proteomics 7:215-246.
Fry B G, Vidal N, Norman J A, et al. 2006. Early evolution of the venom system in lizards and snakes. Nature 439:584-588.
Fry B G, Wuster W. 2004. Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. Mol Biol Evol 21:870-883.
Fry B G, Wüster W, Kini R M, Brusic V, Khan A, Venkataraman D, Rooney A P. 2003a. Molecular evolution and phylogeny of elapid snake venom three-finger toxins. Journal of Molecular Evolution 57:110-129.
Fry B G, Wuster W, Ryan Ramjan S F, Jackson T, Martelli P, Kini R M. 2003b. Analysis of Colubroidea snake venoms by liquid chromatography with mass spectrometry: evolutionary and toxinological implications. Rapid Commun Mass Spectrom 17:2047-2062.
Gibbs G M, Bianco D M, Jamsai D, Herlihy A, Ristevski S, Aitken R J, De Kretser D M, O'Bryan M K. 2007. Cysteine-rich secretory protein 2 binds to mitogen-activated protein kinase kinase kinase 11 in mouse sperm. Biology of Reproduction 77:108-114.
Gibbs G M, O'Bryan M K. 2007. Cysteine rich secretory proteins in reproduction and venom. Society of Reproduction and Fertility supplement 65:261-267.
Gibbs G M, Scanlon M J, Swarbrick J, Curtis S, Gallant E, Dulhunty A F, O'Bryan M K. 2006. The cysteine-rich secretory protein domain of Tpx-1 is related to ion channel toxins and regulates ryanodine receptor Ca2+ signaling. J Biol Chem 281:4156-4163.
Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Molecular Biology and Evolution 11:725-736.
Gu X. 1999. Statistical methods for testing functional divergence after gene duplication. Mol Biol Evol 16:1664-1674.
Guindon S, Dufayard J F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307-321.

Guo M, Teng M, Niu L, Liu Q, Huang Q, Hao Q. 2005. Crystal structure of the cysteine-rich secretory protein stecrisp reveals that the cysteine-rich domain has a K+ channel inhibitor-like fold. Journal of Biological Chemistry 280:12405-12412.
Haendler B, Habenicht U F, Schwidetzky U, Schuttke I, Schleuning W D. 1997. Differential androgen regulation of the murine genes for cysteine-rich secretory proteins (CRISP). Eur J Biochem 250:440-446.
Haendler B, Kratzschmar J, Theuring F, Schleuning W D. 1993. Transcripts for cysteine-rich secretory protein-1 (CRISP-1; DE/AEG) and the novel related CRISP-3 are expressed under androgen control in the mouse salivary gland. Endocrinology 133:192-198.
Hill R E, Mackessy S P. 2000. Characterization of venom (Duvernoy's secretion) from twelve species of colubrid snakes and partial sequence of four venom proteins. Toxicon 38:1663-1687.
Huelsenbeck J P, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.
Huson D H, Richter D C, Rausch C, Dezulian T, Franz M, Rupp R. 2007. Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics 8:460.
Kelly C M, Barker N P, Villet M H. 2003. Phylogenetics of advanced snakes (Caenophidia) based on four mitochondrial genes. Syst Biol 52:439-459.
Kierszenbaum A L, Lea O, Petrusz P, French F S, Tres L L. 1981. Isolation, culture, and immunocytochemical characterization of epididymal epithelial cells from pubertal and adult rats. Proc Natl Acad Sci U S A 78:1675-1679.
Kimura M. 1983. The Neutral Theory of Molecular Evolution: Cambridge University Press.
Kini R M. 2004. Platelet aggregation and exogenous factors from animal sources. Curr Drug Targets Cardiovasc Haematol Disord 4:301-325.
Kini R M, Chan Y M. 1999. Accelerated evolution and molecular surface of venom phospholipase A2 enzymes. J Mol Evol 48:125-132.
Kitajima S, Sato F. 1999. Plant pathogenesis-related proteins: Molecular mechanisms of gene expression and protein function. Journal of Biochemistry 125:1-8.
Kosakovsky Pond S L, Posada D, Gravenor M B, Woelk C H, Frost S D. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23:1891-1901.
Lange A, Giller K, Hornig S, Martin-Eauclaire M F, Pongs O, Becker S, Baldus M. 2006. Toxin-induced conformational changes in a potassium channel revealed by solid-state NMR. Nature 440:959-962.
Lynch V J. 2007. Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. BMC Evol Biol 7:2.
Mackessy S P. 2002. Biochemistry and pharmacology of colubrid snake venoms. Journal of Toxicology - Toxin Reviews 21:43-83.
Miyata T, Suga H. 2001. Divergence pattern of animal gene families and relationship with the Cambrian explosion. Bioessays 23:1018-1027.
Morita T. 2005. Structures and functions of snake venom CLPs (C-type lectin-like proteins) with anticoagulant-, procoagulant-, and platelet-modulating activities. Toxicon 45:1099-1114.
Morrissette J, Kratzschmar J, Haendler B, et al. 1995. Primary structure and properties of helothermine, a peptide toxin that blocks ryanodine receptors. Biophys J 68:2280-2288.
Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929-936.
Nobile M, Magnelli V, Lagostena L, Mochca-Morales J, Possani L D, Prestipino G. 1994. The toxin helothermine affects potassium currents in newborn rat cerebellar granule cells. J Membr Biol 139:49-55.
Peichoto M E, Leme A F, Pauletti B A, Batista I C, Mackessy S P, Acosta O, Santoro M L. 2010. Autolysis at the disintegrin domain of patagonfibrase, a metalloproteinase from Philodryas patagoniensis (Patagonia Green Racer; Dipsadidae) venom. Biochim Biophys Acta 1804:1937-1942.

