Squamata phylogenomics and molecular evolution of venom proteins in Toxicofera

Masterthesis

in the Master of Science in Molecular Biology and Evolution

of the Faculty of Mathematics and Natural Sciences of the Christian-Albrechts-University of Kiel

submitted by

Kim Joffroy

First examiner: Dr. Julien Dutheil

Second examiner: Prof. Dr. Tal Dagan

Kiel, August 2022

Table of Contents

1. Abstract

How frequent is convergent evolution? This fundamental question of evolutionary biology is challenging to address as it requires mapping innovations on a phylogeny. Phylogeny reconstruction methods, however, aim at reconstructing the tree with the minimum number of such events. Squamata the order of scaled reptiles composed of lizards, snakes, and amphisbaenians offers a striking example of such a conundrum. The Toxicofera hypothesis states that all venomous squamates such as iguanas, anguimorphs, and snakes are a monophyletic group, and that venom evolved only once in their last common ancestor, therefore constituting the only synapomorphy legitimating this group. Morphological and molecular phylogenetics of squamates in particular those of mitochondrial genes, however, result in distinct phylogenies supporting multiple convergent evolution of venomousness also because not all Toxicofera are venomous. Venom is composed of different proteins that are recruited into the venom from their original function after gene duplication. Thus, homologs of venom proteins are also found in non-venomous taxa. Thereby, the composition of Toxicofera venom resembles those of various other taxa which evolved venomousness multiple times convergently. Here, I aim for studying the molecular evolution of two venom proteins by first establishing a phylogenetic framework for the squamates group with a phylogenomic approach that makes use of all protein families in the RefSeq database of the NCBI that are available for at least 15 squamates resulting in a dataset containing 768 protein families for 272 species. I then use the resulting phylogeny to study the molecular evolution of two venom proteins independent of their single-gene phylogenies. I perform selection models of codon sequence evolution to detect variations in selection pressure between venomous and non-venomous clades. Additionally, I expect to find positively selected sites to be fast-evolving surface proteins that are co-adapting. Even though mitochondrial and nuclear phylogenies diverge a lot the results reveal evidence for multiple convergent evolutions of venom in Colubroidea, Anguimorpha, and Iguania. Venom proteins experience positive selection in snakes and anguimorphs but not in iguanas. Among positively selected sites are fast-evolving surface residues that are co-adapting with other residues. I conclude selection pressure acting on venom proteins is stronger in all Toxicofera except for Iguania compared to other squamates. This difference is not necessarily a consequence of heritability but to some extent affected by ecological factors like differences in diet.

2. Introduction

Squamata is the highly diverse order of scaled reptiles containing about 11,000 extant species that emerged in the Middle Triassic 242 mya [1]. The closest living relative of squamates is the tuatara (*Sphenodon punctatus*) which is the only living species of the order Sphenodontia [2,3]. Squamates gain a particular research interest especially because of the medical relevance of snake venom. Nevertheless, besides several well-known clades of geckos, chameleons, iguanas, skinks, monitor lizards, and snakes squamates contain inconspicuous clades like the worm-like dibamids and amphisbaenians or the venomous helodermatids. Although they are highly medically relevant and already contributed to ancient horror movies like “The Giant Gila Monster – 1959”, these species are not frequently recognized among non-specialists. All these clades morphologically differ significantly because of specialized adaptations to their environments or lifestyles. [2]. Additionally, squamates are not only morphologically but also genetically very diverse. Compared to mammals having a genome size of 2.2 to 6.0 Gbp and the tuatara (~5 Gbp) squamates have on average significantly smaller genomes, which range from 1.3 to 2.8 Gbp [3,4]. Even though their genomes are comparatively small, they have the highest genetic diversity of all non-avian reptilian lineages [2]. Therefore integrating these various diverse squamate taxa into a systematic and phylogenetic framework has been everlasting challenging, but undeniable to study the evolution of venomousness.

2.1 Hypotheses regarding the phylogenetic history of squamates

In 2005, with the emergence of genome data, Vidal and Hedges established the to-date best-accepted hypothesis about the phylogeny of squamates and thereby also about the evolution of venomousness. While especially the position of snakes in the phylogeny was unclear before they now sustain a reasonable position in the phylogeny, giving rise to an innovative hypothesis about a single early origin of venomousness in squamate history. The Toxicofera hypothesis states that venomousness evolved only once in the last common ancestor of a monophyletic group named Toxicofera composed of Iguania, Anguimorpha, and Serpentes [5,6]. Since only these three taxa possess toxin-secreting oral glands [7]. Despite the overwhelming support from many other molecular studies [2,5,7–10] morphological studies, as well as some molecular studies, find evidence against the Toxicofera hypothesis [11–18]. A major difficulty regarding a monophyletic venomous clade is the distribution of venomous clades across Toxicofera. Only the most derived and, therefore, youngest snake families are venomous while all more basal snakes like henophidians and blind snakes are non-venomous [19]. Additionally also within venomous snake families are non-venomous species as in Colubridae [20]. Within Anguimorpha varanids and helodermatids are venomous but venomousness in anguids has so far not been studied in detail. For decades iguanas were assumed not to be venomous but only recently venom could be detected in a few species like the bearded dragon and the green iguana [7,21]. Apparently, revealing monophyletic Toxicofera depends a lot on taxon sampling, gene sampling, evolutionary rates, and probably many other factors [8,16,22].

2.1.1 Morphological phylogenies

The first computer-generated morphological phylogeny of squamates was established by Estes et al. in 1988 [11]. A simplified version of this phylogeny which was generated from 130 osteological and 18 soft anatomy characters is displayed in Figure 1A [11]. In this study, the positions of Dibamidae, Serpentes, and Amphisbaenia could not be confidently solved. It is assumed that the limblessness of these clades had a misleading effect in this morphological study [1,11]. Besides, all higher-level taxa are monophyletic with Iguania being the most basal clade, followed by Gekkota. Here Toxicofera is not monophyletic since Anguimorpha and Scincomorpha have a sister relationship.

An approach to refine morphological phylogenies and to overall date back clades is adding data for extinct taxa. In his study from 2008 [15], Conrad added osteological characters of fossil taxa notably Mosasauria which are assumed to be closely related to snakes [10,12,23], and Paramacellodidae an extinct Scincomorpha family [24]. In this phylogeny summarized in Figure 1B the limbless clades of dibamids, snakes and amphisbaenians cluster within Scincomorpha as the sister group to Scincidae. Surprisingly even though Paramacellodidae already were proved to be scincomorphs at that time in this phylogeny they appear within anguimorphs as the sister taxon to all other Anguimorpha [15]. Additionally, also Mosasauria appears within anguimorphs most terminal as the sister taxon to Varanoidea. Also here Toxicofera does not appear to be monophyletic although this study includes fossil taxa and overall more taxa than previous studies [1].

A more recent study from 2018 by Simões et al. [23] summarized in Figure 1C is the first morphological study that finds Gekkota as the most basal squamate clade and not Iguania. Iguanas appear as the sister taxon to a group that in addition to all other toxicoferans contains the extinct Paramacellodidae and Mosasauria but also Scincidae, Dibamidae, and Amphisbaenia which unambiguously do not belong to Toxicofera. Therefore, Toxicofera does not appear to be monophyletic. Nevertheless, this phylogeny is one of the few morphological studies to indicate a close relationship between snakes and venomous anguimorphs respectively Varanidae and Helodermatidae, and between snakes and Mosasauria. However, this study could not solve the clade of the limbless squamates Dibamidae, Amphisbaenia, and Serpentes.

A series of studies that were able to account for the limblessness of dibamids, amphisbaenians, and snakes was performed by Lee et al. from 1998 to 2005 [12,13]. The phylogeny in Figure 1D summarizes the outcomes of these studies. Even though Iguania appear as the most basal squamates again making Toxicofera paraphyletic snakes appear most terminal within anguimorphs indicating a close relationship to venomous monitor lizards and heloderms. This relationship appeared in particular when only soft tissue characters were included [12]. In contrast, Dibamidae and Amphisbaenia appear in a close relationship to geckos.

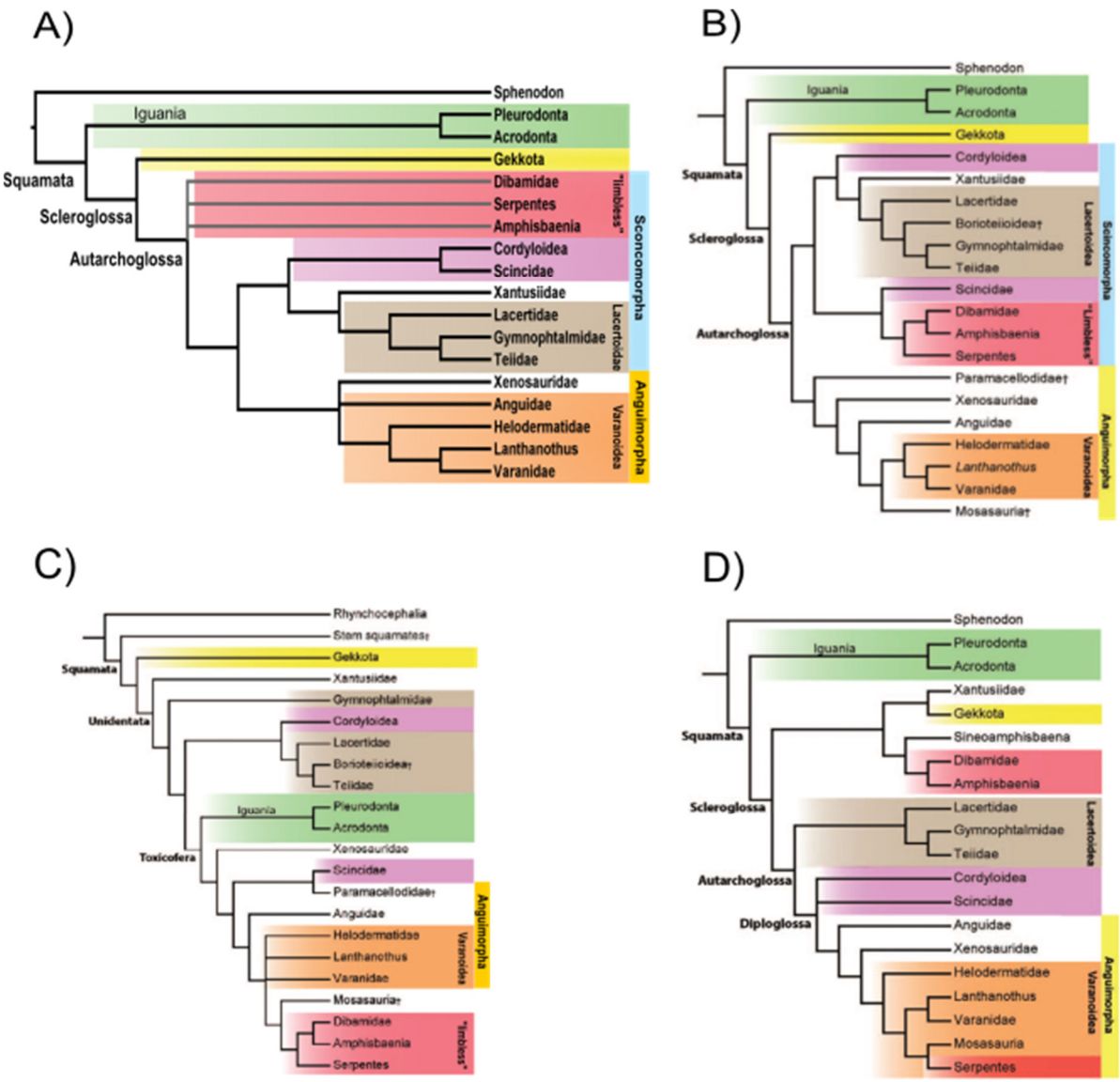

Fig.1: Morphological phylogenies obtained from [1]. A) Summarized and modified from Estes et al.1988 [11] grey branches indicate unsolved relationship, B) Summarized and modified from Conrad et al. 2008 [15], C) Summarized and modified from Simões et al. 2018 [23], D) Summarized and modified from Lee et al. 1998-2005 [12,13].

In summary, morphological phylogenies do overall not support monophyletic Toxicofera even though fossil taxa are included. Additionally, most morphological phylogenies place Iguania as the basal squamates followed by Gekkota. Higher-level taxa like Varanoidea, Lacertoidea, and Iguania always appear to be monophyletic. In contrast, the most variable is the positions of the limbless clades Serpentes, Dibamidae, and Amphisbaenia. Furthermore, the position of Scincoidea remains unsolved in morphological phylogenies since they do not appear to be monophyletic. Most of the incongruence between morphological phylogenies seems to result from an overestimation of limblessness in Dibamidae, Amphisbaenia, and Serpentes which can be overcome by including a higher proportion of soft tissue characters [12].

2.1.2 Molecular phylogenies

The original tree topology from Vidal & Hedges 2005 [5] can be found in Figure 2A. This study is one of the first molecular studies on squamates and simultaneously the origin of the Toxicofera hypothesis. The phylogeny was inferred from merely nine nuclear genes for 18 squamates and *Sphenodon punctatus* as the outgroup [5]. It shows monophyletic Toxicofera most terminal in the phylogeny while dibamids and geckos appear to be the most basal taxa similar to the morphological phylogenies Figure 1 that also suggest a very basal position for geckos. Additionally, molecular studies are able to separate the limbless clades of dibamids, snakes, and amphisbaenians by clustering Amphisbaenia with lacertids. Overall this early molecular study revolutionized the squamate phylogeny but still, the Toxicofera hypothesis is controversial since this study includes only nine nuclear genes and a small number of taxa [25]. Another controversy could be taxon sampling because Vidal & Hedges did not include venomous iguanas or snakes [5,16]. Besides, it was shown that taxon sampling can affect the position of Amphisbaenia in the tree topology in nuclear phylogenies by appearing within Toxicofera [16].

A different study from 2007 by Böhme et al. contrariwise only includes mitochondrial genes [14]. In this study summarized in Figure 2B whole mitochondrial genomes for 32 species (including outgroups) were used to infer the phylogeny [14]. The mitochondrial phylogeny differs clearly from the nuclear phylogeny. Iguania became paraphyletic by splitting into Acrodonta as the sister group of snakes and Iguanidae as the sister taxon of a group composed of Scincidae and Lacertidae. Amphisbaenia however became separated from lacertids clustering with snakes and acrodonts recovering the close relationship between amphisbaenians and snakes hypothesized by morphological studies (Figure 1B-C). These circumstances engender the breakup of the Toxicofera monophyly. The position of Gekkota though stays most basal at the root of Squamata. The effects of mitochondrial genes on squamate phylogeny are frequently discussed [16,17,25,26]. A problem that occurs very often is that mitochondrial genes tend to uncover a sister relationship between snakes and acrodonts [14,16,26] because of potential convergent molecular evolution between snakes and Agamidae [17,26] and high evolutionary rates [22]. The relationship between acrodonts and snakes does probably not occur from taxon sampling since increasing the number of representative species per clade does not break up this branch [16]. A biological explanation for the difference between mitochondrial and nuclear phylogenies could be that mitochondrial genes evolve independently from nuclear genes [27]. Mitochondrial DNA is favored over nuclear DNA for inferring species trees because the mitochondrial tree has a higher probability of tracking short internodes due to its haploidy and therefore smaller effective population size [28].

Studies combining mitochondrial and nuclear data produce a mixture of mitochondrial and nuclear phylogenies. The example in Figure 2C shows the summarized phylogeny from Pyron et al. 2013 which was inferred from seven nuclear and five mitochondrial genes [9]. This study appears to be reasonably balanced when it comes to the proportion of mitochondrial and

nuclear genes and additionally has a higher number of representative species per clade than studies discussed before. Overall the tree topology agrees with Vidal & Hedges 2005 (Figure 1A). This led to a high acceptance of the Toxicofera hypothesis. However, is it still under debate how much the overall small number of genes impacts the phylogeny since increasing gene number is correlated with increasing phylogenetic accuracy [29]. Therefore, in a smaller set of genes, the effects of a single gene on the whole phylogeny are greater than in a set containing many genes. These effects can be long branch attraction, short divergence times, and incomplete lineage sorting [8].

But how does combining morphological and molecular data affect the phylogeny? Figure 2D shows a summarized phylogeny by Reeder et al. 2015 [10]. This study combines a morphological dataset of 691 characters with a molecular dataset of 46 genes for 161 extant and 49 fossil species [10]. The tree topology agrees with molecular studies that support the Toxicofera hypothesis. Extant toxicoferans are monophyletic but additionally, contain two fossil taxa Mosasauria as the sister group of snakes and Polyglyphanodontia as the sister group of iguanas. The close relationship between snakes and mosasaurs has already been suggested by other morphological studies (Figure 1C-D) [10,12,23]. But the sister relationship between Iguania and Polyglyphanodontia is surprising because previous studies suggested they have a closer relationship with lacertids than with iguanas as indicated in the morphological studies in Figures 1B and 1C where Polyglyphanodontia appear with their original name Barioteiioidea within Lacertoidea [10,15,23].

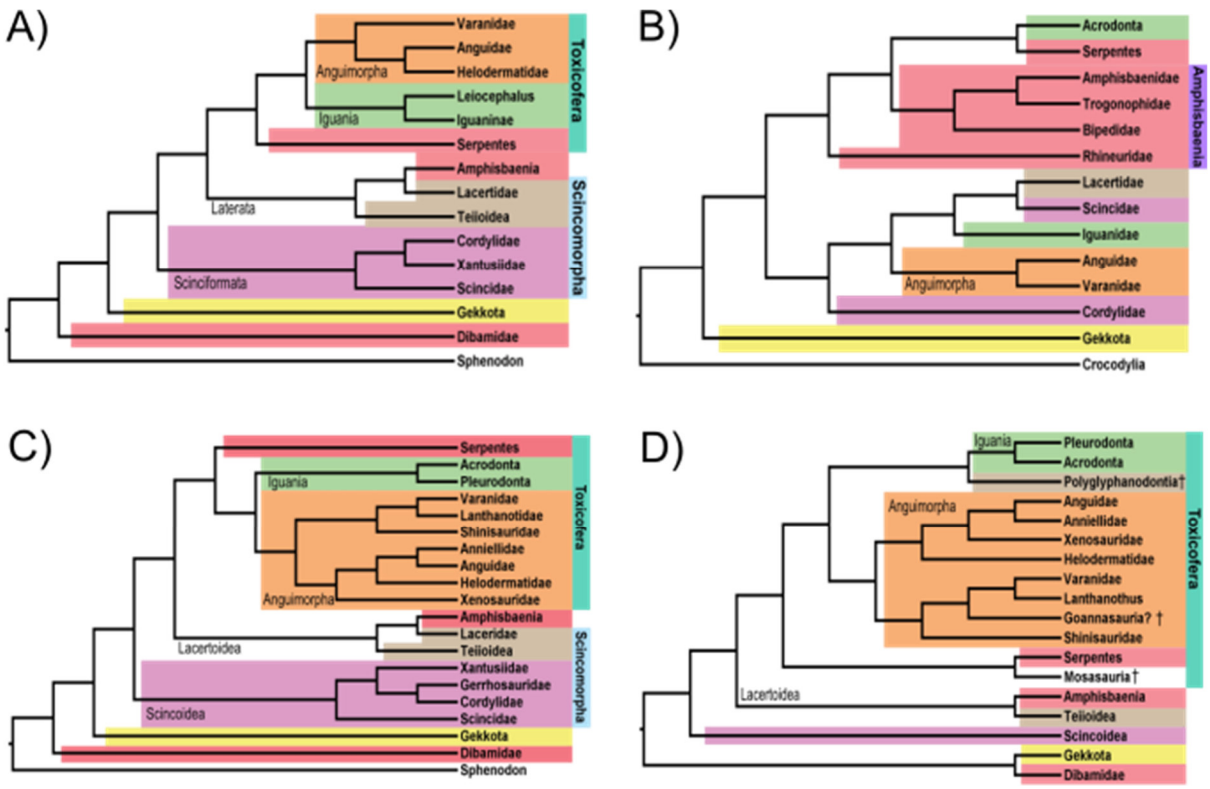

Fig.2: Molecular phylogenies obtained from A) Nuclear genes modified from Vidal & Hedges 2005 [5], B) Mitochondrial genomes modified from Böhme et al. 2007 [14], C) Combined nuclear and mitochondrial genes modified from Pyron et al. 2013 [9], and D) Combined molecular and morphological data modified from Reeder et al. 2015 [10].