Pfisterer P, Konig H, Hess J, Lipowsky G, Haendler B, Schleuning W D, Wirth T. 1996. CRISP-3, a protein with homology to plant defense proteins, is expressed in mouse B cells under the control of Oct2. Mol Cell Biol 16:6160-6168.
Pond S L, Frost S D. 2005. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. Mol Biol Evol 22:478-485.
Pond S L, Frost S D, Muse S V. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676-679.
Pond S L, Scheffler K, Gravenor M B, Poon A F, Frost S D. 2010. Evolutionary fingerprinting of genes. Mol Biol Evol 27:520-536.
Posada D. 2008. jModelTest: Phylogenetic model averaging. Molecular Biology and Evolution 25:1253-1256.
Pupko T, Pe'er I, Shamir R, Graur D. 2000. A fast algorithm for joint reconstruction of ancestral amino acid sequences. Mol Biol Evol 17:890-896.
Ronquist F, Huelsenbeck J P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.
Rosenberg N A. 2002. The probability of topological concordance of gene trees and species trees. Theor Popul Biol 61:225-247.
Sasa M. 1999. Diet and snake venom evolution: can local selection alone explain intraspecific venom variation? Toxicon 37:249-252; author reply 253-260.
Schambony A, Hess O, Gentzel M, Topfer-Petersen E. 1998. Expression of CRISP proteins in the male equine genital tract. J Reprod Fertil Suppl 53:67-72.
Shikamoto Y, Suto K, Yamazaki Y, Morita T, Mizuno H. 2005. Crystal structure of a CRISP family Ca2+ -channel blocker derived from snake venom. J Mol Biol 350:735-743.
Soto J G, White S A, Reyes S R, Regalado R, Sanchez E E, Perez J C. 2007. Molecular evolution of PIII-SVMP and RGD disintegrin genes from the genus Crotalus. Gene 389:66-72.
Suzuki N, Yamazaki Y, Brown R L, Fujimoto Z, Morita T, Mizuno H. 2008. Structures of pseudechetoxin and pseudecin, two snake-venom cysteine-rich secretory proteins that target cyclic nucleotide-gated ion channels: implications for movement of the C-terminal cysteine-rich domain. Acta Crystallographica Section D 64:1034-1042.
Suzuki Y, Nei M. 2004. False-positive selection identified by ML-based methods: examples from the Sig1 gene of the diatom Thalassiosira weissflogii and the tax gene of a human T-cell lymphotropic virus. Mol Biol Evol 21:914-921.
Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology 56:564-577.
Todd A E, Orengo C A, Thornton J M. 1999. Evolution of protein function, from a structural perspective. Curr Opin Chem Biol 3:548-556.
Tsai I H, Tsai H Y, Wang Y M, Tun P, Warrell D A. 2007. Venom phospholipases of Russell's vipers from Myanmar and eastern India--cloning, characterization and phylogeographic analysis. Biochim Biophys Acta 1774:1020-1028.
Udby L, Calafat J, Sørensen O E, Borregaard N, Kjeldsen L. 2002. Identification of human cysteine-rich secretory protein 3 (CRISP-3) as a matrix protein in a subset of peroxidase-negative granules of neutrophils and in the granules of eosinophils. Journal of Leukocyte Biology 72:462-469.
Wang F, Li H, Liu M N, et al. 2006. Structural and functional analysis of natrin, a venom protein that targets various ion channels. Biochem Biophys Res Commun 351:443-448.
Wang J, Shen B, Guo M, et al. 2005. Blocking effect and crystal structure of natrin toxin, a cysteine-rich secretory protein from Naja atra venom that targets the BKCa channel. Biochemistry 44:10145-10152.
Woolley S, Johnson J, Smith M J, Crandall K A, McClellan D A. 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics 19:671-672.

Xia X, Xie Z. 2001. DAMBE: Software package for data analysis in molecular biology and evolution. Journal of Heredity 92:371-373.
Xia X, Xie Z, Salemi M, Chen L, Wang Y. 2003. An index of substitution saturation and its application. Molecular Phylogenetics and Evolution 26:1-7.
Yamazaki Y, Hyodo F, Morita T. 2003. Wide distribution of cysteine-rich secretory proteins in snake venoms: isolation and cloning of novel snake venom cysteine-rich secretory proteins. Arch Biochem Biophys 412:133-141.
Yamazaki Y, Morita T. 2004. Structure and function of snake venom cysteine-rich secretory proteins. Toxicon 44:227-231.
Yamazaki Y, Morita T. 2007. Snake venom components affecting blood coagulation and the vascular system: structural similarities and marked diversity. Curr Pharm Des 13:2872-2886.
Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Molecular Biology and Evolution 15:568-573.
Yang Z. 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. Journal of Molecular Evolution 51:423-432.
Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24:1586-1591.
Yang Z, Bielawski J R. 2000. Statistical methods for detecting molecular adaptation. Trends in Ecology and Evolution 15:496-503.
Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908-917.
Yang Z, Nielsen R, Goldman N, Pedersen A M. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431-449.
Yang Z, Wong W S W, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Molecular Biology and Evolution 22:1107-1118.
Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol 22:2472-2479.

…

Tables

Table 1.1. Lineage specific maximum-likelihood parameter estimates for toxicoferan-reptilian CRISPs.

Model	ω[a]	Likelihood (ι)	Prop. of sites with ω > 1[b]	No. of sites with ω > 1[c]	Sign[d]
REPTILES					
SNAKES					
Two-ratio Model	1.21	-10370.88844	-	-	***P << 0.001**
Branch-site Model A	3.76	-9915.548688	25.3%	23 (PP ≥ 0.99) 13 (PP ≥ 0.95)	***P << 0.001**
Clade Model C	**3.84**	-9876.069435	20.0%	-	**P << 0.001**
LIZARDS					
Two-ratio Model	0.64	-10465.99893	-	-	***P << 0.001**
Branch-site Model A	3.21	-10093.64418	18.1%	3 (PP ≥ 0.99) 12 (PP ≥ 0.95)	***P << 0.001**
Clade Model C	**2.33**	-9876.069435	20.0%	-	**P << 0.001**
SNAKE LINEAGES					
Colubridae					
Site Model 8	1.42	-3508.159988	28.0% (ω_2= 3.89)	22 (PP ≥ 0.99) 16 (PP ≥ 0.95)	**P << 0.001**
Two-ratio Model	1.16	-10382.84638	-	-	N.S
Branch-site Model A	6.73	-9997.185149	8.1%	3 (PP ≥ 0.99) 3 (PP ≥ 0.95)	***P << 0.001**
Clade Model C	**4.10**	-6578.072786	23.7%	-	**P << 0.001**
Viperidae					
Site Model 8	1.42	-2077.819161	22.7% (ω_2= 5.24)	18 (PP ≥ 0.99) 7 (PP ≥ 0.95)	**P << 0.001**
Two-ratio Model	1.05	-10383.8539	-	-	N.S
Branch-site Model A	5.66	-9998.09742	15.0%	3 (PP ≥ 0.99) 6 (PP ≥ 0.95)	***P << 0.001**
Clade Model C	**4.19**	-6578.072786	23.7%	-	**P << 0.001**
Elapidae					
Site Model 8	1.50	-2874.216899	25.3% (ω_2= 4.45)	15 (PP ≥ 0.99) 14 (PP ≥ 0.95)	**P << 0.001**
Two-ratio Model	1.28	-10381.65348	-	-	N.S
Branch-site Model A	5.52	-9981.464478	16.1%	4 (PP ≥ 0.99) 6 (PP ≥ 0.95)	***P << 0.001**
Clade Model C	**2.86**	-6578.072786	23.7%	-	**P << 0.001**

Legend:

a: dN/dS (weighted average)

b: Proportion of sites with ω>1

c: Number of sites with ω>1 under Bayes empirical Bayes approach with a posterior probability (PP) of more than or equal to 0.99 and 0.95

d: Significance of the model in comparison with the null model

***:** Significant at 0.001after Bonferroni correction

NS: Not significant

Table 1.2. Site-specific maximum-likelihood parameter estimates for toxicoferan-reptilian CRISPs.