To recapitulate, the squamate phylogeny could be considered unresolved because not only do morphological and molecular studies disagree but there are also conflicting hypotheses within molecular studies since phylogenetic accuracy depends on taxon sampling, gene sampling, long branch attraction, short divergence times, incomplete lineage sorting and probably also the genetic divergence between outgroup and ingroup [8,16,18,29]. Therefore especially the monophylies of Toxicofera, Iguania, and Scincoidea highly depend on the data source. Iguania is either monophyletic and in a sister-relationship to Anguimorpha or is split in Acrodonta and Pleurodonta while acrodonts appear in a close relationship to snakes and pleurodonts to Scincidae and Lacertidae. Scincoidea is either monophyletic or split in Scincidae and Cordylidae. Furthermore, the sister-relationship of Lacertidae and Amphisbaenia is incongruent between different data sources. Amphisbaenia is either included in Lacertoidea or closely related to snakes and acrodonts. In contrast, the overall basal position of Gekkota and Dibamidae and the monophyly of Anguimorpha are consistent. It can be beneficial to implement a phylogenomic approach that in contrast to phylogenetics aims for using a major part of the genome to overcome these discrepancies because in a phylogenomic approach difficulties within single genes affects a large-scale phylogeny less than phylogenies that rely on only a few genes and taxa.

2.2 Venomousness in squamates

The venom of Toxicofera is of high medical relevance. However, to tap the full medical potential of toxicoferan venom it is necessary to understand the underlying molecular evolution of the toxins contributing to these venoms. The venom of vipers is used in drugs that are involved in the treatment of for example Angina Pectoris (hypertension), cardiac failure, and acute coronary syndrome [30]. The Gila monster (*Heloderma suspectum*) venom is involved in the treatment of Type II diabetes mellitus and the venom of elapid snakes notably the Chinese cobra (*Naja atra*) is used in chronic arthralgia (joint pain) and sciatica (nerve pain) [30]. But while snake venom is frequently used for medical applications only a few studies on lizard venom have been done [31]. Besides the Gila monster, Varanidae is a second venomous anguimorph family. For a long time, it was suggested that the Komodo dragon (*Varanus komodoensis*) and its relatives kill their prey with the help of toxic bacteria in the saliva but then it was shown that monitor lizards are truly venomous [7,32]. Additionally, it was shown that also some iguana species have the ability to produce venom for example the green iguana (*Iguana iguana*) and the bearded dragon (*Pogona vitticeps*) [7,21]. However, since iguanas are mostly herbivores and insectivores their venom system has not specialized to kill prey, and toxins that are contained in the venom are only expressed at low levels [21]. The morphological evolution of the venom system was described by Fry et al. [7,33]. The venom glands of iguanas are supposed to be in basal condition consisting of protein secreting glands in mandibular (lower jaw) and maxillary (upper jaw) regions [7]. In contrast, snakes have only specialized maxillary venom glands, and anguimorphs only specialized mandibular venom glands [7]. The protein composition of the toxicoferan venom is overall the same, and subject to strong natural selection [34]. The venom consists of very similar proteins to the venoms of amphibians, cephalopods, insects, arachnids, and cone snails [35]. Due to the strong selection pressure in venom proteins adapting to the prey, especially surface residues are under positive selection [35–37]. These sets of positively selected sites tend to be very similar between different snake families [36]. Therefore, they are likely functionally important. Previous studies found that many functionally important sites are co-adapting positions [38]. Prominent examples of toxins in the toxicoferan venom are Phospholipase A2 (PLA_2) complexes, cysteine-rich secretory proteins (CRISP), and tree-finger toxin complexes [35,39,40]. Genes coding for these toxins were recruited into the venom after gene duplication from protein families with physiological functions during squamate history [39,41]. Thereby the phylogenetic history of venomous squamates plays an important role in answering the question when and how often these genes were recruited into the venoms. In the following two important venom proteins analyzed in this study are further examined.

2.2.1 Phospholipase A2

Phospholipase A2 is a superfamily of enzymes that catalyze the hydrolysis of sn-2 ester bonds in phospholipids [42]. These enzymes can be assigned to five main types with squamate toxins belonging to the first type consisting of secreted PLA_2s [42]. The type of secreted PLA_2s is

subdivided into three major groups: Group I – vipers, Group II – elapids, and Group III – lizards, bees, jellyfish, and scorpions [42,43]. Snake venom PLA_2s have many different pharmacological effects: Neurotoxicity, Myotoxicity, Cardiotoxicity, and anticoagulant effects [44] to name only a few of them. Additionally, PLA_2s of snakes, heloderms, and varanids are able to inhibit platelet aggregation [45]. Non-venomous homologs of these toxins play a role in the digestion of phospholipids in the stomach, signal transduction, and host defense [42,43]. It is assumed venom PLA_2s undergo rapid evolution after gene duplication due to changes in expression levels that alter the strength of selection pressure and thereby drive functional diversification [43,46]. This mechanism can be triggered by the need to respond to changes in the diet, therefore similarities in selection pressure between different clades can be justified by phylogenetic history or diet [46]. Group III genes must have been recruited several times independently into the venoms of lizards, bees, jellyfish, and scorpions [43]. Their overall structure is displayed in Figure 3 which shows the tertiary structure of the human group III PLA_2 [47]. The protein structure consists of three alpha helices, two beta sheets, an extended C-terminus, and a calcium-binding loop at the N-terminus. The active site is highly conserved across all lineages comprised of a catalytic histidine followed by an aspartic acid [47]. Only the third position of the active site is more or less variable while a tyrosine takes that place in 28 out of 50 analyzed sequences [47]. Most rapidly evolving residues appear at the protein surface because they interact directly with target molecules [46].

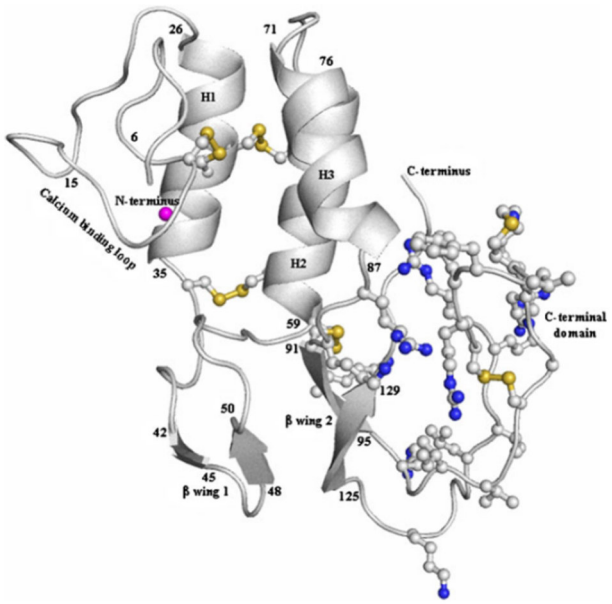

Fig.3: Protein structure of human group III PLA_2 from Hariprasad et al. 2013 [47]

2.2.2 Cysteine-rich secretory protein

Cysteine-rich secretory proteins (CRISP) are glycoproteins that were primarily found in the epididymis of mammals [37,48]. In mammals, they are involved in gamete fusion, cancer, and immune response [48,49]. Additionally, they were found in many other clades of the animal kingdom. In *C. elegans* they contribute to longevity and stress resistance and in snails, they act as a sperm chemoattractant [48]. Furthermore, they act as toxins in the venom glands of snakes, lizards, arachnids, and cone snails [49]. Venom CRISPs of Toxicofera cause paralysis of smooth muscles and thereby block smooth muscle contraction [32,48]. Besides, they are able to induce hypothermia by blocking ion channels and in snake venom, they also inhibit the growth of new blood vessels, promote inflammatory response, and can increase vascular permeability in the prey [32,49]. Previous studies found that toxicoferan CRISPs must have undergone strong positive selection which is stronger in snakes than in lizards [37,49]. Non-venomous homologs in mammals however experience strong purifying selection [49]. In general, CRISPs are single chained polypeptides that contain 16 conserved cysteines of which 10 are clustered in the C-terminal third [48]. They comprise a PR-1 domain at the N-terminus which is built of five alpha helices and eight beta sheets and contains five cysteines and a cysteine-rich domain at the C-terminus which is built of a few short helices and three disulfide bonds and a hinge region that connects the two domains (Figure 4) [49]. So far the active site could not be confidently predicted [48]. Sites that are interacting with other molecules are located at the protein surface [37]. Because these sites play a functional role they tend to be under positive selection [37].

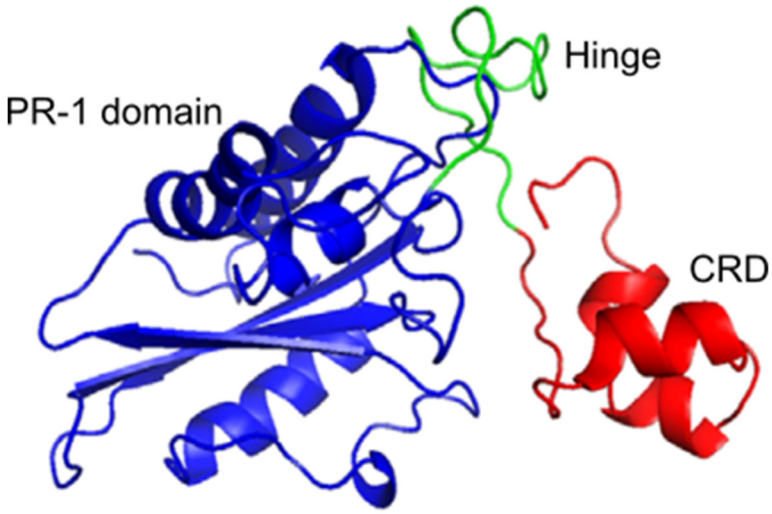

Fig.4: Venom CRISP protein structure of *Naja atra* modified from Wang et al. 2006 [50].

In summary, venom proteins were once proteins with physiological functions that were recruited into the venom after gene duplication. Phospholipase A_2 and Cysteine-rich secretory proteins contribute to venoms all over the animal kingdom and are present in the venoms of all venomous toxicoferans snakes, anguimorphs, and iguanas. Previous studies suggest these proteins underwent rapid molecular evolution in Toxicofera promoted by strong positive selection especially in surface residues since they interact with the target molecules and are therefore essential for the toxin function. Furthermore, functionally important residues tend to coevolve.

3. Objectives

Here I aim for studying the molecular evolution of venomousness in squamates by implementing two approaches that address the origin of venomousness and the molecular evolution of venom proteins within Toxicofera. To study the origin of venomousness in squamates phylogenomic analyses are conducted to account for uncertainties regarding low gene and species numbers. Additionally, the impact of gene sources needs to be examined. To study the molecular evolution of venom proteins population genomic methods are used to analyze selection pressure and coevolution acting on two different proteins phospholipase A_2 and venom cysteine-rich secretory protein

3.1 Hypotheses

i) Since venoms consisting of overall the same proteins evolved multiple times independently across the animal kingdom and because of the absence of venom in major toxicoferan lineages like basal snakes and anguids, and because of different conflicting phylogenetic hypotheses between different data sources multiple convergent origins of venomousness within the squamate phylogeny are likely.

ii) Venomous toxicoferans have undergone positive selection because sites that interact with molecules in the prey are essential for the toxin function and are therefore under positive selection.

iii) Positively selected sites are fast-evolving surface residues that constantly need to adapt to the target molecule of the prey.

iv) Furthermore, the same set of sites is assumed to be under positive selection in several clades because they are necessary for the toxin function. Therefore positively selected sites do not evolve independently but are co-adapting.

3.2 Methodological approach

To analyze the origin of venomousness on a large scale that accounts for gene and taxon sampling a phylogenomic approach is used that includes as many taxa and genes as possible to minimize the risk of overestimating single squamate lineages or genes. Additionally, it is

necessary to include genes from different sources notably mitochondrial and nuclear genes to account for the effects of independent evolution and to increase the probability of tracking short internodes [28]. Obtaining a reliable maximum-likelihood phylogeny depends on extensive substitution models followed by elaborate bootstrapping. Tracking differences between mitochondrial and nuclear phylogenies is achieved by performing separate phylogenetic analyses on subsets of the original dataset that include only mitochondrial and only nuclear genes. Selection pressure acting on venom proteins in venomous squamates is estimated by performing selection models of codon sequence evolution that analyze the ratio of substitution rates at non-synonymous and synonymous sites (dN/dS). The dN/dS ratio is estimated by counting the synonymous and non-synonymous substitutions between two sequences [51]. Thereby a dN/dS above 1 indicates positive selection while a dN/dS below 1 indicates purifying selection and a dN/dS around 1 represents neutral evolution [52]. Because selection models of codon sequence evolution depend on a reliable phylogeny the large-scale phylogeny inferred previously is used to not run the risk of using a gene tree that does not conform with the species tree since gene trees tend to vary from species trees due to for example incomplete lineage sorting [8]. To figure out if positively selected sites are indeed located at the protein surface the protein structure has to be predicted for all clades that yield sites that are significantly under positive selection. Furthermore, to find surface residues, the relative solvent accessibility which indicates how much a residue is exposed to the protein surface and evolutionary rates need to be examined. Finally, coevolving groups of sites and evolutionary rates per site are predicted to examine whether positively selected sites are co-adapting under high evolutionary rates.

4. Material and Methods

4.1 Phylogenetic inference

4.1.1 Dataset creation:

This study aims to infer the squamate phylogeny from as many proteins and for as many species as possible to obtain a reliable phylogeny that can be used in studies on individual venom proteins for various subsets of the phylogeny. Therefore, in the first step, all protein sequences were retrieved that were available for squamates on the NCBI's RefSeq database. These 587102 sequences were sorted into protein families making use of SiLiX a software package that clusters sequences from similarity networks with help of the single-linkage clustering approach [53]. The parameters sequence identity and alignment coverage were examined by investigating all possible pairs of values in 0.1 steps to find the optimal thresholds for this dataset. Afterward, all 81 resulting datasets were analyzed with an R script that calculates all necessary properties of the datasets e.g. the number of protein families, the mean number of sequences per family, and the sum of sequences in families that contain at least four non-duplicated sequences (Supplementary Table 1). Additionally, the calculated properties from the table were evaluated in heatmaps to find the sequence identity and alignment coverage that

result in the highest number of protein families containing at least four non-duplicated sequences and the highest proportion of the total number of sequences in families containing at least four non-duplicated sequences over the total number of sequences in families containing at least four sequences. Furthermore, the optimal dataset with a sequence identity of 0.7 and an alignment coverage of 0.9 was filtered in R by first removing paralog sequences and extracting only protein families containing sequences for at least 15 squamate species, afterward. The final dataset consists of 768 protein families for 272 squamate species (Supplementary Tables 2 and 3).

4.1.2 Phylogenomics:

To address the uncertainty of the data sequences were aligned with two distinct methods MUSCLE [54] and ClustalO [55]. Afterward, a consensus alignment was established by retaining only sites that match both alignments using the BppAlnScore program of the Bio++ Program Suite software compares two alignments and calculates the column scores [56]. Finally, the alignments were concatenated into a single data. The maximum-likelihood phylogeny for the full dataset was retrieved using RAxML [57]. 500 bootstrap replicates were performed. A bootstrap convergence test implemented in RAxML was performed. This test aims to find the optimal number of bootstrap replicates, by splitting the set of replicates in half after every 50 replicates to test if support values differ significantly between the sets, if they do not, adding more bootstrap replicates will most likely not change the support values [58]. The performed bootstrapping test converged after 450 trees for a cutoff of 0.02. Bootstrap support values of Transfer Bootstrap Expectation (TBE) [59] and Felsenstein Bootstrap Proportion (FBP) [60] were mapped onto the phylogeny. At last, the phylogeny was rooted using the minimal ancestor deviation method (MAD) [61].

The same procedure was performed on subsets of the dataset containing only mitochondrial protein families and only nuclear protein families. Subsets were generated using R by filtering according to protein names. The resulting mitochondrial dataset consists of 36 protein families for 264 squamate species with an alignment size of 10131bp. The resulting nuclear dataset consists of 732 protein families for 16 squamate species with an alignment site of 293130bp. Maximum-likelihood phylogenies for both subsets of the superalignment were generated using RAxML. For the mitochondrial phylogeny, 500 bootstrap replicates were performed which converged at a cutoff of 0.015. For the nuclear phylogeny, 100 bootstrap replicates were performed. These converged already after 50 trees at a cutoff of 0.01. TBE and FBP support values were mapped onto both phylogenies before they were rooted with MAD. The phylogenies were visualized and colored in FigTree (http://tree.bio.ed.ac.uk/).

These analyses were conducted using R scripts available at https://gitlab.gwdg.de/joffroy/squamata_phylogenomics/-/tree/main/R-Scripts, using the packages 'ape 5.5' [62], 'dplyr 1.0.7' [63], and 'purrr 0.3.4' [64]. Graphical representations were done using the 'ggplot2 3.3.6' package [65].

4.2 Phospholipase A_2

4.2.1 Dataset creation and phylogeny inference:

The dataset for the venom protein phospholipase A_2 (PLA_2) was initiated by compiling the results of four blast searches to cover all clades of venomous squamates. Each one tBLASTx search against the GenBank nucleotide database was performed for Viperidae (*Daboia russelii*, DQ365975), Elapidae (*Naja atra*, AM492700.1), Helodermatidae (*Heloderma suspectum*, EU790968.1), and Varanidae (*Varanus komodoensis*, EU195460.1). To decrease the number of blast results for the two snake searches the e-value was set to 1e-10. Contrariwise to not decrease the number of blast results too much for the two anguimorphs the e-value was set to 1e-06. These four datasets were combined into a single file containing 703 sequences for 44 squamate species. Sequences that were retrieved in more than one blast search were removed using the helper tool "Remove Duplicates from a Fasta File" from "sRNAtoolbox" (https://arn.ugr.es/srnatoolbox/helper/removedup/, 05.2022). Afterward, the remaining sequences were filtered for the species list of the large-scale phylogeny (Supplementary Table 2) using the 'seqinr 4.2.8' package in R [66]. Additionally, four sequences containing unresolved nucleotides had to be removed manually resulting in a dataset of 462 sequences for 44 species. To filter the data for paralogs I used SiLiX with sequence identity and alignment coverage of 0.1 to obtain a final dataset that contains sequences for the venomous families of Elapidae, Viperidae, Helodermatidae, and Varanidae. SiLiX suggested three different protein families that have at least five sequences. I decided on the only protein family containing a sequence for *Heloderma suspectum* that contains 34 sequences for 17 species to generate a quick PhyML phylogeny in SeaView [67] using default parameters to filter the remaining paralogs. Since all paralogs clustered in a separate branch of the phylogeny the paralog branch was removed from the dataset to receive a single sequence for each of the 17 species. The remaining four duplicated species were transcript variants and therefore the shorter sequences were removed manually after comparison of their NCBI entries. The final set of squamate species used here and their corresponding taxonomy can be found in Supplementary Table 4.

The previously inferred phylogeny was used in subsequent analyses on this dataset, restricting to the 17 selected squamate species in the alignment. Sequence and phylogeny processing were performed using R scripts, with the following packages: 'seqinr', 'ape', and 'castor 1.6.9' [68].

4.2.2 Selection models of codon sequence evolution

In order to prepare the dataset for positive selection analyses, the coding regions (CDS) of the sequences were extracted using a Python script available at https://gitlab.gwdg.de/joffroy/squamata_phylogenomics/-/tree/main/R-Scripts that uses the module Biopython [69]. Afterward, the sequences were aligned using the PRANK algorithm [70] at the codon level on the GUIDANCE2 2.02 web server (http://guidance.tau.ac.il/, 05.2022) [71]. Ambiguous sites were filtered using gBlocks [72] implemented in SeaView resulting in an alignment length of 729bp. Using this alignment and the subset phylogeny primarily branch models were performed with codeml in PAML [73] to get an overview of the average dN/dS ratios for various squamate clades at three different taxonomic levels. Here, the dN/dS ratios were also calculated for the roots of monophyletic groups in the branch models to test if a possible signal comes from the root or arises within a clade. In two additional models, I analyzed the difference in dN/dS for different squamate clades according to their venomousness regardless of their systematic classification, since not all toxicoferans are venomous. To enable the dN/dS to vary not only between clades but also between sites I performed 16 different branch-site models at different taxonomic levels in PAML (Figure 5). Since Branch-Site analyses can only deal with two different clades, namely foreground and background lineages I created subsets of the alignment and phylogeny according to each model containing a non-venomous group as the background lineage and a venomous group as the foreground lineage (Fig. 5). Also here groups were tested according to their venomousness regardless of their systematic classification. To test for the significance of positive selection the Branch-Site test of positive selection was performed for each of the models [74]. While the alternative hypothesis of this test allows ω_2 to take values above 1 in foreground branches enabling positive selection and suggesting purifying selection or neutral evolution for background branches the null hypothesis fixes ω_2 at 1 for foreground branches assuming neutral evolution but still allowing sites to experience negative selection in the background branches. P-values were calculated with the Chi-square test in R from the log-likelihoods of the corresponding pairs of Branch-Site results to reject the null hypothesis. Finally, sites that are significantly under positive selection in the alternative hypothesis were extracted using the Bayes Empirical Bayes (BEB) approach that calculates the posterior probabilities for sites to come from the site class allowing dN/dS above 1 [75] which is implemented in PAML.

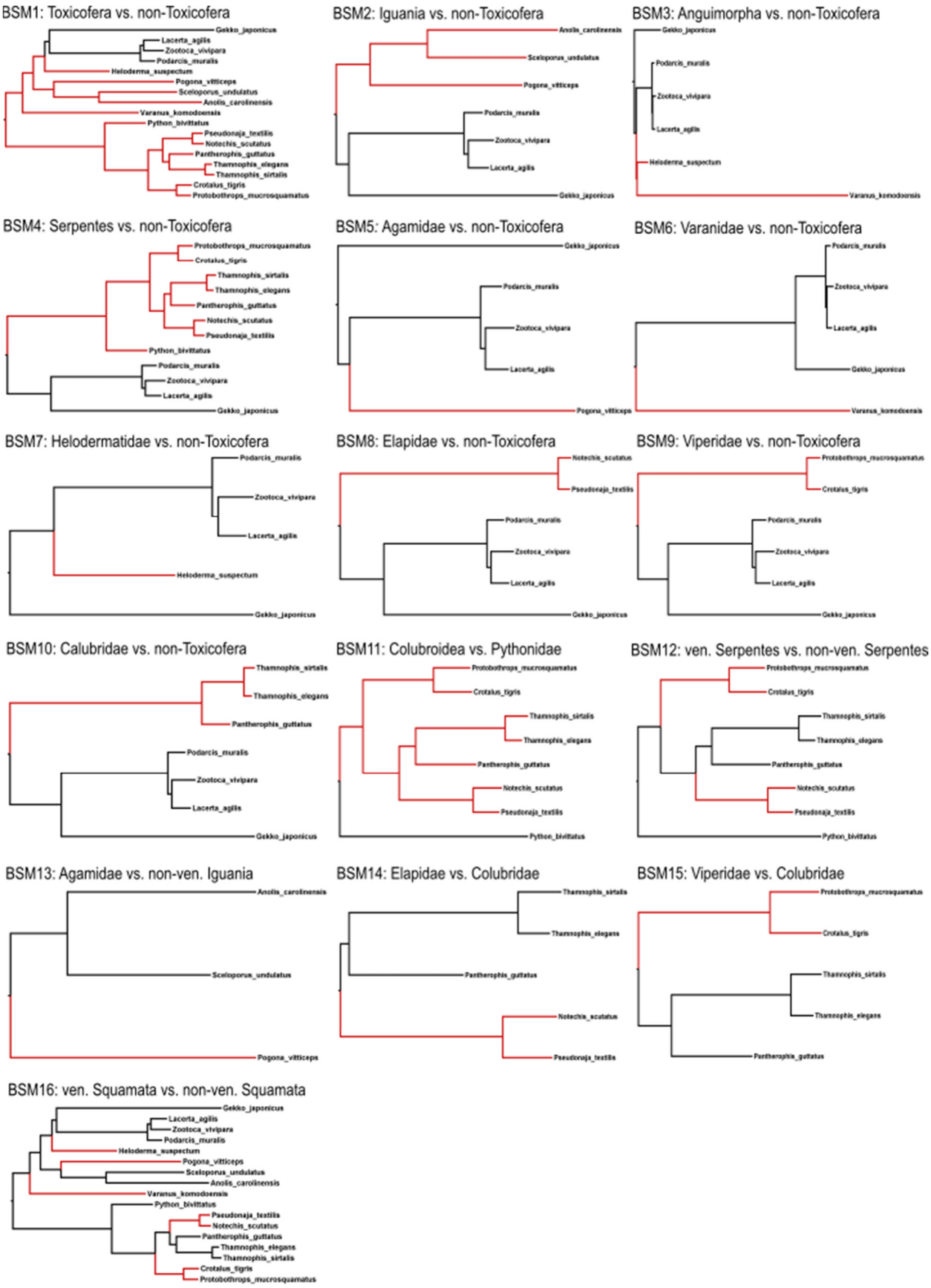

Fig.5: Subsets of the PLA_2 phylogeny for 16 Branch-Site models showing the relationships between foreground lineages (red) and background lineages (black).

4.2.3 Protein structure prediction and visualization:

Sites that are significantly under positive selection in *Varanus komodoensis* were tracked in the protein structure retrieved from the full PLA_2 sequences of *V. komodoensis* (XM_044444354.1) using a simplified version of AlphaFold [76]. To identify the matching residues in the structure of the positively selected sites in the alignment a Python script was used that aligns the sequences from the PDB file retrieved with AlphaFold on the alignment to translate the coordinates accordingly. Corresponding residues were then visualized in VMD [77]. To receive further information about the location of these sites in the protein structure the Relative Solvent Accessibility (RSA) was calculated with another Python script that runs a local version of DSSP [78] with the dedicated Biopython module. Python scripts available at https://github.com/jydu/sgedtools.

4.2.4 Coevolution analyses:

Coevolving groups of sites and evolutionary rates were retrieved using CoMap a software that can compute coevolution statistics and cluster sites [79]. Afterward, coevolving groups were visualized using the PLA_2 protein structure of *Thamnophis elegans* (XM_032238399) that was predicted using the simplified version of AlphaFold and displayed in VMD. Before visualization sites of coevolving groups were translated from alignment positions to residue positions in the structure by using a Python script of sgedtools. Evolutionary rates (PR = posterior rate) were taken from the info file of the CoMap output.

4.3 Cysteine-rich secretory protein (CRISP)

4.3.1 Dataset creation and phylogeny inference:

The dataset for the venom cysteine-rich secretory protein consists of two tBLASTx searches against the GenBank nucleotide database. Blast searches were performed for Varanidae (*Varanus komodoensis*, EU195455) and Helodermatidae (*Heloderma suspectum*, EU790958) both setting the e-Value to 1e-10. These two datasets were combined, and coding regions were extracted using a Python script. After removing duplicated sequences with the helper tool “Remove Duplicates from a Fasta File” from “sRNAtoolbox” remaining sequences were filtered according to the species list of the large-scale phylogeny (Supplementary Table 2) in R resulting in a dataset of 272 sequences for 32 species of the list. These sequences were sorted into 5 protein families that contain at least 5 sequences by SiLiX with default parameters of 0.35 sequence identity and 0.8 alignment coverage. The most appropriate protein family consists of 120 sequences for 32 squamate species. After discarding 15 obvious transcript variants by comparing NCBI entries and leaving only the longest variant four sequences containing internal stop codons and four duplicated sequences were removed manually. Unlike the PLA_2 dataset here the remaining paralogs did not appear in a distinct branch of the fast PhyML tree, therefore a gene tree was inferred with Maximum-likelihood in RAxML for the remaining 97 sequences. 750 bootstrap replicates were generated, and TBE support values were mapped on the

phylogeny. Finally, the resulting gene tree was reconciled with the corresponding subset of the large-scale phylogeny for the 32 species from the alignment. Reconciliation was performed with Treerecs [80] implemented in SeaView using a Branch support threshold of 60%. Taxonomic classification of the squamate species used here can be found in Supplementary Table 4.

4.3.2 Selection models of codon sequence evolution:

To perform positive selection analyses in PAML the dataset was aligned at the codon level using the PRANK algorithm implemented in the GUIDANCE2 web server. Ambiguous sites were filtered with gBlocks in SeaView resulting in an alignment length of 519bp. This alignment and the reconciled phylogeny were used to perform branch models with codeml in PAML to get an overview of the average dN/dS for various squamate clades at three different taxonomic levels and non-taxonomic groups to detect possible differences between venomous and non-venomous species regardless of their systematic classification. In this case, it was not possible to calculate the average dN/dS at the roots of different squamate clades because they were not monophyletic due to several paralogs and multifurcations in the reconciled phylogeny. Additionally, I performed 17 different branch-site models with codeml in PAML on this dataset to find potential sites under positive selection in different clades (Figure 6). The additional model compared to PLA_2 analyses is a model that accounts for venomous and non-venomous Colubridae, since *Rhabdophis tigrinus* is a venomous colubrid species. Also here I created subsets of the alignment and phylogeny according to each model containing a non-venomous group as the background lineage and a venomous group as the foreground lineage. To test for significance I ran the Branch-Site analyses under the alternative and Null-hypothesis and calculated the p-Values performing the Chi-Square test on the log-likelihoods in R.

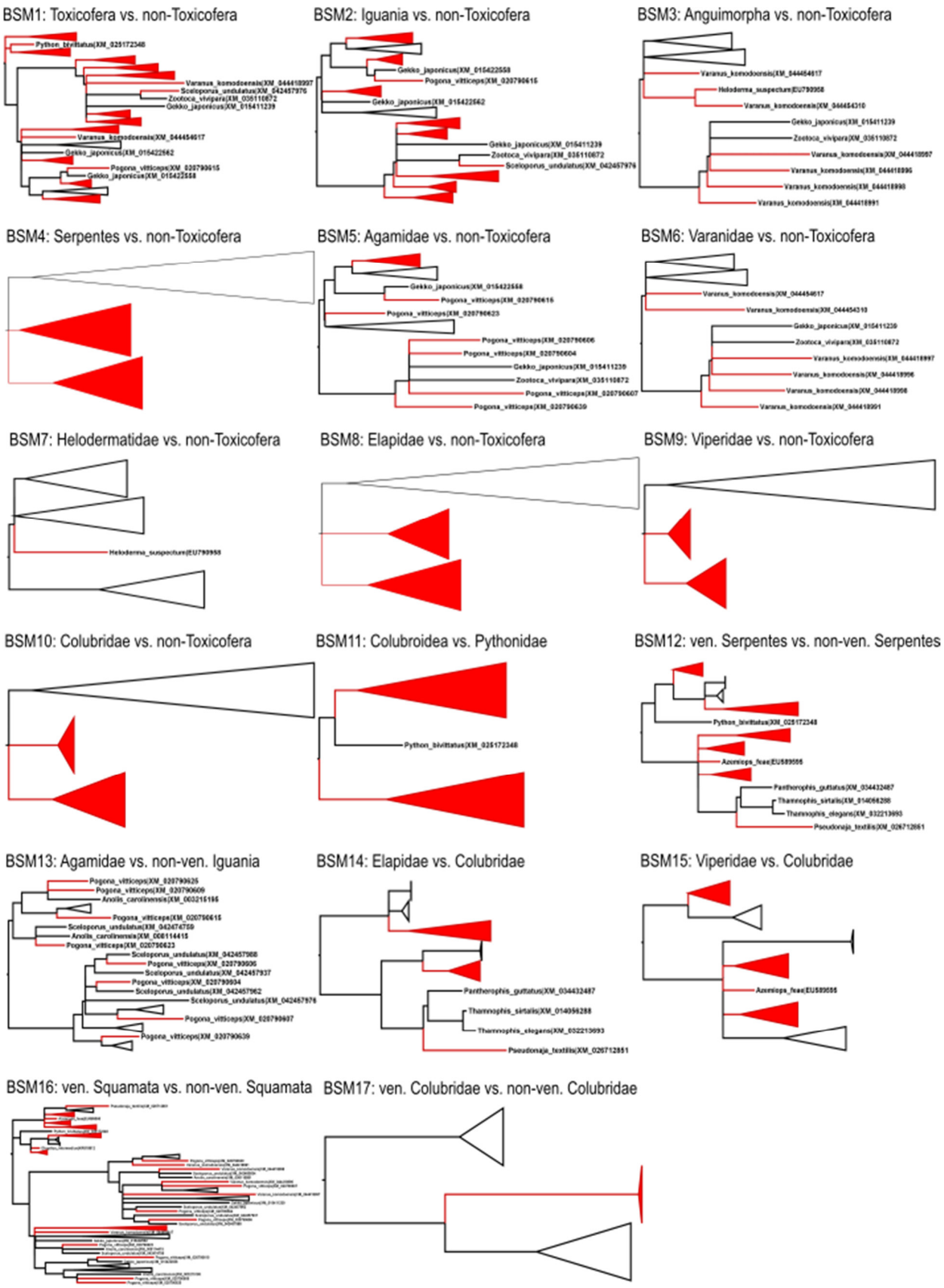

Fig.6: Subsets of the CRISP phylogeny for 17 Branch-Site models showing the relationships between foreground lineages (red) and background lineages (black). Monophyletic groups of foreground/background branches were collapsed.

4.3.3 Protein structure prediction and visualization:

Since the AlphaFold approach used for the PLA_2 dataset could not predict the protein structures for venom CRISPs structures were predicted using the protein structure prediction service Robetta (https://robetta.bakerlab.org/, 05.2022) that uses the modeling method RoseTTAFold [81]. To cover all squamate families that exhibit sites that are significantly under positive selection I used each one representative species to predict protein structure for Helodermatidae, Elapidae, Viperidae, and Colubridae. Structures were predicted using full CRISP protein sequences of *Heloderma suspectum* (EU790958), *Pseudonaja textilis* (XM_026712848), *Ovophis okinavensis* (AB848276), and *Pantherophis guttatus* (XM_034404849). Positively selected sites were visualized in the structures in VMD after translating alignment positions into structure residues using a Python script from sgedtools. Additionally, the RSA for these sites was calculated with a Python script from sgedtools.

4.3.4 Coevolution analyses:

Coevolution analyses were performed with CoMap and resulting coevolving groups were visualized in the protein structure of *Pantherophis guttatus* using VMD. Finally, evolutionary rates retrieved with CoMap have been plotted in histograms in R and positively selected sites and coevolving groups were highlighted.

5. Results

Here I aim for studying the evolution of venomousness in squamates. First, a dataset containing all protein sequences of the NCBI's RefSeq protein database that are available for at least 15 squamate species is generated. The resulting dataset containing 768 protein families for 272 squamate species is then used to reconstruct a large-scale maximum-likelihood phylogeny. This phylogeny is furthermore used on one hand to study differences in the tree topology according to mitochondrial and nuclear data and to evaluate scenarios of multiple convergent evolution of venomousness in squamates. On the other hand, it serves as the underlying phylogeny for analyses of the molecular evolution of two venom proteins since gene trees often deviate from the species tree. Sequences of two venom proteins notably Phospholipase A_2 and Cysteine-rich secretory protein are retrieved for all available species from the list (Supplementary Table 1) to first analyze models of codon sequence evolution to detect signals of positive selection in venomous squamates. Afterward, sites under positive selection are visualized in the protein structures and the relative solvent accessibility (RSA) is calculated to find surface proteins that might be interacting with target molecules in the prey. Evolutionary rates are studied to find evidence for adaptive evolution in surface residues. Finally, groups of coevolving sites are predicted to detect co-adaptation between positively selected sites.