Model	Likelihood (ι)	ω_0[a]	Parameters	Sign.[b]	No. of Sites with $\omega > 1$[c]
SNAKES					**B.E.B**
M0 (One ratio)	-6964.393838	**1.14**	**=** ω_0		-
M0 (constrained)	-6965.771794	1.0	ω_0 constrained to 1		-
M1 (Neutral)	-6711.896261	0.55	**P_0:** 0.48460 **ω_0:** 0.07 **P_1:** 0.48460 **ω_1:** 1.0		-
M2 (Selection)*	-6595.953989	**1.24**	**P_0:** 0.39687 **ω_0:** 0.05 **P_1:** 0.36260 **ω_1:** 1.0 **P_2:** 0.24054 **ω_2:** 3.58	**P << 0.001**	24(PP ≥ 0.99) 13 (PP ≥ 0.95)
M3 (Discrete)*	-6593.227401	**1.37**	**P_0:** 0.43544 **ω_0:** 0.08 **P_1:** 0.38890 **ω_1:** 1.0 **P_2:** 0.17566 **ω_2:** 1.0	**P << 0.001**	-
M7 (beta)	-6713.705160	0.52	**p:** 0.16753 **q:** 0.15459		-
M8 (beta and ω)*	-6598.579842	**1.22**	**p_0:** 0.745 **p:** 0.162 **q:** 0.183 **p_1:** 0.255 **ω:** 3.43	**P << 0.001**	**26** (PP ≥ 0.99) **11** (PP ≥ 0.95)
LIZARDS					
M0 (One ratio)	-3647.864377	0.78	**=** ω_0		-
M0 (constrained)	-3650.682815	1.0	ω_0 constrained to 1		-
M1 (Neutral)	-3588.206161		**P_0:** 0.45529 **ω_0:** 0.10 **P_1:** 0.54471 **ω_1:** 1.0		-
M2 (Selection)*	-3566.671365	**1.04**	**P_0:** 0.38959 **ω_0:** 0.10 **P_1:** 0.47093 **ω_1:** 1.0 **P_2:** 0.47093 **ω_2:** 3.73	**P << 0.001**	2 (PP ≥ 0.99) 8 (PP ≥ 0.95)
M3 (Discrete)*	-3566.221975	**1.01**	**P_0:** 0.23786 **ω_0:** 0.00 **P_1:** 0.56320 **ω_1:** 0.68 **P_2:** 0.19894 **ω_2:** 3.15	**P << 0.001**	-
M7 (beta)	-3590.931288	0.59	**p:** 0.18663 **q:** 0.12454		-
M8 (beta and ω)*	-3566.284106	**1.02**	**p_0:** 0.836 **p:** 0.304 **q:** 0.183 **p_1:** 0.253 **ω:** 3.44	**P << 0.001**	**6** (PP ≥ 0.99) **8** (PP ≥ 0.95)

Legend:

a: dn/ds (weighted average)

b: Significance of the model in comparison with the null model

c: Number of sites with $\omega > 1$ under the Bayes empirical Bayes approach with a posterior probability (PP) more than or equal to 0.99 and 0.95

***** Models which allow $\boldsymbol{\omega} > 1$

Table 1.3. Maximum-likelihood parameter estimates for mammalian CRISPs.

	FEL[a]		REL[b]		SLAC[c]			Codeml	
	ω>1[d]	ω<1[e]	ω>1[d]	ω<1[e]	ω>1[d]	ω<1[e]	ω	**M8 (ω)**	ω>1[f]
CRISP1	0	32	0	All	0	10	0.50	0.55 [N.S]	5
CRISP2	3	49	3	49	0	5	0.40	0.40 [N.S]	-
CRISP3	1	21	1	21	0	8	0.61	0.68 [N.S]	4

Legend:

a: Fixed-effects likelihood

b: Random-effects likelihood

c: Single Likelihood Ancestor Counting

d: Number of positively selected sites at 0.05 significance

e: Number of negatively selected sites at 0.05 significance

ω: mean dN/dS

f: Number of positively selected sites under B.E.B approach with posterior probability greater than 0.95

N.S: Not significant in comparison with the null model M7 (beta)

Table 2. Amino-acid sites under positive selection.