5.1 Receiving the dataset with as many large protein families as possible

Homologous gene families contain paralogs. Paralog sequences of the same species in the dataset can cause a problem because they might be more similar to the sequences of a different species than to those of the same species. Protein family reconstruction relies on sequence similarity, which is defined by the percent identity of a certain region length. Strict criteria (high identity on a long region) will tend to assign paralogs to different protein families which increases the number of families but decreases the number of sequences per family. Conversely, less strict criteria (low identity and short region) will assign paralogs to the same family increasing the family size but decreasing the number of families. The goal is to find the optimal thresholds that return the highest number of protein families with as many sequences as possible in each family. To find these optimal thresholds datasets for all possible combinations of these two parameters are generated. The distribution of data shows that increasing sequence identity and alignment coverage increases the mean proportion of non-duplicated sequences per family over the total number of sequences per family (Supplementary Table 1). The proportion of the mean number of non-duplicated sequences per family over the mean number of sequences per family behaves similarly with a maximum at a sequence identity of 0.9 and alignment coverage of 0.9 (Supplementary Table 1). However, on average the mean of proportions is lower than the proportion of means which indicates an extremely skewed distribution of sequences. Indeed a lot of protein families have a very low proportion of non-duplicated sequences over the total number of sequences. Figure 7 shows the distribution of the number of protein families with at least four non-duplicated sequences across the datasets. It is important to restrict to protein families that have at least four non-duplicated sequences since a phylogenetic signal can only arise if at least four sequences are involved. With in total of 14359 families having at least four non-duplicated sequences the dataset having a sequence identity of 0.7 and an alignment coverage of 0.9 has the highest number of families. The number of protein families increases as identity and coverage increase but starts to decrease at an identity that is higher than 0.7. Figure 8 displays the distribution of proportions of the number of sequences in protein families that contain at least four non-duplicated sequences over the number of sequences in protein families with at least four sequences. This proportion is at 0.25 highest at an identity of 0.7 and coverage of 0.9 and decreases at identities higher than 0.7. For these reasons in the end the favored dataset has a sequence identity of 0.7 and an alignment coverage of 0.9. Removing all paralog sequences and extracting only protein families containing sequences for at least 15 squamate species results in the final dataset consisting of 768 protein families for 272 squamate species (Supplementary Tables 2 and 3). In the next step, the species phylogeny is inferred from this dataset.

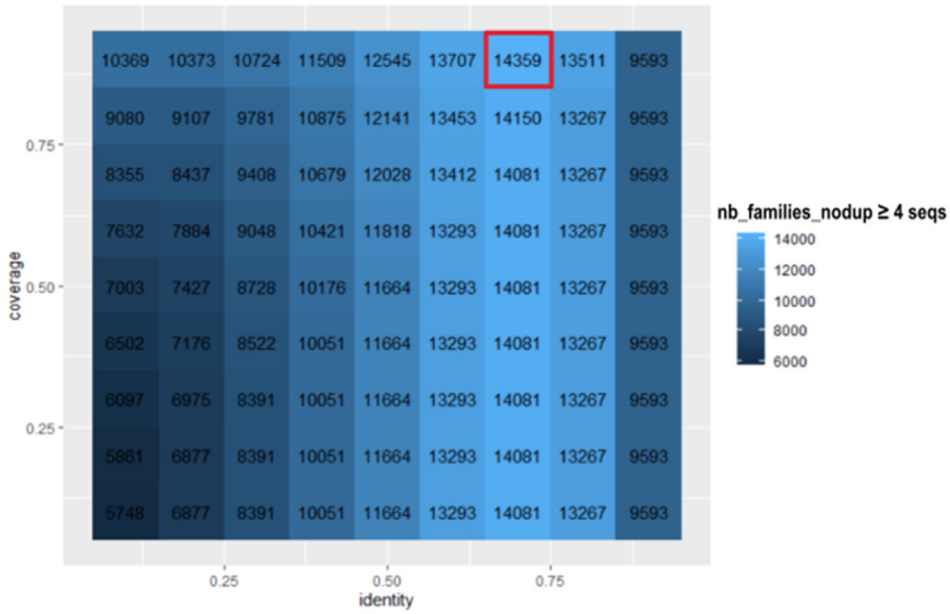

Fig.7: Heatmap showing the distribution of the number of protein families containing at least four non-duplicated sequences. The red rectangle indicates the maximum.

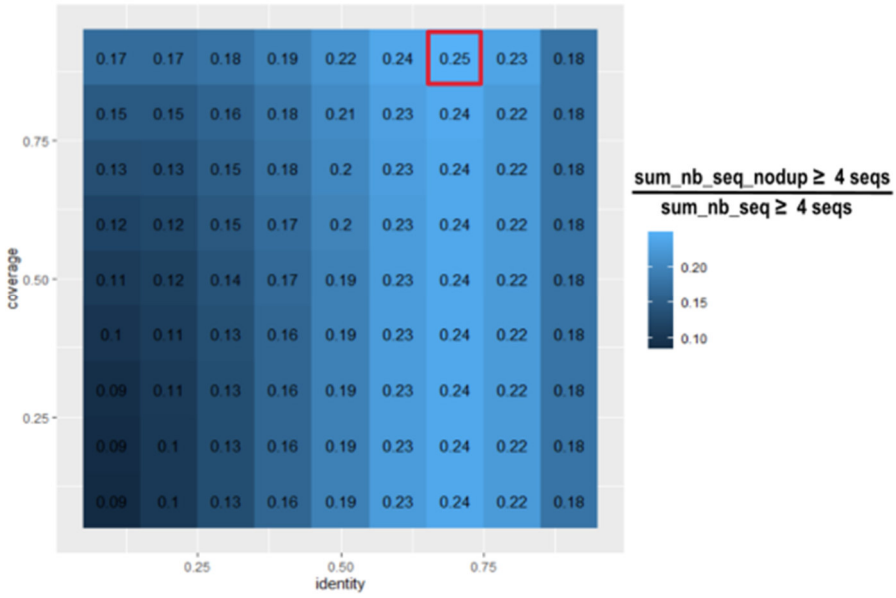

Fig.8: Heatmap showing the distribution of the number of non-duplicated sequences in protein families containing at least four non-duplicated sequences over the number of sequences in protein families containing at least four sequences. The red rectangle indicates the maximum.

5.2 Squamate phylogeny inference

The phylogeny inferred from the optimal dataset is analyzed according to higher-level relationships in squamates and the impact of different data sources notably mitochondrial and nuclear data on the tree topology. Furthermore, scenarios of multiple convergent evolution of venomousness are evaluated also concerning different data sources.

5.2.1 The large-scale phylogeny is well supported but indicates polyphyletic Toxicofera

The large-scale phylogeny inferred from this dataset is summarized in Figure 9. The root in this phylogeny is placed between snakes and all other squamates. The next branching clades are Varanidae followed by a group of iguanas that contains acrodonts as the sister taxon to two non-acrodont iguanas *Anolis carolinensis* and *Sceloporus undulatus*. The following node branches into Teiioidea and the remaining squamates. Then all remaining anguimorphs branch off with Shinisauridae being most basal in this clade and Anguidae and Helodermatidae being sister taxa. Geckos are the sister taxon to a group of Scincoidea, Pleurodonta, Amphisbaenia, and Lacertidae while Scincoidea is the most basal of these clades. Pleurodonta is the sister group of amphisbaenians and lacertids. Toxicofera is not monophyletic in this phylogeny since the root was placed between Serpentes and all other squamates. Additionally, none of the toxicoferan clades is monophyletic except for snakes because Varanidae is separated from all other anguimorphs and pleurodonts from the other iguanas. Another clade that appears to be polyphyletic in this phylogeny is Lacertoidea because instead of clustering with Teiioidea Lacertidae are placed more terminal with amphisbaenians and pleurodonts. The only higher-level monophyletic group is Scincoidea. Finally, also the limbless clade consisting of snakes and amphisbaenians is polyphyletic. The statistic of the minimal ancestor deviation (MAD) method which was used to root the phylogeny yield the root was placed at this position in 19.6% of the trees which is a reasonable amount according to the literature [61]. Overall deep nodes are very well supported with FBPs of over 0.95 except for the node branching into Gekkota and its sister group which has an FBP of circa 0.9. All more terminal nodes are less supported with most FBPs being less than 0.95 except for the sister relationship between acrodonts and the clade comprised of *Anolis carolinensis* and *Sceloporus undulatus* which has high support of 0.99. Even though all other nodes are less supported their support values still range between 0.6 and 0.7. How many convergent origins of venomousness in squamates do the higher-level relationships suggest?

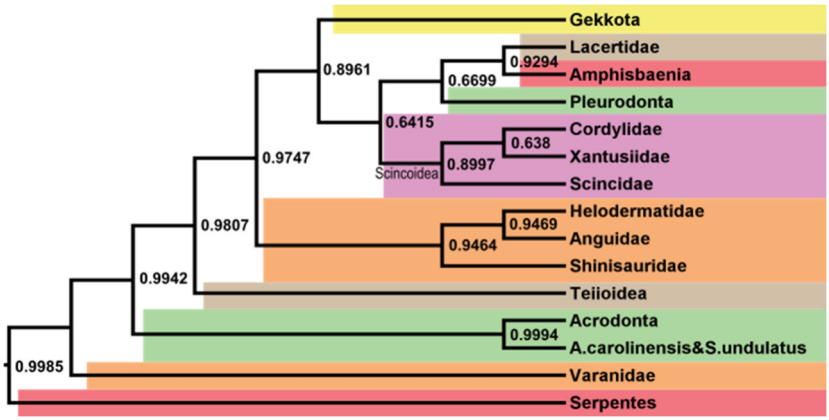

Fig.9: Molecular cladogram of squamate relationships summarized from the large-scale phylogeny. FBP support values as node labels. The Colour scheme is according to phylogenies in the introduction: red – "limbless", orange – Anguimorpha, green - Iguania, brown – Lacertoidea, pink – Scincoidea, yellow – Gekkota.

5.2.2 Venomousness evolved at least five times independently

The evolution of venomousness in squamates is tracked in Figure 10 which displays the large-scale phylogeny with collapsed monophyletic groups according to their venomousness (red). Within snakes, only advanced snakes of the clade Colubroidea are venomous Solecophidia and Henophidia are not venomous. Venomous monophyletic families of Colubroidea in this phylogeny are Pareatidae, Viperidae, Hydrophiidae, Dipsadidae, and Xenodermatidae. Elapidae are separated into three distinct clades. However, Colubridae is even more split up and does not contain only venomous species (Supplementary Table 4). Among Anguimorpha venomous taxa are Varanidae and Helodermatidae. So far Shinisauridae and Anguidae are presumed to be non-venomous. Within iguanas, only one acrodont genus *Pogona* and only one genus *Iguana* of Iguanidae are confidently considered to be venomous. According to this phylogeny the most parsimonious scenario of venom evolution involves at least five independent origins:

i) At the root of Colubroidea since all solecophidians and henophidians are not venomous. This implies several independent losses of venom in different clades of colubrids.
ii) At the root of Varanidae, because all monitor lizards are venomous and in this phylogeny, they are separated from the other anguimorphs.
iii) In the acrodont genus *Pogona* very terminal in the monophyletic Agamidae.
iv) At the root of Helodermatidae since both helodermatids are venomous but Shinisauridae and Anguidae are not considered to be venomous yet.
v) In the genus *Iguana* of the paraphyletic Iguanidae.

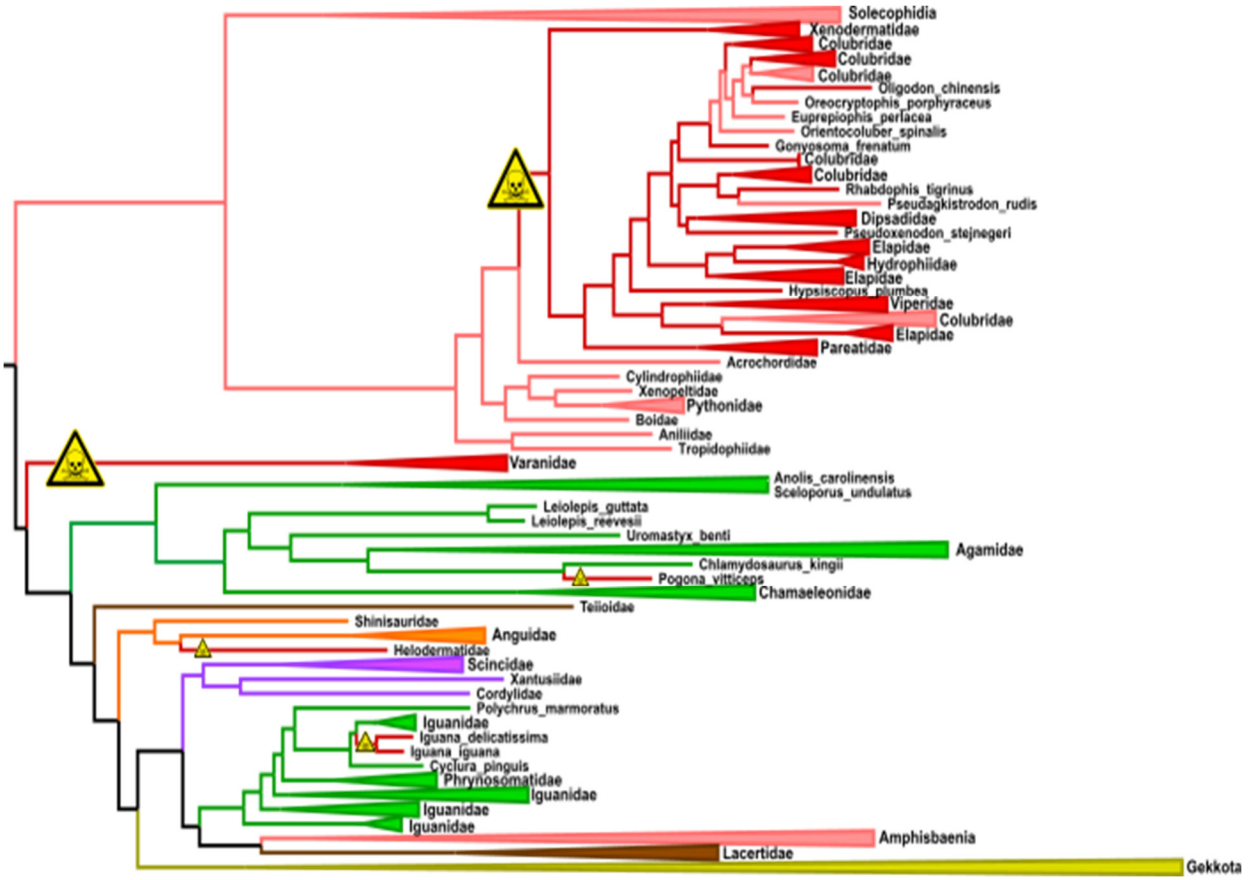

Fig.10: Large-scale maximum-likelihood phylogeny. Monophyletic groups collapsed. Colour scheme according to phylogenies in the introduction: pink – "limbless" (Serpentes + Amphisbaenia), green – Iguania, brown – Lacertoidea, orange – Anguimorpha, purple – Scincoidea, yellow – Gekkota. Venomous branches are labeled in red. Possible origins of venomousness are labeled with yellow signs.

5.2.3 The mt-phylogeny indicates a close relationship between snakes and acrodonts

Are phylogenies inferred from distinct datasets for mitochondrial and nuclear genes able to decrease the number of convergent origins of venomousness? The phylogeny inferred from a subset of the full dataset for mitochondrial genes displayed in Figure 11 consists of 36 protein families for 264 squamate species because sequences for each of the 13 mitochondrial genes were assorted to more than one protein family. Overall the tree topology matches the topology of the large-scale phylogeny but misses eight species. Six of them are snakes: *Pantherophis guttatus*, *Thamnophis elegans*, *Thamnophis sirtalis*, *Pseudonaja textilis*, *Notechis scuttatus*, and *Crotalus tigris*. The other two are iguanas: *Anolis carolinensis* and *Sceloporus undulatus.* Also here the root was confidently placed between Serpentes and all other squamates. The MAD statistics indicate this root appeared in 15.9% of the trees indicating the root with the highest likelihood is placed between Serpentes and all other squamates. In this phylogeny not Varanidae but Acrodonta is placed most basal in the sister group of snakes followed by the only species of Rhineuridae belonging to amphisbaenians which is well supported by an FBP of 0.93 but disperses the monophyly of Amphisbaenia. Lacertidae branch off after all other amphisbaenians supported by an FBP of 0.93 indicating a close relationship between lacertids and amphisbaenians similar to the large-scale phylogeny. The next branch in this phylogeny is

monophyletic Anguimorpha with Shinisauridae branching first (FBP 0.98) then Helodermatidae (FBP 0.97) and a sister relationship between Varanidae and Anguidae (FBP 0.97). Also here Lacertoidea is not monophyletic since Teiioidea branches off next as the sister group to Scincidae (FBP 0.94) which additionally breaks the monophyly of Scincoidea. The resulting clade appears as the sister group to Pleurodonta supported by a slightly lower FBP of 0.74. Finally the remaining Scincoidea namely Cordylidae and Xantusiidae (FBP 0.72) show a sister relationship with Gekkota well supported by an FBP of 0.95. Overall all higher taxa relationships are well supported with an FBP above 0.7 but also here Toxicofera are polyphyletic since they are separated into four distinct clades snakes, acrodonts, anguimorphs, and pleurodonts. Tracing back the evolution of venomousness according to this mitochondrial phylogeny four independent origins appear to be most parsimonious:

i) At the root of Colubroidea implying several independent losses in colubrid lineages, but fewer losses than in the large-scale phylogeny.
ii) In the acrodont genus *Pogona* since other acrodonts have not been confirmed to be venomous so far.
iii) At the root of Helodermatidae, Varanidae, and Anguimorpha but likely with a secondary loss in anguids.
iv) In the pleurodont genus *Iguana* since venomousness was so far not confirmed for all Iguanidae.

The difference to the large-scale phylogeny is the monophyly of anguimorphs that implicates a single origin of venomousness in Anguimorpha is more parsimonious than two when assuming losses are more likely than gains. Overall the phylogeny inferred from the mitochondrial dataset is able to erase one of the five hypothesizes convergent evolutions of venomousness in squamates.

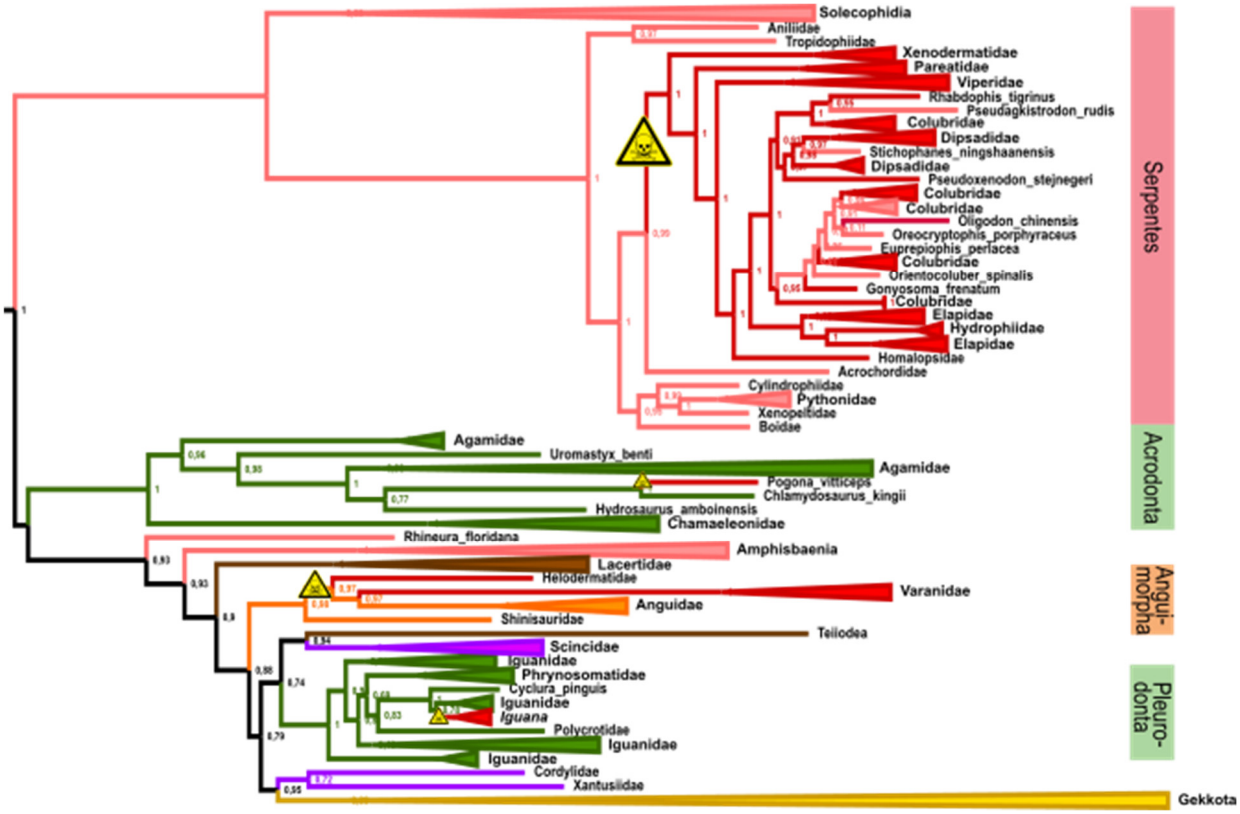

Fig.11: Mitochondrial maximum-likelihood phylogeny. Monophyletic groups collapsed. Colour scheme according to phylogenies in the introduction: pink – "limbless" (Serpentes + Amphisbaenia), green – Iguania, brown – Lacertoidea, orange – Anguimorpha, purple – Scincoidea, yellow – Gekkota. FBP support values as node labels. Venomous branches are labeled in red. Possible origins of venomousness are labeled with yellow signs.

5.2.4 Nuclear genes suggest three convergent evolutions of venomousness

The phylogeny inferred from the subset of the original dataset for only nuclear genes consists of 732 protein families for only 16 squamate species. Among these 16 species are the eight that are missing in the mitochondrial phylogeny (Figure 12). Some higher-level squamate lineages got completely lost due to the lack of data namely Pleurodonta, Amphisbaenia, and Scincoidea. Additionally, anguimorphs are represented by only one species *Varanus komodoensis*. Also, most basal snakes got lost therefore *Python bivittatus* is the only species representing all non-Colubroidea. Nonetheless, the root was confidently placed between snakes and all other squamates. The MAD statistics reveal that this root appears in 12.7% of the phylogenies. In this nuclear phylogeny, all present snake families Viperidae, Elapidae, and Colubridae are monophyletic. The relationships within snakes are highly supported with an FBP of 1. Regarding the branch containing all other squamates Varanidae representing anguimorphs branches off first. The remaining taxa show a sister relationship between acrodonts which represent Iguania and a clade consisting of Lacertidae and Gekkota. All nodes are extremely well supported with an FBP of 1 except for the node branching into *Zootoca vivipara* and *Lacerta agilis* which is supported by an FBP of 0.99. Here Toxicofera is paraphyletic since lacertids, and geckos would be included because of the position of the root. Therefore, venomousness could have evolved only once according to this tree topology at the root of Squamata. But this would indicate

several losses of venomousness. It would have been lost in some Iguania lineages and the branch leading to Lacertidae and Gekkota. In snakes, venomousness would have been lost in all basal snake families represented by *Python bivittatus* but then regained in Colubroidea and secondarily lost in Colubridae. An alternative for this scenario is displayed in Figure 12. It suggests three independent origins for venomousness in Squamata and the only loss would appear within Colubridae:

i) At the root of Colubroidea with a loss in Colubridae.
ii) In Varanidae.
iii) In the acrodont genus *Pogona.*

Which scenario of the evolution of venomousness is more parsimonious depends on the costs and the likelihood of losing and regaining venom. Because the datasets for mitochondrial and nuclear genes are varying a lot in taxon sampling and the number of genes it is necessary to analyze the impact of the data source on the consensus of the species from all three phylogenies.

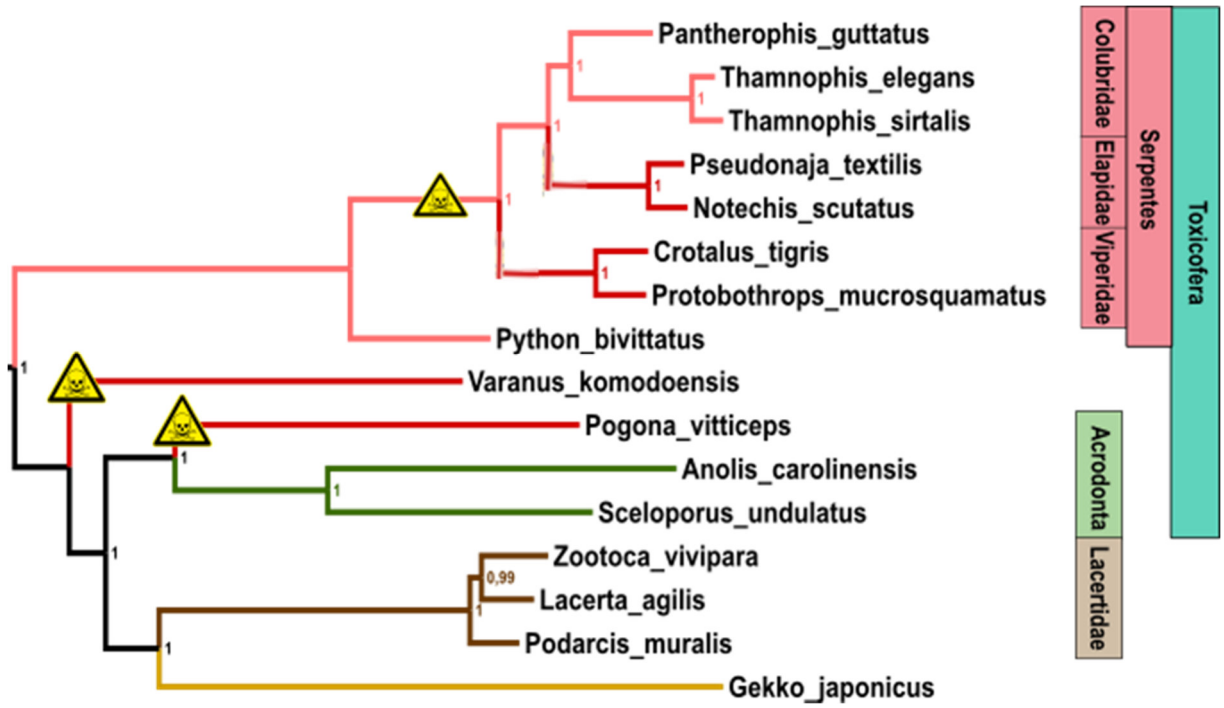

Fig.12: Nuclear maximum-likelihood phylogeny. Colour scheme according to phylogenies in the introduction: pink – "limbless" (Serpentes), green – Iguania, brown – Lacertidae, yellow – Gekkota. FBP support values as node labels. Venomous branches are labeled in red. Possible origins of venomousness are labeled with yellow signs.

5.2.5 The position of *Varanus komodoensis* depends on the data source

To analyze the impact on the tree topology of using either only mitochondrial genes or only nuclear genes subsets of the full phylogenies for the eight species they have in common are displayed in Figure 13. Even though just eight species are included all major squamate lineages are present. *Protobothrops mucrosquamatus* represents Colubroidea, *Python bivittatus* represents all basal snakes of Henophidia and Solecophidia, *Varanus komodoensis* represents Anguimorpha, and *Pogona vitticeps* represents Iguania. In addition to these Toxicofera, non-Toxicofera are represented by three lacertids and *Gekko japonicus* which represents Gekkota. The topology of the large-scale phylogeny for these eight species (A) matches the topology of the nuclear tree (C) with Toxicofera being paraphyletic by including Lacertidae and Gekkota. In these two phylogenies, *Varanus komodoensis* is placed most basal after the root that separates Serpentes from the others, followed by *Pogona vitticeps* which is placed as the sister taxon to all non-Toxicofera. However, in the mitochondrial phylogeny (B) Toxicofera is polyphyletic because Varanus komodoensis appears as the sister taxon to Gekkota.

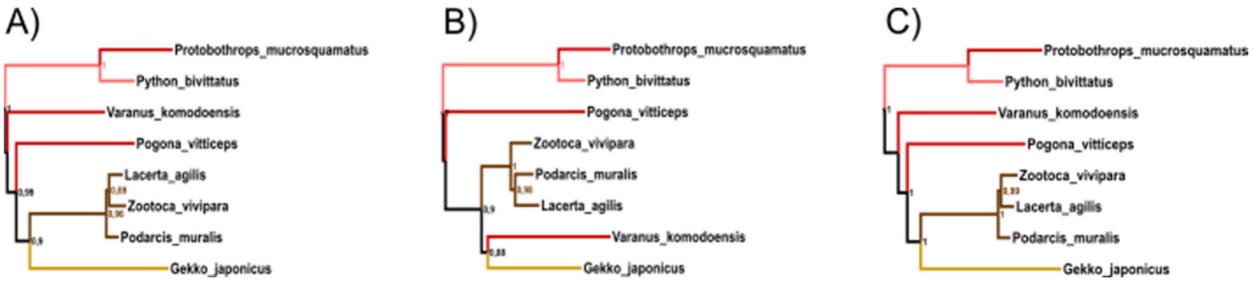

Fig.13: Subsets of the maximum-likelihood phylogenies for A) the whole dataset, B) mitochondrial genes only, and C) nuclear genes only showing the relationships of eight squamate species that were included in all datasets. Colour scheme according to phylogenies in the introduction: pink – "limbless" (Serpentes), brown – Lacertidae, yellow – Gekkota. FBP support values as node labels. Venomous branches are labeled in red.

To recapitulate, in all three distinct phylogenies the root was reliably placed between snakes and all other squamates, therefore Toxicofera does not appear to be monophyletic according to any of the tree topologies. Additionally, Gekkota and Lacertoidea always appear very terminal in the phylogenies. Tracing back the evolution of venomousness in squamates reveals multiple independent origins of venomousness in all tree topologies. Restricting to subsets of the phylogenies for only those species they have in common indicates the topology of the nuclear phylogeny is more similar to the topology of the large-scale phylogeny than the mitochondrial phylogeny even though the dataset is biased toward mitochondrial data. Furthermore, analyses of the molecular evolution of distinct venom proteins can shed light on the question about multiple convergent evolution of venomous in squamates

5.3 Molecular evolution of venom Phospholipase A_2

Phospholipase A_2 is a venom protein frequently found in snakes and lizards [42,43] that inhibits platelet aggregation in the prey [45]. Here, it is analyzed according to differences in selection pressure between venomous and non-venomous squamates which can give an insight into when PLA_2 was recruited into the venom. Additionally, sites under positive selection can elucidate which residues are involved in toxin function if they can be considered fast-evolving surface residues. Finally, evaluation of coevolving groups can reveal if these functionally important sites are co-adapting.

5.3.1 PLA_2 was recruited four times into the venom of Toxicofera

The dataset for venom Phospholipase A_2 containing sequences for 17 squamate species was analyzed to study the molecular evolution of venom proteins in Toxicofera by implementing approaches of phylogenetics, selection models of codon sequence evolution, coevolution, and structural analyses. Figure 14 displays a subset of the large-scale phylogeny from the previous phylogenomic approach for the set of species used in PLA_2 analyses. According to this tree topology phospholipase A2 was recruited into the venom of four distinct squamate lineages:

i) At the root of Colubroidea with a loss in Colubridae.
ii) At the root of Varanidae.
iii) In the acrodont genus *Pogona*.
iv) At the root of Helodermatidae.

It is also possible that PLA_2 evolved a toxin function only once at the root of squamates, but this function got lost in several lineages. In snakes it would have been lost in all basal sake families and was then regained by Colubroidea and got secondarily lost in some colubrids. In the branch of all other squamates, PLA_2 would have lost its toxin function in some iguanas and the sister group of Helodermatidae. Selection models of codon sequence evolution could support the hypothesis of four independent recruitments of PLA_2 into the venom by signals of positive selection in these clades.

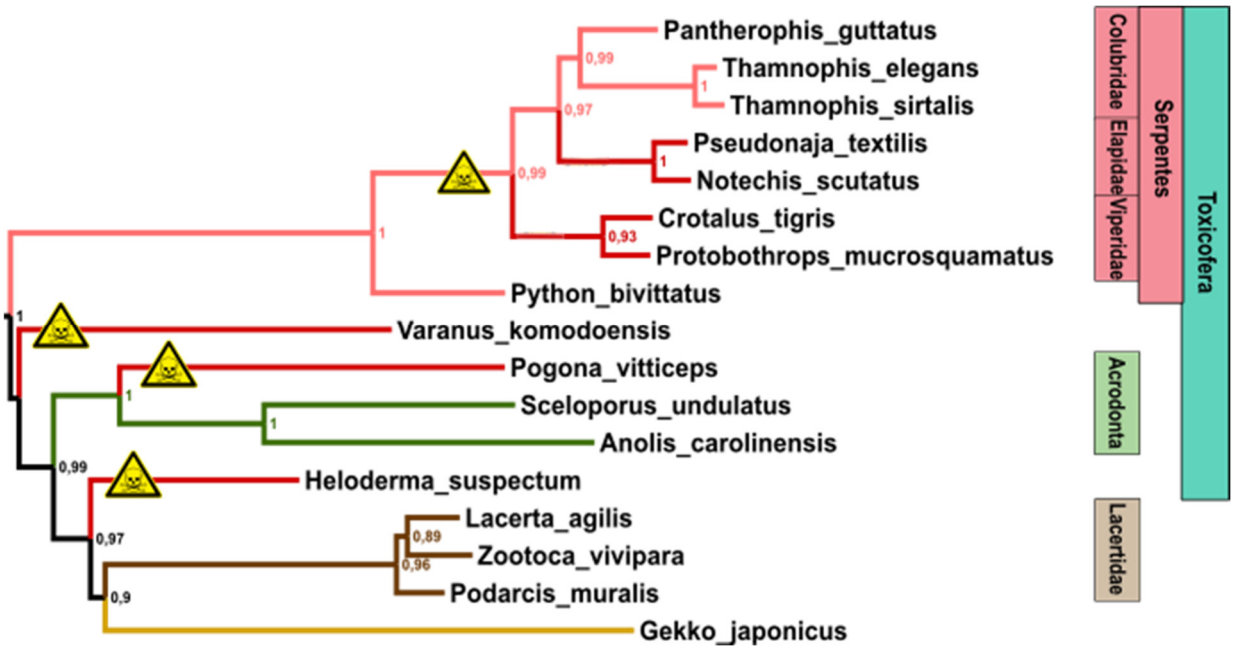

Fig.14: Subset of the large-scale phylogeny for 17 species that are included in PLA_2 analyses. Colour scheme according to phylogenies in the introduction: pink – "limbless" (Serpentes), green – Iguania, brown – Lacertidae, yellow – Gekkota. FBP support values as node labels. Venomous branches are labeled in red. Possible origins of venomousness are labeled with yellow signs.

5.3.2 PLA_2 Branch-Models reveal signal of positive selection in Anguimorpha

Selection models of codon sequence evolution were performed on the dataset for PLA_2 using the corresponding subset of the large-scale phylogeny as the underlying tree topology. These models conduce to analyze the difference in selection pressure acting on PLA_2 between venomous and non-venomous clades. Branch-Models performed in PAML to obtain the average dN/dS for different squamate lineages are illustrated in Table 1. Species were assigned to groups at different taxonomic levels to detect at which level (family or higher) signals of positive selection appear. Additionally, models included a distinct dN/dS parameter for the branch leading to a specific group, in order to distinguish scenarios where adaptation occurred in the ancestor or was recurrent within the group. Model 0 computes a single dN/dS ratio for the whole alignment for comparison. The first model compares the selection pressure acting on PLA_2 between Toxicofera, Lacertidae, and Gekkota. All calculated dN/dS values are clearly below 1 and range between 0.1 and 0.2. This indicated all three clades experience purifying selection. The dN/dS of the venomous clade (Toxicofera) is not higher than the dN/dS of non-venomous clades. Model 2 compares clades at a lower taxonomic level. Here Iguania, Anguimorpha, and Serpentes are compared to Gekkota and Lacertidae. Overall the dN/dS ratios are still clearly below 1 but slightly higher than in the previous model in anguimorphs (0.56) and the root of Iguania (0.26). In the third model, clades are analyzed at the family level to localize the sources of signals of possible positive selection deeper in the phylogeny. Iguania families and snake families show globally very low dN/dS which is highest in Pythonidae and the root of Elapidae. However, anguimorph families show overall high dN/dS with the highest value coming from

Helodermatidae (0.89) being close to 1 indicating neutral evolution of PLA_2 in *Heloderma*. Model 4 and Model 5 aim to find differences in selection pressure between venomous and non-venomous squamates regardless of their systematic classification. Model 4 compares all venomous species with all non-venomous species. Both clades reveal very low dN/dS values of 0.13 for non-venomous and 0.23 for venomous species. Model 5 additionally differentiates between venomous and non-venomous Toxicofera indicating a slightly higher dN/dS in venomous Toxicofera (0.23) than in non-venomous Toxicofera (0.11). In general Branch-Model for PLA_2 indicate purifying selection for all squamate lineages except for Anguimorpha where in particular Helodermatidae give evidence for neutral evolution of venom PLA_2.

Table 1: PLA_2 Branch Models from PAML

	Model 0		Model 1		Model 2		Model 3		Model 4		Model 5	
Branch Label	Clade	dN/dS	Clade	dN/dS	Clade	dN/dS	Clade	dN/dS	Clade	dN/dS	Clade	dN/dS
0	Squamata	0.15	Gekkota	0.13	Gekkota	0.22	Gekkota	0.17	all non-venomous	0.13	non-Toxicofera	0.14
1			Lacertidae	0.18	Lacertidae	0.18	Lacertidae	0.18	all venomous	0.23	non-venomous Toxicofera	0.11
2			Toxicofera	0.16	Iguania	0.11	Dactyloidae	0.11			venomous Toxicofera	0.23
3			root Lacertidae	0.13	Anguimorpha	0.56	Phrynosomatidae	0.14				
4					Serpentes	0.17	Agamidae	0.08				
5					roor Lacertidae	0.12	Varanidae	0.32				
6					root Iguania	0.26	Helodermatidae	0.89				
7					root Serpentes	0.07	Elapidae	0.12				
8							Viperidae	0.21				
9							Colubridae	0.08				
10							Pythonidae	0.66				
11							root Lacertidae	0.12				
12							root Elapidae	0.38				
13							root Viperidae	0.22				
14							root Colubridae	0.13				

Models were performed at different taxonomic levels. Venomous clades are labeled in red.