Sites		PAML		TreeSAAP			
				Radical changes in amino-acid properties			
Position	A.A	M2A	M8	Chemical	Structural	Total	A.S.A
Snakes							
30	Q	3.495 ± 0.12	3.496 ± 0.095	-	α_C V^0	2	**92.9%**
73	N	3.392 ± 0.512	3.417 ± 0.453	-	α_C	1	**57.6%**
75	N	3.31 ± 0.664	-	-	α_C	1	35.0%
76	L	3.396 ± 0.444	3.421 ±0.444	-	α_C	1	7.0%
81	D	3.501 ±0.011	3.5 ±0.011	-	α_C	1	**75.9%**
82	Y	3.501 ±0.016	3.5 ±0.016	-	α_C B_I	2	**78.9%**
83	S	3.454 ±0.297	3.465 ±0.297	-	α_C B_I	2	37.5%
87	E	3.499 ±0.053	3.499 ±0.053	-	α_C B_I	2	**57.7%**
100	N	3.496 ±0.092	3.497 ±0.092	pK'	α_C	2	**82.6%**
102	R	3.5 ± 0.031	3.5 ± 0.031	pK'	α_C	2	28.2%
103	A	3.472 ±0.226	3.48 ± 0.226	pK'	α_C	2	**64.3%**
106	E	3.492 ±0.121	3.494 ±0.121	pK'	α_C	2	**47.2%**
110	L	3.501 ±0.005	3.5 ±0.005	pK'	α_C R_a N_S H_P	5	32.9%
115	**Y**	**3.497 ±0.081**	**3.497 ±0.081**	**pK'**	**α_C R_a P_β N_S H_P**	**6**	**47.1%**
119	**V**	**3.471 ±0.237**	**3.478 ±0.237**	**pK'**	**α_C R_a P_β N_S H_P**	**6**	**53.2%**
145	I	3.501 ±0.011	3.5 ±0.011	pK'	α_C R_a P_β H_P	5	0%
150	N	3.492 ±0.124	3.494 ±0.124	pK'	α_C R_a P_β H_P	5	22.9%
156	E	3.501 ±0.005	3.5 ±0.005	pK'	α_C R_a P_β H_P	5	**63.9%**
168	**S**	**3.282 ± 0.706**	**-**	**pK'**	**α_C R_a P_β H_{nc} N_S H_P**	**7**	21.7%
171	**M**	**3.5 ±0.025**	**3.5 ±0.025**	**pK'**	**A_C R_F R_a P_β H_{nc} N_S H_P**	**8**	**45.9%**
172	**R**	**3.501 ±0.004**	**3.5 ±0.004**	**-**	**α_C B_I R_F R_a P_β H_{nc} N_S H_P**	**8**	**87.0%**
174	**S**	**3.499 ±0.057**	**3.499 ±0.057**	**-**	**α_C B_I R_F R_a P_β H_{nc} N_S H_P**	**8**	**54.4%**
186	**G**	**3.386 ±0.475**	**3.409 ±0.475**	**-**	**α_C B_I P_C R_F R_a P_β H_{nc} N_S H_P**	**9**	**41.7%**
202	**T**	**3.402 ±0.429**	**3.426 ±0.429**	**-**	**α_C B_I P_C R_F R_a P_β N_S α_m H_P**	**9**	**97.9%**
203	**L**	**3.501 ±0.004**	**3.5 ±0.004**	**-**	**α_C B_I P_C R_F R_a P_β N_S α_m H_P**	**9**	**65.1%**
204	**Y**	**3.499 ±0.058**	**3.499 ±0.058**	**-**	**α_C B_I P_C R_F R_a P_β α_m**	**7**	33.6%
206	E	3.501 ±0.004	3.5 ±0.004	-	α_C B_I R_a P_β α_m	5	**40.0%**
207	Y	3.45 ±0.308	3.462 ±0.308	-	α_C P_β α_m	3	35.3%
211	D	3.445 ±0.322	3.459 ± 0.322	-	α_C α_m	2	**88.7%**
212	S	3.499 ±0.05	3.499 ±0.05	-	α_C α_m	2	**55.9%**
214	V	3.356 ±0.535	3.383 ±0.535	-	α_C α_m	2	**47.6%**
215	K	3.432 ±0.36	3.448 ±0.36	-	α_C α_m	2	**78.1%**
217	S	3.501 ±0.005	3.5 ±0.005	-	α_C α_m	2	**45.0%**
218	S	3.5 ±0.035	3.5 ±0.035	-	α_C	1	**60.0%**
220	Q	3.46 ±0.278	3.469 ±0.278	-	α_C	1	**95.6%**
222	E	3.254 ± 0.746	-	-	α_C	1	**90.8%**
223	W	3.488 ±0.148	3.491 ±0.148	-	α_C	1	**55.7%**
224	I	3.481 ±0.19	3.486 ±0.19	-	α_C	1	4.1%
226	S	3.465 ±0.259	3.473 ±0.259	-	α_C	1	**60.2%**
231	S	3.394 ±0.445	3.42 ±0.445	-	α_C	1	19.9%
235	H	3.501 ±0.019	3.5 ±0.019	-	α_C	1	**84.3%**

Lizard CRISPs							
18	H	3.634 ± 0.736	2.016 +- 1.129	-	$_pH_I$ α_C	2	-
65	T	3.678 ± 0.661	3.086 +- 0.581	-	α_C	1	**59.1%**
85	T	3.684 ± 0.649	3.089 +- 0.575	pK'	$_pH_I$ α_C	3	**86.3%**
86	S	3.681 ± 0.657	3.087 +- 0.579	pK'	c $_pH_I$ α_C	3	**86.1%**
130	T	3.586 +- 0.806	3.059 +- 0.623	-	$_pH_I$ α_C	2	**91.5%**
149	T	3.600 +- 0.790	3.060 +- 0.623	-	α_C	1	0.0%
158	A	-	3.000 +- 0.703	-	α_C	1	**80.3%**
160	R	3.706 +- 0.605	3.096 +- 0.563	-	α_C	1	**71.9%**
176	E	3.671 +- 0.676	3.084 +- 0.585	-	α_C	1	**100.0%**
181	E	-	3.015 +- 0.680	-	α_C	1	33.3%
184	A	3.693 +- 0.633	3.091 +- 0.572	-	α_C	1	**56.5%**
187	E	-	2.987 +- 0.718	-	$_pH_I$ α_C	2	**87.0%**
207	H	-	3.017 +- 0.682	-	$_pH_I$ α_C	2	**41.0%**
210	Q	3.610 +- 0.778	3.060 +- 0.624	-	$_pH_I$ α_C	2	**72.2%**

Legend:

Amino-acid sites detected by PAML and TreeSAAP as under positive selection along with the ω estimation and Bayesian (BEB) analysis posterior probabilities for sites with P ≥ 95% under M2 and M8 models.

TreeSAAP: Radical changes in amino-acid properties (chemical, structural and other property changes) under category 6 and/or 7 and/or 8. Amino-acid sites that belong to the Type II class (greater than 6 property changes) are represented in bold letters.
Amino-acid property symbols used: Average number of surrounding residues ***(Ns)***, β-structure tendencies ***(Pβ)***, Bulkiness ***(Bl)***, Composition ***(c),*** Chromatographic index ***(RF)***, Coil tendencies ***(Pc)***, Equilibrium constant for ionization of COOH ***(pK')***, Isoelectric point ***(pHi)***, Normalized consensus hydrophobicity ***(Hnc)***, Partial specific volume ***(V^0)***, Polar requirement ***(Pr),*** Power to be at C-terminus of the *α*-helix **(*αc*)**, Power to be in the middle of an *α*-helix ***(αm),*** Solvent accessible reduction ratio ***(Ra)***, Surrounding hydrophobicity **(*Hp*)**