Branch-Models compute the average dN/dS over all sites and are therefore influenced by both positive and purifying selection. If positively selected sites are rare the average ratio will still be below 1. In Table 1 some venomous clades appear to have higher dN/dS ratios. This could be due to the presence of only a few sites under positive selection or relaxation of selection. To distinguish between those two possibilities models that allow dN/dS ratios to vary also between sites are implemented.

5.3.3 PLA_2 Branch-Site models suggest six sites under positive selection in Varanidae

Branch-Site Models that aimed to find sites that are significantly under positive selection in venomous squamates were performed on 16 different subsets of the PLA_2 dataset (Table 2 and Figure 5 in 4.2.1). The Bayes Empirical Bayes (BEB) approach revealed sites that are significantly under positive selection for Branch-Site Models 3, 6, 7, and 16 (Table 2). However, the Null-Hypothesis that does not allow dN/dS above 1 in foreground lineages is only rejected for Branch-Site Model 6 since the Chi-Square Test yields a p-value below 0.05 only in the model that compares Varanidae with non-Toxicofera ($p \sim 0.002$). The six significantly positive selected sites selected in Varanus komodoensis are visualized in the protein structure predicted with AlphaFold in Figure 15. The Relative Solvent Accessibility (RSA) displayed in the legend indicates to what extent a residue is exposed to the protein surface. Residues with an RSA of 1

are completely exposed to the protein surface while residues with an RSA of 0 are completely buried in the structure. Half of the positively selected sites in the Komodo dragon PLA_2 are Serines of which all have RSA values higher than 0.25 indicating these sites are considered to be surface residues [82] but when looking at the average RSA of the sites in the *Varanus* sequence it appears only sites with an RSA above 0.36 are more exposed to the protein surface than the average (Figure 16). PHE55 and TYR100 are slightly more exposed to the protein surface than the average but ALA112 is with an RSA of 0.08 completely buried in the structure. In general these positive selected sited do not tend to be surface residues.

Table 2: PLA_2 Branch-Site Models from PAML

Model	Foreground	Background	Num. sites p<0.05	lnL HA	lnL H0	p Value
Branch-Site 1	Toxicofera	non-Toxicofera	0	-4710.454515	-4710.454515	1
Branch-Site 2	Iguania	non-Toxicofera	0	-2565.241842	-2565.241842	1
Branch-Site 3	Anguimorpha	non-Toxicofera	5	-1933.836207	-1934.660636	0.1991144
Branch-Site 4	Serpentes	non-Toxicofera	0	-3003.050457	-3003.050457	1
Branch-Site 5	Agamidae	non-Toxicofera	0	-1793.503844	-1793.503844	1
Branch-Site 6	Varanidae	non-Toxicofera	6	-1643.437859	-1648.276051	0.001866516
Branch-Site 7	Helodermatidae	non-Toxicofera	6	-1682.257728	-1682.328593	0.7065672
Branch-Site 8	Elapidae	non-Toxicofera	0	-1961.106967	-1961.106967	1
Branch-Site 9	Viperidae	non-Toxicofera	0	-1993.376211	-1994.205693	0.1977433
Branch-Site 10	Colubridae	non-Toxicofera	0	-2139.189012	-2139.189012	1
Branch-Site 11	Colubroidea	Pythonidae	0	-2117.810354	-2117.810354	1
Branch-Site 12	ven. Serpentes	non-ven. Serpentes	0	-2115.919708	-2115.919708	1
Branch-Site 13	ven. Iguania	non-ven. Iguania	0	-1794.911195	-1794.925013	0.8679675
Branch-Site 14	Elapidae	Colubridae	0	-1503.873274	-1504.002941	0.6105781
Branch-Site 15	Viperidae	Colubridae	0	-1558.129024	-1558.261788	0.6063475
Branch-Site 16	all venomous	non-all venomous	2	-4704.391765	-4704.391765	1

Significant p-values marked in red.

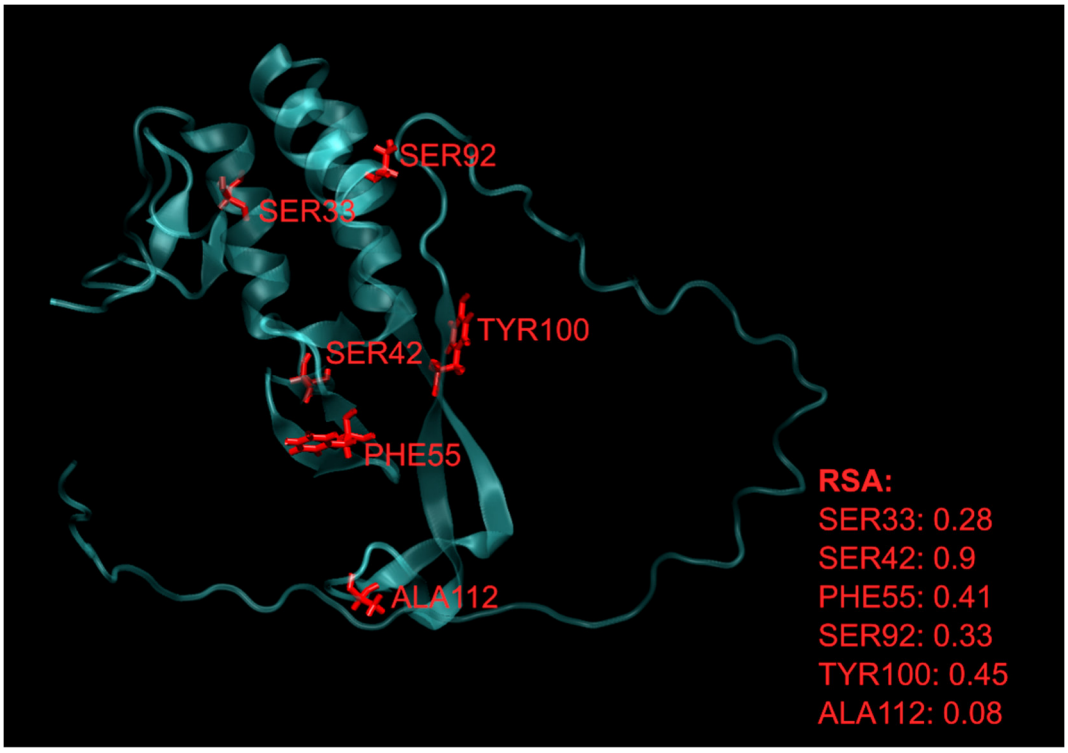

Fig.15: PLA_2 structure of *Varanus komodoensis*. Positively selected sites are labeled in red. The legend shows the relative solvent accessibility.

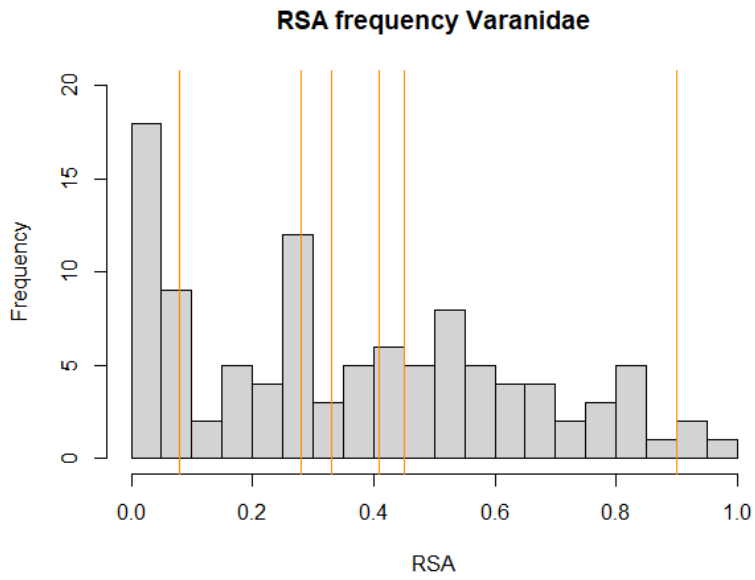

Fig.16: Histogram: RSA frequencies of PLA_2 for *Varanus komodoensis*. On the X-Axis is the RSA and on the Y-Axis is the frequency. RSAs of positively selected sites are labeled in orange. Mean = 0.36, median = 0.34

Substitution rates were calculated to detect a pattern in the evolutionary rate frequency of positively selected sites. The majority of positively selected residues in *Varanus komodoensis* were found to have an average substitution rate, only one site SER42 had a comparatively high rate of 3.3 (Figure 17). Co-adaptation of residues under positive selection could support the importance of these sites for toxin function and fast adaptation to changes in the target molecules in the prey.

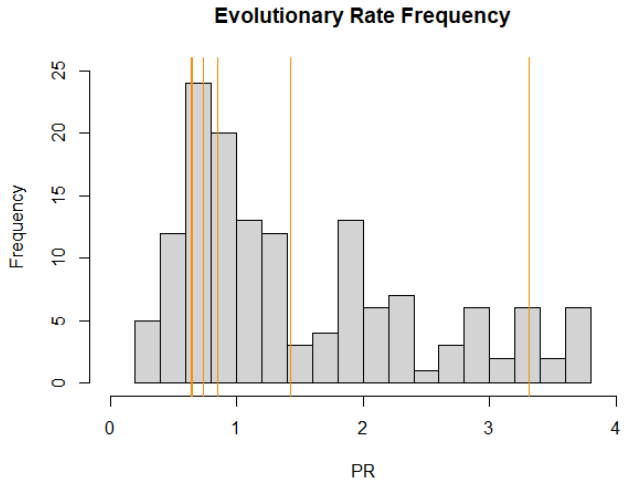

Fig.17: Histogram of the evolutionary rate frequency of *Varanus komodoensis* PLA_2. On the X-Axis is the posterior rate and on the Y-Axis is the frequency. PRs of positively selected sites are labeled in orange. Mean = 1.51, median = 1.19.

5.3.4 One residue of PLA_2 under positive selection is co-adapting with six other sites

Coevolution analyses aimed to find coevolving groups of sites that might include the positively selected sites from *Varanus komodoensis*. CoMap [79] revealed five groups of coevolving sites with the 'Simple' approach that does not account for the biochemical properties of amino acids and all substitutions are weighted equally (Table 3). These groups are visualized in the protein structure of *Thamnophis elegans* in Figure 18. Sizes of coevolving groups range from two to seven sites. Among the largest group of coevolving sites is one site that is significantly under positive selection in the Komodo dragon. ASN234 in the protein structure of *Thamnophis elegans* matches SER92 in the protein structure of *Varanus komodoensis*. ASN234 is coevolving with VAL56, GLU237, LEU235, ALA185, LEU266, and TYR261 which correspond with ALA15, GLN95, VAL93, VAL43, MET125, and ARG120 in the *Varanus* protein structure. Even though these sites are not among the significantly positive selected they are included in the BEB results for Branch-Site Model 6 except for MET125 but are not considered to be significant. This indicated the posterior probabilities of these sites of coming from the site class with dN/dS > 1 are between 50% and 95% because BEB analyses filter sites for all that have a more than 50% chance of coming from the class with dN/dS > 1 [75]. Therefore, these sites may be under positive selection too.

Table 3: PLA_2 Coevolving groups

Group	Alignment positions	Structure residues	Size	Method	p-value
1	[115;140]	PRO355; PHE425	2	Simple	0.011
2	[181;226]	HIS466; ALA514	2	Simple	0.012
3	[44;223;147]	ASN211; ARG408; SER432	3	Simple	0.031
4	[3;12;82]	PHE170; GLU179; GLU251	3	Simple	0.039
5	[2;65;68;66;18;95;90]	VAL56; ASN234; GLU237; LEU235; ALA185; LEU266; TYR261	7	Simple	0.012

Site under positive selection marked in red.

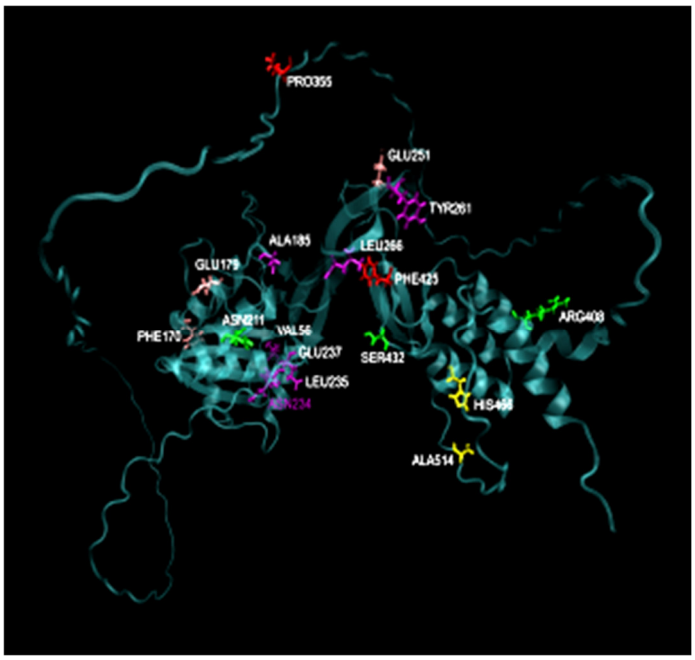

Fig.18: PLA_2 structure of *Thamnophis elegans*. Coevolving groups sites are labeled with different colors: red – PHE425; PRO355, yellow – HIS466; ALA514, green – ASN211; ARG408; SER432, pink – PHE170; GLU179; GLU251, and purple – VAL256; ASN234; GLU237; LEU235; ALA185 LEU266; TYR261.

Analyzing the evolutionary rate frequency of the coevolving groups of sites reveals that they overall follow the pattern of posterior probabilities across the whole PLA_2 sequence of *Varanus komodoensis* (Figure 19). Two of the coevolving groups range around the PR maximum at 1. Among these is the large groups of coevolving sites that match the positively selected sites marked in yellow in the histogram in Figure 19. The other three groups experience higher evolutionary rates but are still consistent with peaks in the overall distribution of evolutionary rate frequencies. Overall co-adapting groups do not evolve faster than the average indicating they are more conserved.

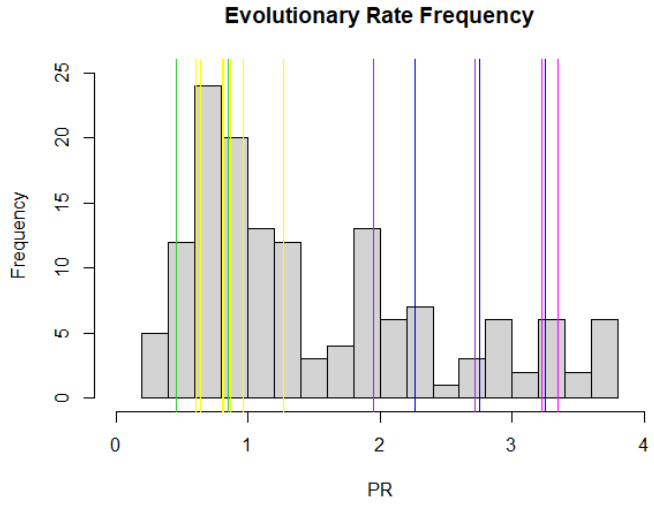

Fig.19: Histogram of the evolutionary rate frequency of venom PLA_2. On the X-Axis is the posterior rate and on the Y-Axis is the frequency. PRs of coevolving sites are labeled according to coevolving groups. Mean = 1.51, median = 1.19.

Comparing the Phospholipase A_2 protein structure of *Varanus komodoensis* with the literature [43,44,47] reveals the active site of the protein, which is in the classical arrangement and consistent with the active site of close relatives of *Varanus* namely *Heloderma*. The first of three residues that form the active site is HIS38 which is connected with a hydrogen bond to the second residue ASP68 which is then connected to the third residue TYR90 via a hydrogen bond. This pattern is visualized in the protein structure of the Komodo dragon in Figure 20. The RSAs of these residues show that the active site is buried deep in the structure. None of the residues of the active site is among the coevolving groups or significantly positive selected but the third position TYR90 which is the most variable position is retrieved by the BEB analysis of Branch-Site Model 6.

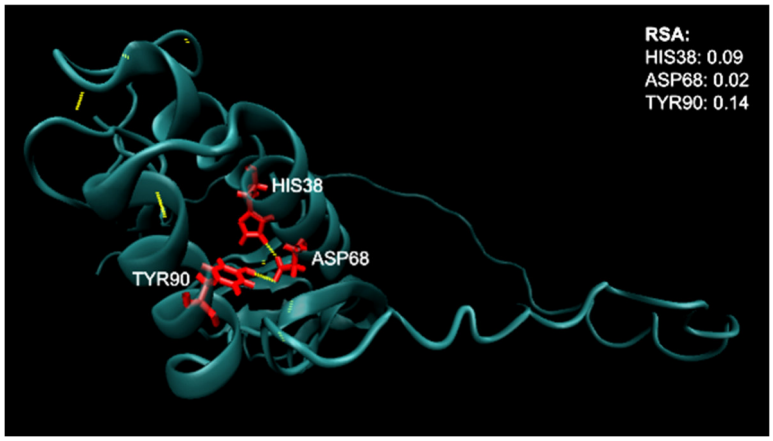

Fig.20: PLA_2 structure of *Varanus komodoensis*. Residues of the active site are labeled in red. Hydrogen bonds are labeled in yellow. The legend shows the relative solvent accessibility.

In summary, Phospholipase A_2 is a frequent venom protein that according to the phylogeny in Figure 14 was recruited into the venom multiple times independently within different Toxicofera lineages. The average dN/dS ratios calculated for all different squamate clades included in the dataset down to the family level overall indicates purifying selection for most squamates regardless of their venomousness except for anguimorphs where Helodermatidae have the highest dN/dS suggesting nearly neutral evolution. Enabling the dN/dS to vary also between sites reveals six sites that are significantly under positive selection in *Varanus komodoensis*. Only half of these sites are overall more exposed to the protein surface than the average but do not have remarkably higher evolutionary rates than other sites. Venom PLA_2 appears to have five groups of coevolving sites of which the largest (7 sites) includes one of the sites that are significantly positively selected in the Komodo dragon and five more sites that were additionally suggested by BEB analyses but are not considered to be significant. Finally, the active site of venom PLA_2 in Varanus komodoensis matches the classical arrangement of HIS-ASP-TYR whereat the third position appears in the BEB analyses within sites where positive selection is not considered to be significant.

5.4 Molecular evolution of venom cysteine-rich secretory proteins

Cysteine-rich secretory protein is a venom protein frequently found in snakes and lizards [49] that blocks smooth muscle contraction [48] and induces hypothermia [32]. Here, it is analyzed according to differences in selection pressure between venomous and non-venomous squamates which can give an insight into when CRISP was recruited into the venom. Additionally, sites under positive selection can elucidate which residues are involved in toxin function if they can be considered fast-evolving surface residues. Finally, evaluation of coevolving groups can reveal if these functionally important sites are co-adapting.

5.4.1 The phylogeny for CRISP reveals distinct paralog branches for snakes and lizards

The phylogeny in Figure 21 is a maximum-likelihood phylogeny for 97 CRISP sequences and 32 squamate species. Since the dataset contains many paralogs that could not be identified confidently the gene tree was reconciled with a subset of the large-scale species tree that corresponds with the CRISP species list (Supplementary Table 4). The CRISP phylogeny has many multifurcations because during reconciliation a TBE support threshold of 60% was applied which turns all weakly supported branches into multifurcations. Therefore it is not possible to trace back the origin of venomousness in the tree topology. The overall structure of the phylogeny sorts the sequences into three distinct groups. The first branch consists of all the snake sequences. The two other branches have a sister relationship while one branch is much longer than the other. Venomous anguimorphs and iguanas more precisely *Varanus komodoensis* and *Pogona vitticeps* appear widespread over the two branches. The only sequence for the second venomous anguimorph family Helodermatidae appears is the short branch as the sister taxon to the Komodo dragon. Non-venomous lizards of the family Lacertidae appear mainly in the short branch where also all anguimorph sequences appear to be very closely related compared to the long branch where all species are more spread.

Fig.21: CRISP maximum-likelihood phylogeny reconciled with a subset of the large-scale phylogeny for 32 species that are included in CRISP analyses. Colour scheme according to phylogenies in the introduction: pink – "limbless" (Serpentes), green – Iguania, brown – Lacertidae, yellow – Gekkota. TBE support values as node labels with a threshold of 60. Monophyletic clades collapsed. Venomous branches are labeled in red.

5.4.2 CRISP Branch-Models indicate signal of positive selection in Colubroidea

The distinct paralog branches in the phylogeny are not treated independently in the following analyses but paralogs from the same species are grouped together. Selection models of codon sequence evolution were performed on the dataset for CRISP using the corresponding reconciled phylogeny as the underlying tree topology. These models aim for studying differences in selection pressure acting on CRISP sequences between venomous and non-venomous squamates by calculating the average dN/dS over the whole sequence for predetermined clades. The models were performed at different taxonomic levels to trace back the signals of positive selection down to the family level (Table 4). The first model compares higher-level squamates Gekkota, Lacertidae, and Toxicofera. While CRISP in geckos appears to be neutrally evolving (0.96) Lacertidae and Toxicofera experience purifying selection with Lacertidae having a higher dN/dS ratio of 0.61 than Toxicofera (0.39). Model 2 aims to find differences between the three toxicoferan clades. Iguania, Serpentes, and Anguimorpha have similar dN/dS values that indicate negative selection. The third model compares taxa at the family level. There is no signal for positive selection in the three Iguania families because their dN/dS ratios are clearly below 1. Also, the dN/dS values of the two anguimorph families indicate purifying selection with a dN/dS of 0.45 for Varanidae and a dN/dS of 0.31 for Helodermatidae. In snakes, the only clear signal for positive selection arises in Elapidae with a dN/dS of 1.36 but

also venom CRISPs in the other two Colubroidea families Viperidae (0.95) and Colubridae (0.93) do not appear to be under purifying selection but rather neutral evolving. Model 4 compares all venomous species with all non-venomous species regardless of their systematic classification. Here, venomous species have only a slightly higher dN/dS ratio than non-venomous species which still indicates purifying selection. The last model behaves similarly, but it compares also venomous and non-venomous Toxicofera. The results are similar to Model 4 since the dN/dS of venomous Toxicofera is only slightly higher than the dN/dS of non-venomous clades.

Table 4: CRISP Branch Models from PAML

	Model 0		Model 1		Model 2		Model 3		Model 4		Model 5	
Branch Label	Clade	dN/dS	Clade	dN/dS	Clade	dN/dS	Clade	dN/dS	Clade	dN/dS	Clade	dN/dS
0	Squamata	0.45	Gekkota	0.96	Gekkota	0.96	Gekkota	0.45	all non-venomous	0.45	non-Toxicofera	0.46
1			Lacertidae	0.61	Lacertidae	0.61	Lacertidae	0.59	all venomous	0.51	non-venomous Toxicofera	0.41
2			Toxicofera	0.39	Iguania	0.39	Dactyloidae	0.34			venomous Toxicofera	0.5
3					Serpentes	0.34	Phrynosomatidae	0.42				
4					Anguimorpha	0.45	Agamidae	0.38				
5							Varanidae	0.45				
6							Helodermatidae	0.31				
7							Elapidae	1.36				
8							Viperidae	0.95				
9							Colubridae	0.93				
10							Pythonidae	0.46				

Models were performed at different taxonomic levels. Venomous clades are labeled in red.

5.4.3 Venom CRISP is under positive selection in Colubroidea and Helodermatidae

Branch-Site Models were implemented to allow the dN/dS to vary also between sites to identify sites under positive selection in different clades of venomous squamates. The 17 models used here are explained in Figure 6 and Table 5. The BEB approach for these models proposes sites that are under significant positive selection for all models except for Branch-Site Models 13 which compares venomous with non-venomous iguanas, 14 which compares Viperidae with Colubridae, and 17 which compares venomous colubrids more precisely *Rhabdophis tigrinus* with non-venomous Colubridae. Anyhow, testing for significance using the Chi-square test reveals that the Null-hypothesis that does not enable positive selection can only be rejected in Branch-Site Models 1, 4, 7-12, and 15 because these models have p-values below 0.05 (Table 5). In general, this means at the family level significant positively selected sites appear only in Helodermatidae, Elapidae, Viperidae, and Colubridae. Figure 22 shows the by RoseTTAFold predicted protein structures of CRISP for each one representative species of these four families. In Helodermatidae (A) only one site is significantly under positive selection. Serine 184 is considered to be a surface residue since the RSA tells it is more than 25% accessible to the solvent. Elapidae (B) has only one positive selected site, too. CYS233 is with an RSA of 0.16 buried deep in the structure. Viperidae (C) is the family having the most positive selected sites. Nine out of ten positively selected residues are accessible to the solvent more than 26% which is more than the average (Figure 23). Therefore they are considered surface proteins. Also, Colubridae (D) have many positively selected sites. Two of these eight significant sites are

buried with RSAs of less than 0.25. The remaining six resecures are exposed to the surface more than the average of 28% (Figure 23).

Table 5: CRISP Branch-Site Models from PAML

Model	Foreground	Background	Num. sites p<0.05	lnL HA	lnL H0	p Value
Branch-Site 1	Toxicofera	non-Toxicofera	9	-19685.85233	-19706.45824	1.36594E-10
Branch-Site 2	Iguania	non-Toxicofera	1	-12419.94546	-12419.94546	1
Branch-Site 3	Anguimorpha	non-Toxicofera	3	-7442.851969	-7442.851969	1
Branch-Site 4	Serpentes	non-Toxicofera	17	-9219.478506	-9296.608343	2.03226E-35
Branch-Site 5	Agamidae	non-Toxicofera	1	-7721.087406	-7721.087406	1
Branch-Site 6	Varanidae	non-Toxicofera	2	-7116.171971	-7116.171971	1
Branch-Site 7	Helodermatidae	non-Toxicofera	1	-5028.010306	-5030.03751	0.04405616
Branch-Site 8	Elapidae	non-Toxicofera	1	-6100.733749	-6104.982612	0.003555906
Branch-Site 9	Viperidae	non-Toxicofera	10	-6925.66128	-6958.022359	8.62396E-16
Branch-Site 10	Colubridae	non-Toxicofera	8	-5867.289017	-5885.115758	2.35726E-09
Branch-Site 11	Colubroidea	Pythonidae	10	-4982.983024	-5030.828115	1.34345E-22
Branch-Site 12	ven. Serpentes	non-ven. Serpentes	5	-5014.906467	-5024.320622	1.43028E-05
Branch-Site 13	ven. Iguania	non-ven. Iguania	0	-8598.666359	-8598.666359	1
Branch-Site 14	Elapidae	Colubridae	0	-2750.391562	-2752.189307	0.05793655
Branch-Site 15	Viperidae	Colubridae	5	-3657.213022	-3661.900626	0.002199397
Branch-Site 16	all venomous	non-all venomous	3	-19708.25913	-19706.37535	1
Branch-Site 17	ven. Colubridae	non-ven. Colubridae	0	-1622.746903	-1622.746903	1

Significant p-values marked in red.

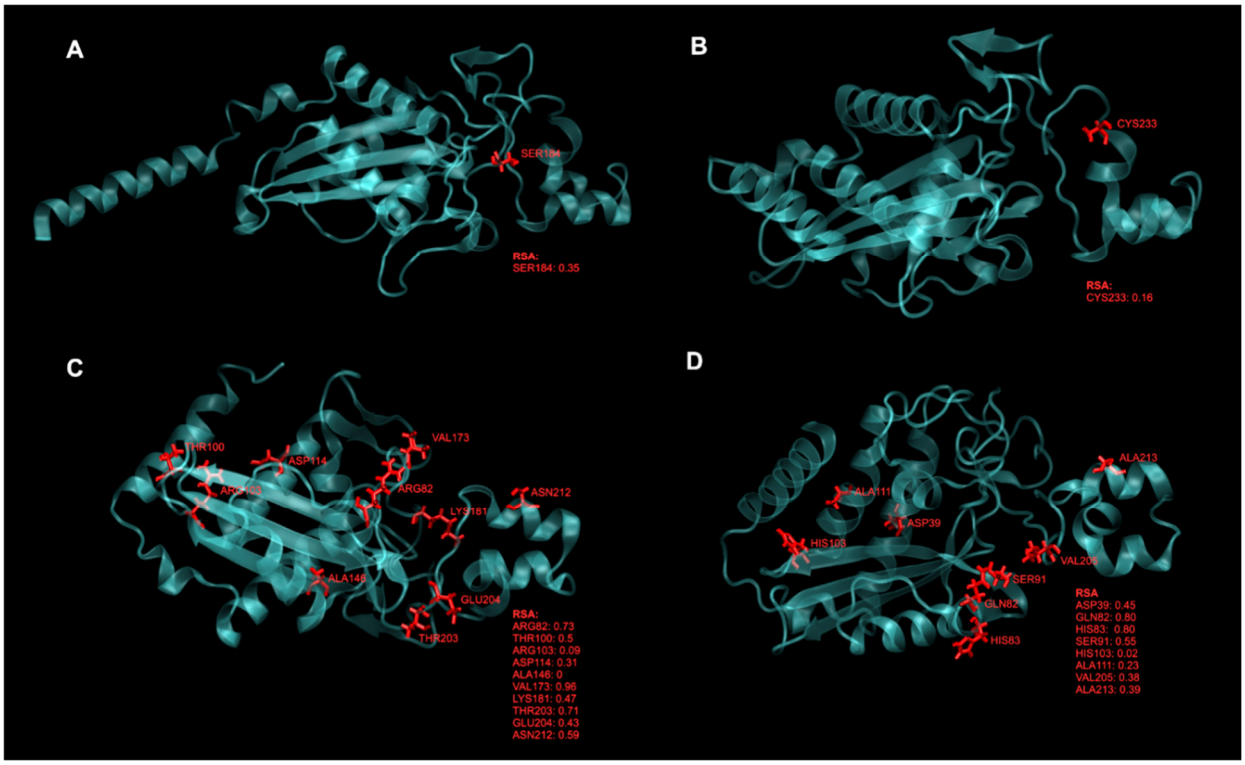

Fig.22: CRSIP structure of A) *Heloderma suspectum* (Helodermatidae), B) *Pseudonaja textilis* (Elapidae), C) *Ovophis okinavensis* (Viperidae), and D) *Pantherophis guttatus* (Colubridae). Positively selected sites are labeled in red. Legends show the relative solvent accessibility.

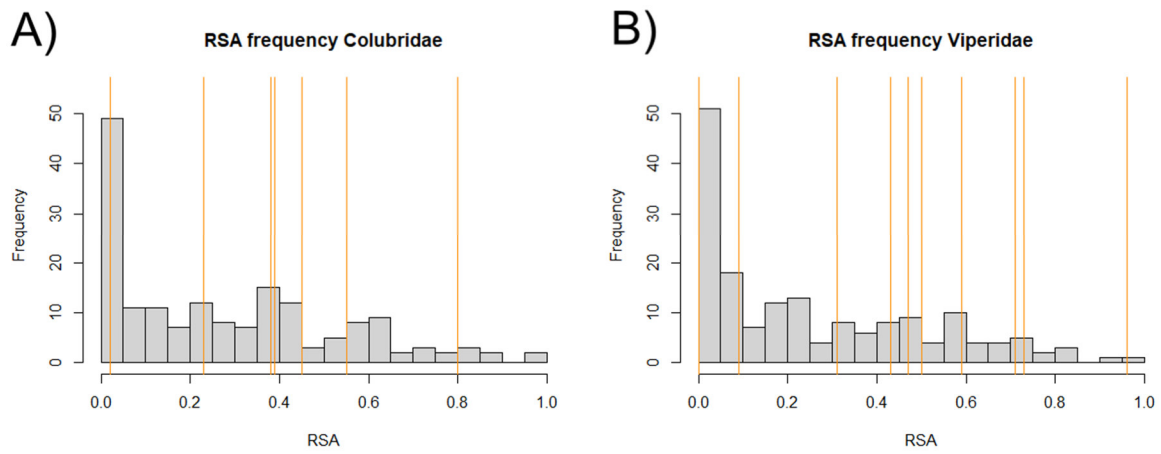

Fig.23: Histogram of the RSA frequency of CRISP for A) *Ovophis okinavensis* mean = 0.26, median = 0.19 and B) *Pantherophis guttatus* mean = 0.28, median = 0.23. On the X-Axis is the RSA and on the Y-Axis is the frequency. RSAs of positively selected sites are labeled in orange.

Posterior rates for venom CRISP of *Ovophis okinavensis* and *Pantherophis guttatus* were calculated to analyze the evolutionary rates of CRISP. Figure 24 shows histograms for the evolutionary rate frequencies in A) Viperidae and B) Colubridae to figure out if the positively selected sites are more or less fast evolving than the average site. The average posterior rate ranges around 1 but positively selected sites evolve remarkably faster ranging around a PR of 2. Co-adaptation between fast-evolving positively selected sites would be beneficial for reacting to changes in the prey's target molecules.

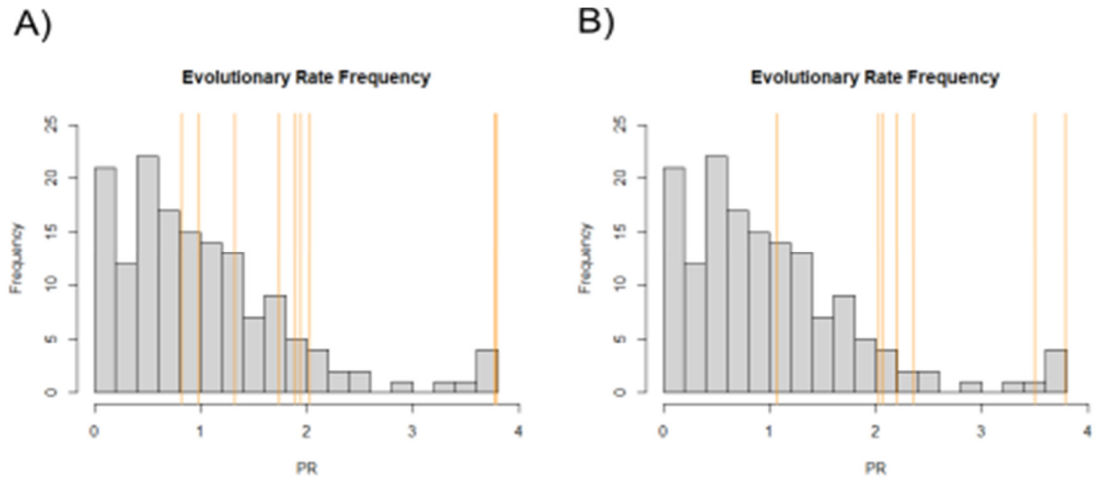

Fig.24: Histogram of the evolutionary rate frequency of CRISP for A) *Ovophis okinavensis* and B) *Pantherophis guttatus.* On the X-Axis is the posterior rate and on the Y-Axis is the frequency. PRs of positively selected sites are labeled in orange. Mean = 1.01, median = 0.83.

5.4.4 One residue of CRISP under positive selection is co-adapting with eight other sites

Coevolution analyses aimed to find coevolving groups of sites that might include the positively selected sites of Helodermatidae, Elapidae, Viperidae, and Colubridae. In total 11 groups of coevolving sites were found by four different approaches of CoMap (Table 6). Coevolving groups are visualized in the protein structure of Pantherophis guttatus in Figure 25 according to the approach that retrieved them. The 'Volume' approach (A) that accounts for the volume of the amino acids suggests two coevolving groups of each two sites. All these sites come from the first 100 amino acids of the CRISP alignment. None of them is positively selected in the four Toxicofera families. Only one coevolving group is retrieved from the 'Grantham' approach (B) which accounts for many different properties of amino acids. Here, GLN110 is coevolving with LYS40 instead of LEU44 as suggested under the 'Volume' approach. These two residues are not among the positive selected as well. The 'Polarity' approach (C) that accounts for the polarity of amino acids found two coevolving groups. One of them is the largest group found for CRISP. It consists of nine residues of which one is LYS40 which was already suggested to be coevolving with GLN110 in the 'Grantham' groups. Additionally, this large group of coevolving sites contains GLN82 which is under significant positive selection in Colubridae and its homolog ARG82 is under significant positive selection in Vipers. BEB analysis of Viperidae (Branch-Site Model 9) does in addition suggest two more positively selected sites that are in this coevolving group. LYS40 and PRO81 are among the non-significant sites in the BEB results. Sites of the second coevolving group from the 'Polarity' results are not among the positive selected. The last approach 'Simple' (D) which does not account for amino acid properties retrieves the most coevolving groups. Each two of these five groups consist of either two residues or three residues. The last group consisting of four residues is the largest of this approach. None of these residues is positively selected on any of the four analyzed families but in total four sites from three different groups were already suggested to be coevolving by other approaches. SER191 and THR51 from the purple group are additionally part of the large group suggested by 'Polarity'. GLN110 which is here coevolving with ASN114 appears in coevolving groups of 'Grantham' and 'Volume'. Finally, LYS117 from the pink group also appears in the large 'Polarity' group.

Table 6: CRISP Coevolving groups

Group	Alignment positions	Structure residues	Size	Method	p-value
1	[14;53;23;25;129;55;149;52;85]	LYS40; GLN82; ARG49; THR51; ASN171; LEU84; SER191; PRO81; LYS117	9	Polarity	0.011
2	[14;78]	LYS40; GLN110	2	Grantham	0.050
3	[18;78]	LEU44; GLN110	2	Volume	0.020
4	[21;135]	SER47; ALA177	2	Simple	0.034
5	[3;149;25]	VAL13; SER191; THR51	3	Simple	0.022
6	[33;162]	GLU59; ASN210	2	Simple	0.001
7	[4;20]	LEU14; ARG46	2	Volume	0.031
8	[46;50;121]	ALA72; HIS79; VAL163	3	Simple	0.015
9	[56;61]	ARG85; GLY93	2	Polarity	0.016
10	[78;82]	GLN110; ASN114	2	Simple	0.028
11	[85;150;117;136]	LYS117; ALA192; ARG152; THR178	4	Simple	0.025

The site under positive selection is marked in red.