Table 3

Functional Divergence

Comparison	Theta (ɵ) [a]	S.E (ɵ) [b]	LRT [c]	P [d]
Lizards vs Elapidae	0.0010	± 0.022361	0	**N.S**
Lizards vs Colubridae	0.1552	± 0.081855	3.594956	**N.S**
Lizards vs Viperidae	**0.3336**	**± 0.130631**	**6.521658**	**p < 0.05**
Colubridae vs Elapidae	**0.4080**	**± 0.097791**	**17.406791**	**p < 0.001**
Colubridae vs Viperidae	**0.3112**	**± 0.103415**	**9.055442**	**p < 0.01**
Elapidae vs Viperidae	**0.5048**	**± 0.14939**	**11.418214**	**p < 0.001**
CRISP1 vs CRISP2	**0.5032**	**0.115574**	**18.956533**	**p < 0.001**
CRISP1 vs CRISP3	**0.7000**	**0.137197**	**26.031785**	**p < 0.001**
CRISP2 vs CRISP3	**0.8064**	**0.122422**	**43.389136**	**p < 0.001**

a: Theta parameter or coefficient of functional divergence
b: Standard Error
c: Likelihood ratio test between the alternate model that allows ɵ > 1 with the null model that does not
d: p-value; significant values are highlighted in bold letters

Figures

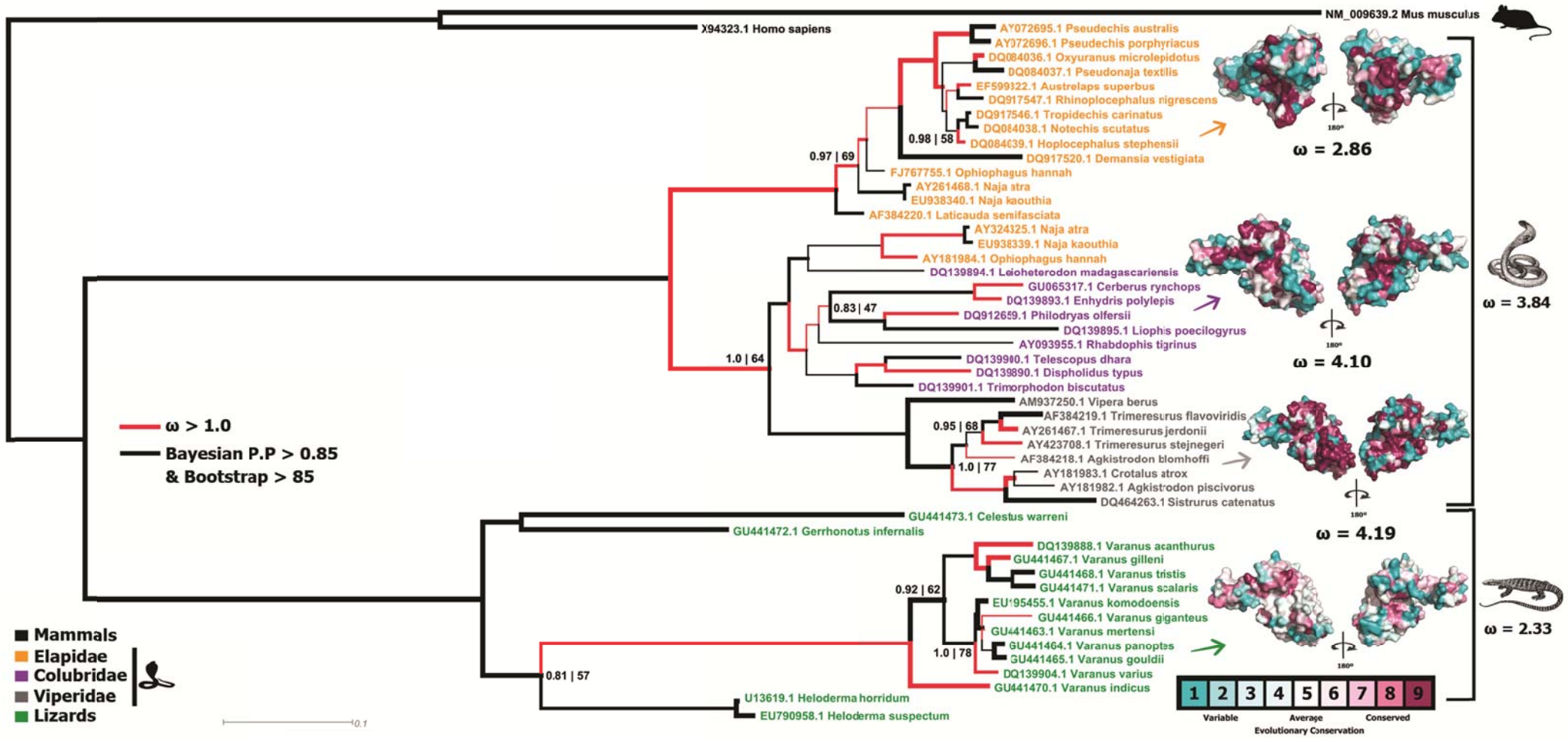

Figure 1. Molecular phylogeny of toxicoferan-reptilian CRISP genes.

Branches with the Bayesian posterior probability (B.P.P) greater than 0.85 and 850 or more (out of 1000) bootstrap replicate support (BS) are presented as thick lines. Branches that have lower BS (<850) but a stronger B.P.P support (> 0.85) are presented as "B.P.P | B.S". Branches under positive selection are represented as gray lines (as red lines in the online version of the publication). The three-dimensional models of CRISPs depicting the amino-acid variability are also shown along with the omega estimate (Clade Model C) of the respective lineages.

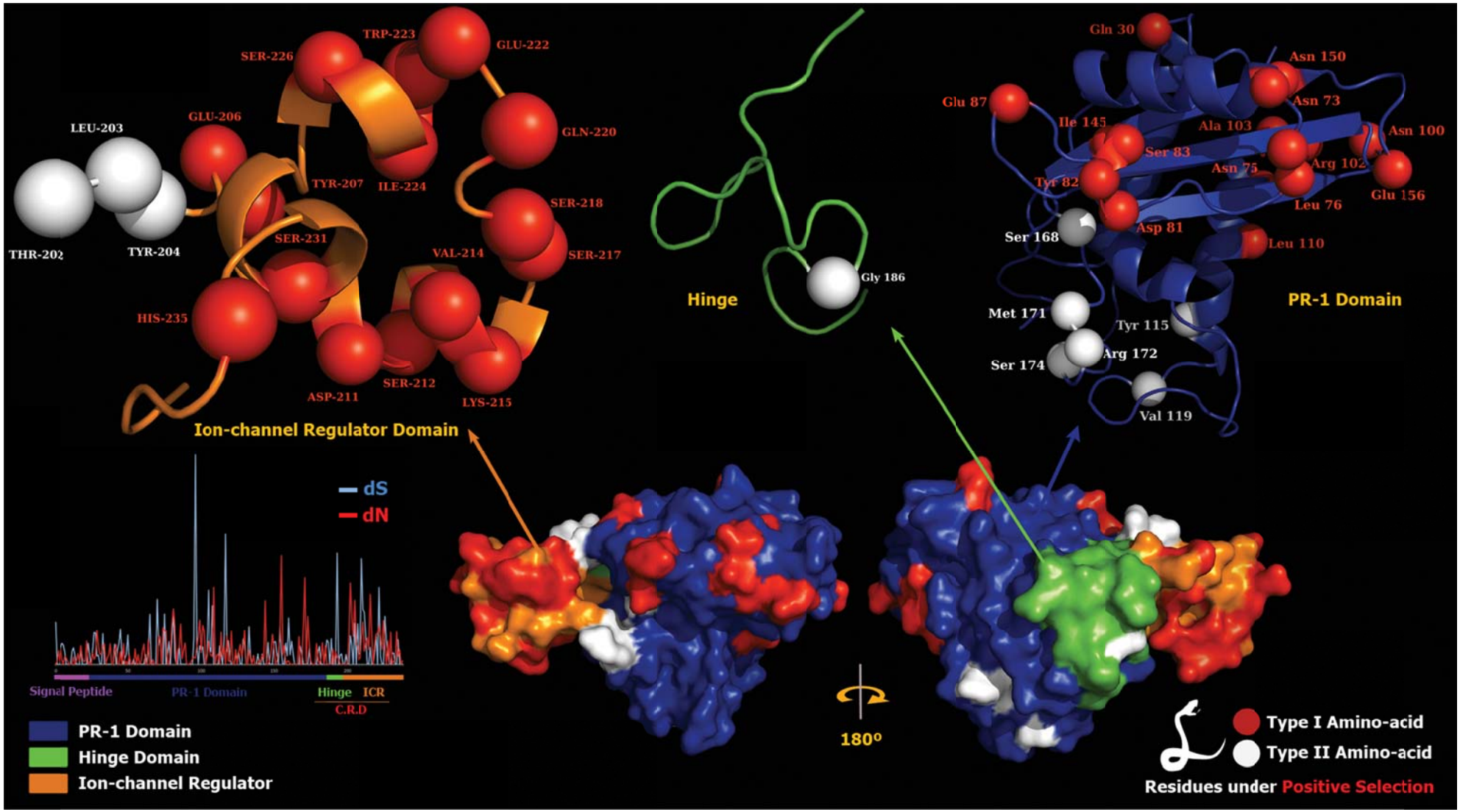

Figure 2.1. Cysteine-rich Secretory Proteins: Snakes.

Three-dimensional model of snake CRISPs depicting the locations of positively selected amino-acid sites detected by PAML (Model 2 and Model 8, PP ≥ 0.95) and TreeSAAP ($p < 0.001$). Type I ($n \leq 6$ amino acid property changes) and type II amino-acid sites ($n \geq 6$) are shown in dark and light colours, respectively.

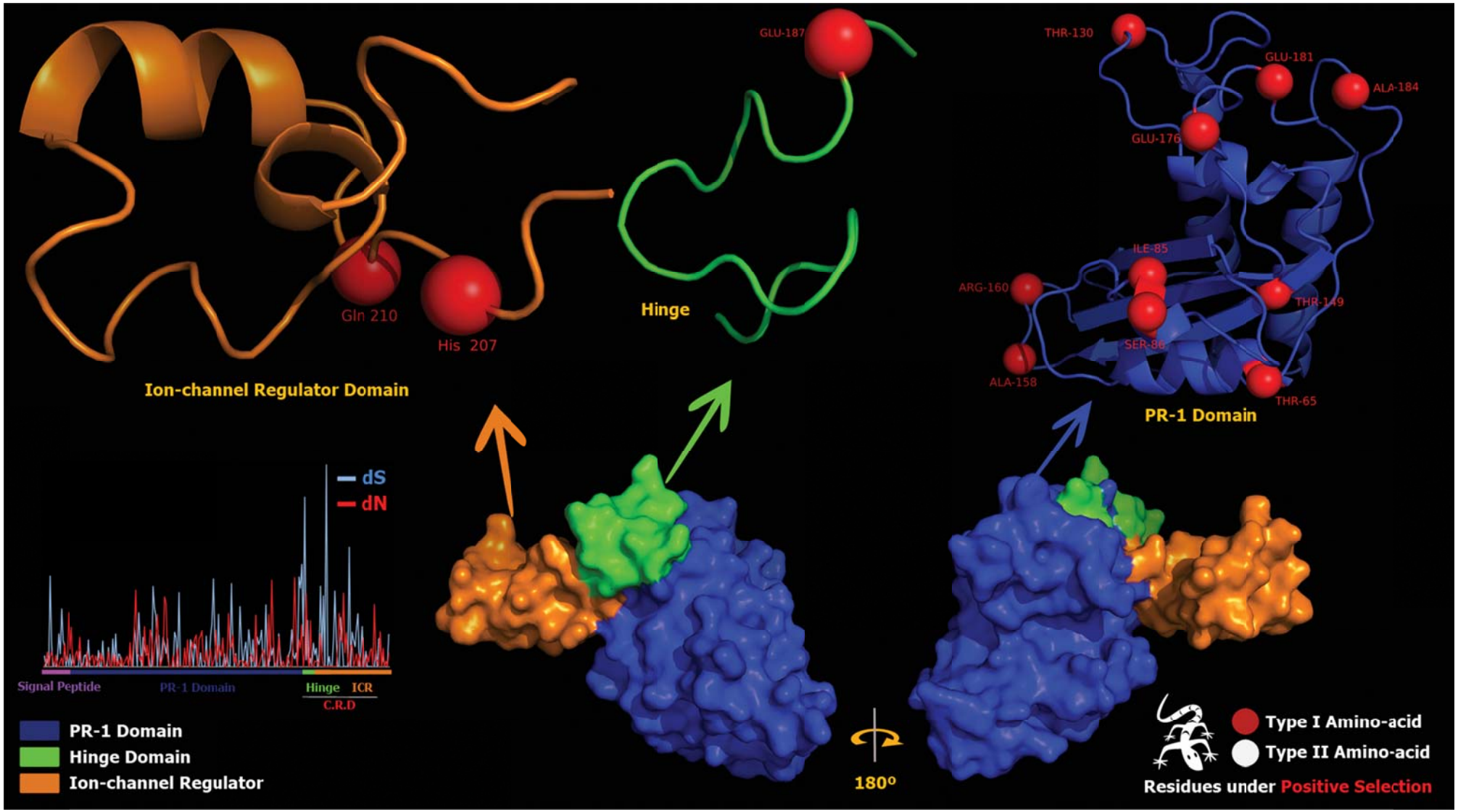

Figure 2.2. Cysteine-rich Secretory Proteins: Lizards.

Three-dimensional model of lizard CRISPs depicting the locations of positively selected amino-acids detected by PAML (Model 2 and Model 8, PP ≥ 0.95) and TreeSAAP ($p \leq 0.001$). Type I amino-acids ($n \leq 6$ amino acid property changes) are shown. There were no type II residues ($n \geq 6$) in the lizard CRISPs.

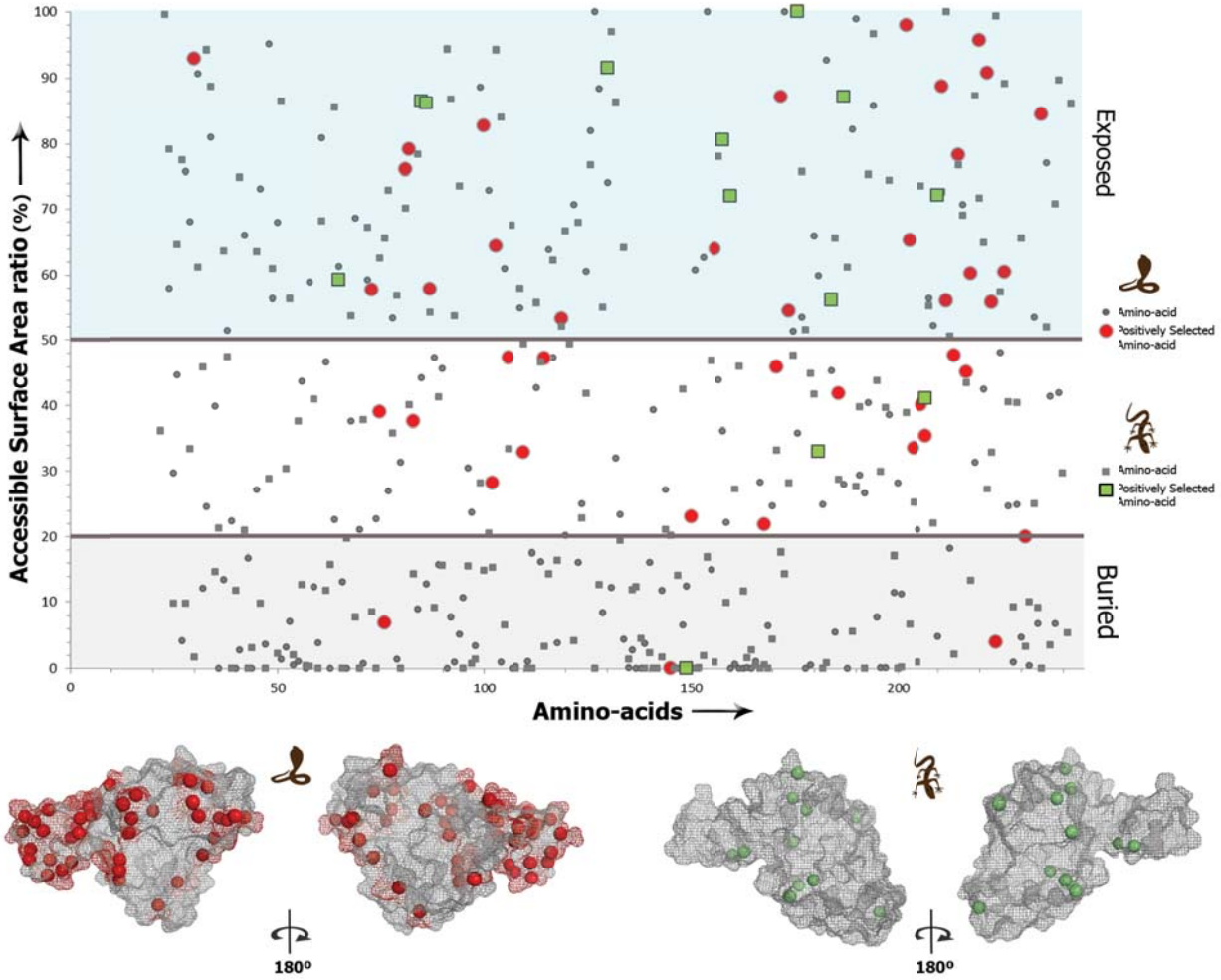

Figure 3. Surface accessibility of Cysteine-rich Secretory Proteins.

A plot of amino-acid positions (x-axis) against accessible surface area (ASA) ratio (y-axis) indicating the positions of amino-acids (exposed or buried) in the crystal structure of toxicoferan-reptilian CRISPs is presented. The snake CRISP amino-acids are represented as small and large (positively selected) dots, while the lizard CRISP amino-acids are represented as small and large (positively selected) squares. Amino-acids with an ASA ratio of more than 50% are considered to be exposed to the surrounding solvent while those with a ratio lesser than 20% are considered to be buried. Three dimensional models of snake and lizard CRISPs depicting the locations of positively selected sites are also presented.

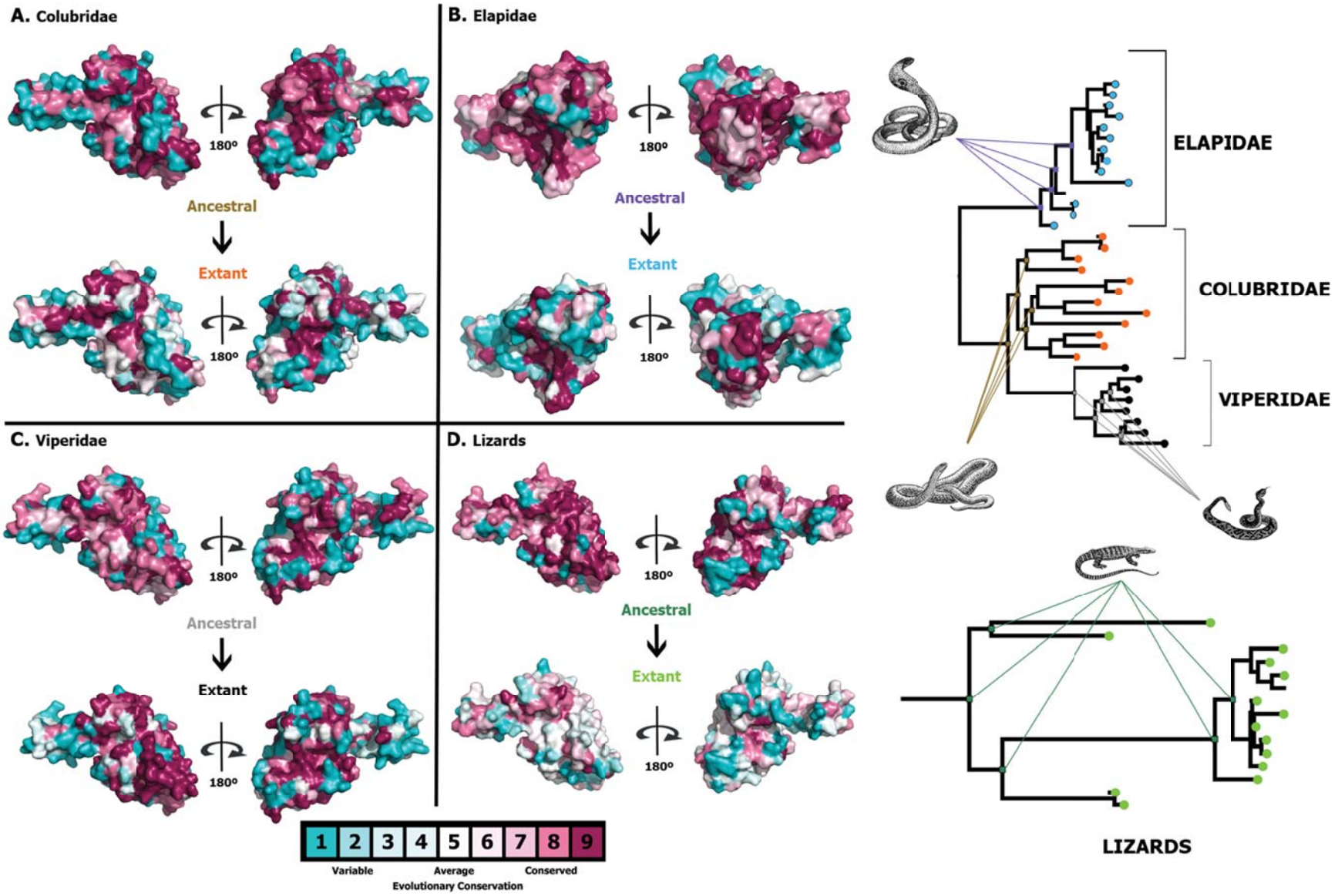

Figure 4. Amino-acid variability in Ancestral and Extant toxicoferan CRISPs.

Three-dimensional models depicting the amino-acid variability in ancestral and extant toxicoferan reptilian CRISPs. Phylogenetic trees showing the nodes that were sampled for creating the three-dimensional models of ancestral toxicoferan CRISPs and for the calculation of selection pressures are shown on the right.

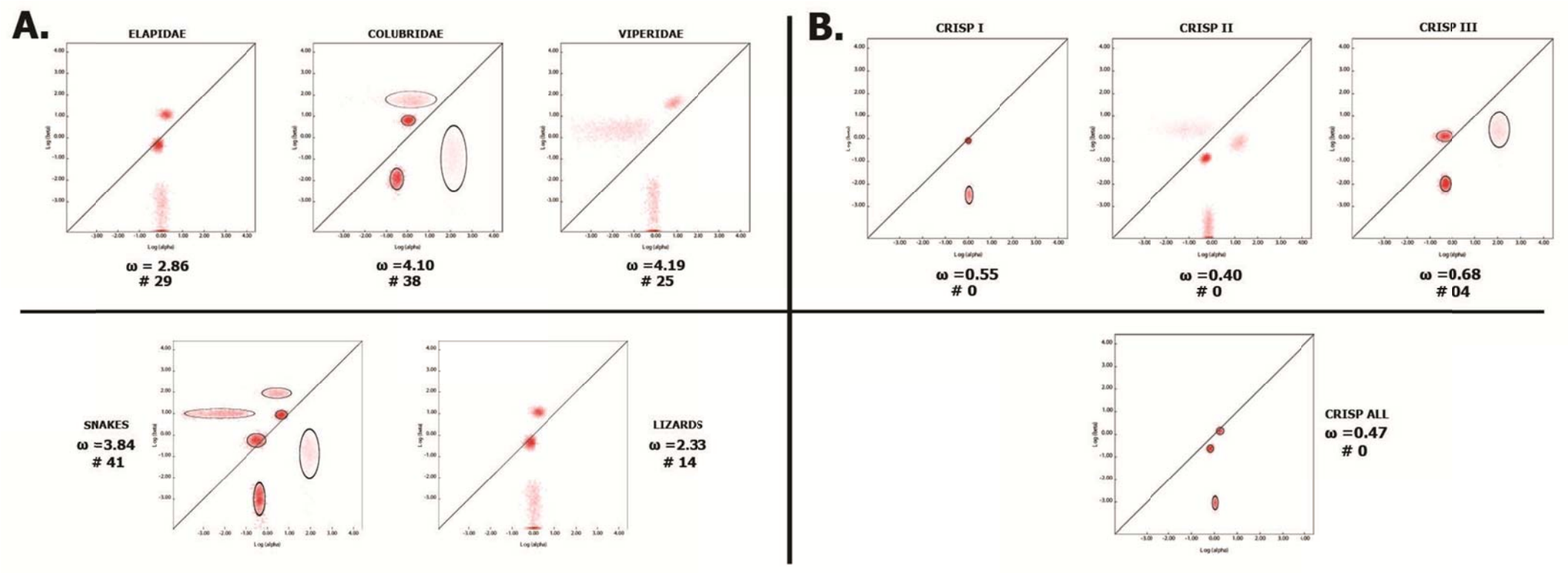

Figure 5. Evolutionary Fingerprint of Reptilian and Mammalian CRISPs.

Estimates of the distribution of synonymous (alpha) and non-synonymous (beta) substitution rates inferred for Reptilian (A) and Mammalian (B) CRISPs. The ellipses reflect a Gaussian-approximated variance in each individual rate estimate, and colored pixels show the density of the posterior sample of the distribution for a given rate. The diagonal line represents the idealized neutral evolution regime (ω = 1), points above and below the line correspond to positive selection (ω>1) and negative selection (ω<1), respectively.