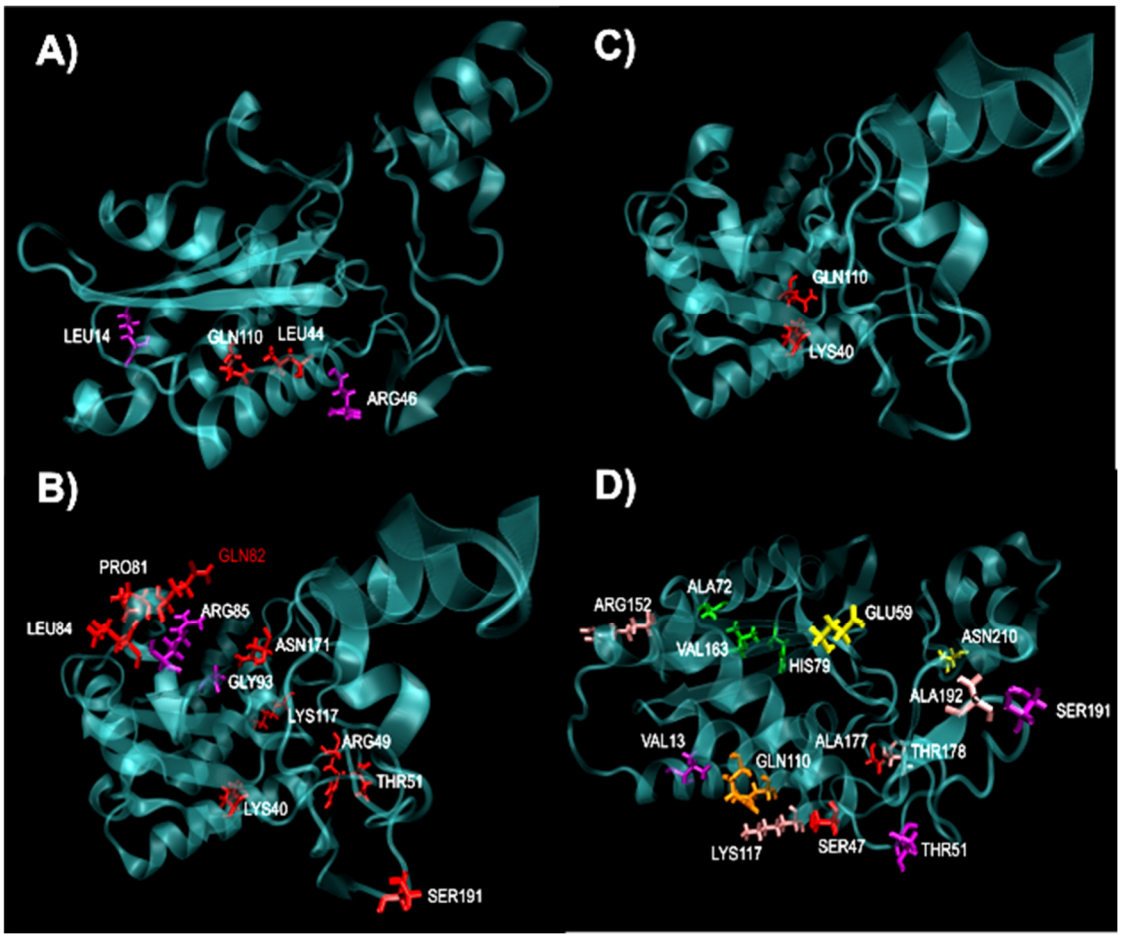

Fig.25: CRISP structure of *Pantherophis guttatus*. Coevolving groups are retrieved from CoMap approaches A) 'Volume', B) 'Grantham', C) 'Polarity', and D) 'Simple'. Coevolving groups are labeled with different colors.

Plotting the posterior rates of the coevolving groups on the histogram of the evolutionary rate frequencies (Figure 26) does not reveal any remarkable deviation from the average PR of around 1. Residues of all coevolving groups range around 1 except for one outlier that appears at the highest maximum at around 3.8 and therefore is very fast evolving. This fast-evolving outlier appears to be the only residue that is coevolving and positively selected in Colubridae and Viperidae GLN82.

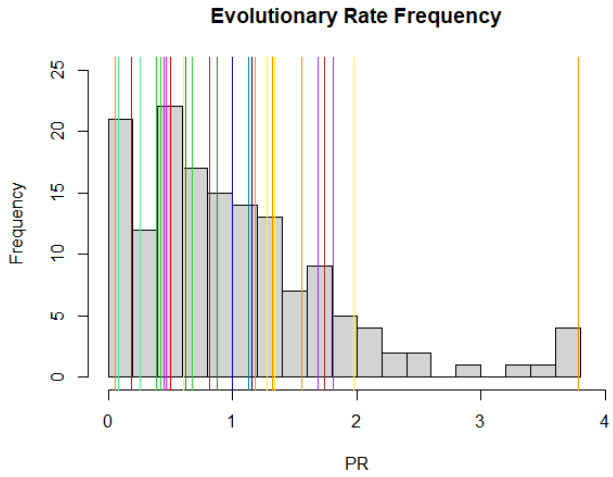

Fig.26: Histogram of the evolutionary rate frequency of venom CRISP. On the X-Axis is the posterior rate and on the Y-Axis is the frequency. PRs of coevolving sites are labeled according to coevolving groups. Mean = 1.01, median = 0.83.

In summary, CRISPs are important venom proteins that were duplicated several times. Therefore the evolution of venomousness could not be traced back on the phylogeny since the gene tree was reconciled with the species tree to handle the multiple paralogs. Selection models of codon sequence evolution were used to calculate the average dN/dS ratios for venom CRISPs of squamate clades down to the family level. These Branch-Models revealed positive selection only for elapid snakes but suggested neutral evolution for Colubridae, Viperidae, and Gekkota. Implementing Branch-Site Models that allow the dN/dS ratios to vary also between sites are able to reject the Null-Hypothesis that does not allow positive selection for Branch-Site Models 1, 4, 7-12, and 15. The extracted sites that are significantly under positive selection from the BEB analyses for Helodermatidae (1), Elapidae (1), Viperidae (10), and Colubridae (8) do overall tend to be surface residues because and are evolving faster than the average site in Viperidae and Colubridae (Figure 22). Coevolution analyses found 11 coevolving groups of sites with four different approaches that account for different properties of the amino acids. One of the coevolving sites (GLN82) of the largest groups that comes from the ‘Polarity’ approach is positively selected in Viperidae and Colubridae. In addition, two more sites of this group are suggested to be positively selected by BEB analyses of Brach-Site Model 9 but are not considered to be significant. Finally, GLN82 which is positively selected and co-adapting with eight other residues is very fast evolving (PR ~ 3.8) compared to the average (PR~1) and other coevolving sites.

6. Discussion

In the following, the results of this study are discussed according to the hypotheses in 3.1. The first major questions about multiple convergent evolution of venomousness in squamates and the impact of using different data sources for inferring phylogenies on the scenarios of venom evolution are addressed also in the context of previous studies. Afterward, the evolution of

venomousness in squamates is discussed at a molecular scale taking the molecular evolution of two venom proteins into account. That section focuses on sites under positive selection and their role in venom acquisition and function.

6.1 How many times did venomousness evolve in squamates?

Recapitulating the Toxicofera hypothesis, the best-accepted hypothesis about the evolution of venomousness in Squamata is a single early origin in the last common ancestor of monophyletic Toxicofera consisting of Iguania, Anguimorpha, and Serpentes [5]. However, detecting monophyletic Toxicofera highly depends on the data source that is used to infer the phylogeny notably morphological traits and mitochondrial or nuclear genes. Here, I discuss the impact of these data sources on the tree topology to evaluate the Toxicofera hypothesis and scenarios of multiple convergent evolution of venomousness in squamates.

6.1.1 Evaluating the impact of different data sources on the squamate phylogeny

The large-scale phylogeny that I inferred in this work (Figure 9) exhibits some similarities to the phylogenies inferred from morphological data in previous works [11–13,15,23] (see also Figure 1). A fundamental similarity is the polyphyly of Toxicofera. In Figure 1B Toxicofera are separated in Anguimorpha, Serpentes and Iguania. Furthermore, the overall position of Acrodonta that is more basal than Gekkota is consistent with the tree topologies in Figures 1A, B, and D. Additionally, a close relationship between Anguimorpha and Scincoidea is indicated in Figures 1A, C, and D. Major differences to the morphological phylogenies are the position of the root that was placed between snakes in all other squamates by the MAD approach [61] but in the morphological phylogenies mostly Iguania appears as the most basal squamates. The large-scale phylogeny provides evidence for monophyletic Scincoidea while in morphological studies it is either not monophyletic (Fig.1B-D) or does not include Xantusiidae (Fig.1A). Regarding molecular phylogenies discussed in 2.1.2, the phylogenomic approach reveals a tree topology that resembles a mixture of the mitochondrial phylogeny in Figure 2B and the phylogeny from combined nuclear and mitochondrial genes in Figure 1C. The paraphyly of Iguania splitting them into Acrodonta that are closely related to snakes and Pleurodonta or Iguanidae that appear in a closer relationship to Lacertidae is a fundamental similarity to the mitochondrial phylogeny. Besides, the polyphyly of Toxicofera appears to be a common feature of morphological, mitochondrial, and the large-scale phylogeny. A remarkable similarity to the mixed data source phylogeny in Figure 2C is the monophyly of Scincoidea that appears in every phylogeny that includes nuclear data. Only in the mitochondrial phylogeny Scincoidea are not monophyletic. This indicates mitochondrial genes of Scincoidea show a polyphyletic history and nuclear genes a monophyletic history that takes hold in phylogenies from mixed datasets even though the amount of mitochondrial data is larger as in Pyron et al. 2013 [9] and this phylogenomic approach. Another clade that appears only when nuclear data is involved is the sister relationship between Amphisbaenia and Lacertidae. Also, this is consistent in all

phylogenies that include nuclear data. Using only mitochondrial data results in paraphyletic amphisbaenians due to the basal but separated position of Rhineuridae and in a close relationship of Amphisbaenia to snakes and acrodonts. This pattern can also be found in the sub-phylogeny for only mitochondrial genes in Figure 11. Additionally, in previous studies [9,10,16] it seems that Toxicofera becomes monophyletic as soon as nuclear genes are involved. Overall it appears that nuclear genes can compensate for the effects of the mitochondrial genes on the phylogeny. But why are Toxicofera still polyphyletic in the large-scale-phylogeny even if nuclear genes are involved? This could be an artifact of the extremely biased data toward mitochondrial genes, meaning the amount of mitochondrial data is simply too big for the few nuclear data to compensate for its effects. The nuclear phylogeny in Figure 12 contains only 16 squamate species. This must be a consequence of the lack of data on the NCBI Protein database because filtering the dataset for protein families that have at least 15 non-duplicated sequences did not discard a single species. Since mixed data was only available for eight species (Figure 13) 256 out of 272 species contribute only mitochondrial data to the large-scale phylogeny. Therefore, the phylogeny may be biased too much toward mitochondrial data that nuclear data for only 16 species cannot compensate for such a big effect on a deep branch like Toxicofera. Besides the different data sources, another aspect could have affected the tree topology. Long-branch attraction was already found to have an impact on the squamate phylogeny in previous studies [8]. Here especially Gekkota, Varanidae and the unusually placed sister group of acrodonts consisting of *Anolis carolinensis* and *Sceloporus undulatus* exhibit long branches that could have led to their extraordinary positions in the phylogeny, meaning the terminal position of Gekkota and the polyphylies of Anguimorpha and Pleurodonta.

6.1.2 Evaluating the impact of the rooting on the higher-level squamate phylogeny

The minimum ancestor deviation approach [61] that was used to root the large-scale phylogeny placed the root between snakes and all other squamates. This has one major effect on the tree topology: Toxicofera are not monophyletic! All previously discussed studies from 2.1 use distinct outgroups notably Sphenodontia to confidently root the phylogenies which were found to be the closest living relative of Squamata [2,3]. Here, since Sphenodontia is genetically very divergent from squamates outgroups are not included for two major reasons: 1. To minimize the risk of long-branch attraction that can be caused by distinct outgroups. 2. To not further restrict the dataset because outgroup sequences might not be available for all used protein families. Most previous studies suggest Gekkota as the most basal taxon that is included here. But even placing the root at Gekkota does not make Toxicofera monophyletic because Teiioidea would appear within Toxicofera and Pleurodonta would not be included in Toxicofera. The position of Teiioidea is questionable anyway, since all previously discussed studies where Teiioidea and Lacertidae find a monophyletic group including these taxa. Unfortunately, the pure mitochondrial phylogeny in Figure 2B does not include Teiioidea at all. But the close relationship of *Iphisa elegans* which is the only species of Teiioidea used here and Toxicofera is consistent with

mitochondrial studies [83]. Removing *Iphisa elegans* in this study and placing the root at Gekkota would still not make Toxicofera monophyletic because of the position of Pleurodonta which resembles that of the mitochondrial phylogeny in Figure 2B. Even though the position of the root might be controversial there is still evidence for this rooting from previous studies. In the mitochondrial phylogeny in Kumazawa et al. 2004 the root was placed between snakes and all other squamates too although distinct outgroups were included [84]. However, the rooting could also have been affected by long-branch attraction notably in Gekkota.

6.1.3. Evaluating convergent evolution of venom in squamates

How many times venomousness evolved independently in Squamata does not only rely on the question if the Toxicofera hypothesis is true. Thereby, a major aspect is venomousness in snakes. In all phylogenies regardless of their data sources and monophyly of Toxicofera, all basal snake lineages of Solecophidia and Henophidia are non-venomous. Colubroidea, the only clade that contain venomous families is the youngest and, therefore, the most terminal of these three clades. But also within Colubroidea, not all species are venomous. In fact, many species of the family Colubridae are non-venomous. Because of the distribution of venomous species across the snake phylogeny, it appears likely that venomousness did not evolve before the emergence of Colubroidea, since a loss of venom in Solecophidia and Henophidia followed by recurrence in Colubroidea and a secondary loss in some Colubridae is not parsimonious. This theory is also supported by the analyses of codon sequence evolution performed here. Overall gains of venomousness should be associated with positive selection because of adaptation to target molecules in the prey while losses of venomousness should be associated with relaxation or purifying selection. Branch-Site analyses of venom CRISP found positive selection in Colubroidea compared to Pythonidae (Table 5) and retrieved sites that are significantly under positive selection in all three analyzed Colubroidea families. Additionally, Branch-Models (Table 4) detect signals of positive selection arising in Colubroidea only after separating them from Pythonidae. In snakes overall, these models indicate purifying selection similar to all non-venomous clades. Additional convergent origins of venomousness in squamates are more difficult to detect without a reliable phylogeny because it is still under debate which families of Anguimorpha and Iguania are venomous at all. Anyhow, the position of venom glands in the three distinct toxicoferan lineages was originally used as evidence for the Toxicofera hypothesis [7]. But it can also serve as evidence for convergent origins of venom in these clades because the position of venom glands contradicts the basal position of snakes in Toxicofera which is consistent in the previously discussed studies that find monophyletic Toxicofera [5,9,14]. It is argued that the constitution of Iguania venom glands is the initial condition and snakes and anguimorphs lost their venom glands in either the lower or upper jaw with time [7] but if snakes are the most basal Toxicofera the initial condition would be the presence of venom glands only in the upper jaw. Another evidence for the convergent evolution of venom in Anguimorpha and Serpentes is that according to the positive selection analyses of PLA_2 this toxin is only under

positive selection in Varanidae not in snakes indicating independent recruitment of PLA_2 into the venom of Varanidae. Furthermore, this study did not find evidence for positive selection of venom protein in Iguania at all even though acrodonts and snakes appear to be closely related regardless of the data source of the phylogenies. An aspect against this argument is that the lack of positive selection could be due to their mostly insectivore or herbivore diet and the resulting low expression levels of toxin genes [21]. On a side note, convergent evolution of venom of the overall same composition has already happened frequently across the whole animal kingdom [35,42,43,49].

6.2 The molecular basis of venom acquisition

Venom is composed of different proteins that are recruited into the venom from their original function after gene duplication [39,41]. Thus, homologs of venom proteins are also found in non-venomous taxa. Since these proteins are essential for venomous species I expect to find differences in selection pressure acting on venom proteins between venomous and non-venomous squamates. Sites that are interacting with the target molecules of the prey are necessary for the toxin function and therefore might be under positive selection because they constantly need to adapt to changes in the target molecules [35–37,43,46]. Here, I discuss the molecular evolution of sites under positive selection and their role in toxin function and venom acquisition in squamates.

6.2.1. Sites under positive selection ensure the toxin function

Sites involved in the evolution of venomousness in squamates must have been under positive selection in venomous clades. Identifying these sites can help to find mutations that lead to the toxic function of the protein in venomous squamates compared to non-venomous squamates. The performed Branch-Models in Tables 1 and 4 indicate signals of positive selection only at the family level. At higher taxonomic levels these signals could be masked by overall purifying selection in non-venomous species that are included in Toxicofera, Iguania, and Serpentes or by regions under purifying selection since Branch-Models calculate the average dN/dS ratio of all sites in the sequences of a clade. A reason why positive selection could not be detected in venomous lineages compared to non-venomous lineages regardless of their systematic classification can be the underlying phylogeny. In the case of PLA_2, the polyphyly of venomous taxa could have caused a problem. In CRISP ortholog branches might have an impact on the Branch-Models because orthologs of the same species were grouped together in these analyses. Therefore, mutations between orthologs could have affected the posterior probabilities of these sites leading to the underestimation of dN/dS ratios. Evidence for multiple convergent recruitments of CRISP can be found in previous studies [35,49]. Another concern of the underlying phylogenies that could have affected the outcome of selection models of codon evolution can be long branch attraction notably in *Varanus komodoensis* and *Gekko japonicus*. These taxa were found to exhibit very long branches elsewhere [26]. Branch-Site analyses

account for the problem of regions under purifying selection because they calculate the dN/dS ratio for every site of the foreground lineage. Sites of venom CRISP that are significantly under positive selection could be identified for Toxicofera, Serpentes, Helodermatidae, Elapidae, Viperidae, Colubridae, and Colubroidea compared to Pythonidae. Additionally, positive selection was detected in venomous snakes compared to non-venomous snakes regardless of their systematic classification. PLA_2 sites under positive selection were only found in Varanidae. Because signals of positive selection arise only in venomous families and these families bear sites that are significantly under positive selection it can be assumed that these sites contribute to the venom function. Comparing the positively selected sites of venom CRISP reveals not always the same set of sites is positively selected. Only one site of Helodermatidae and Viperidae matches and between Viperidae and Colubridae only two sites match. Although sets of positively selected sites differ between snake families sets of higher-level taxa are overall composed of the sets of subordinated taxa. This indicates sets of positively selected sites are family specific, and therefore might constitute adaptations to the target molecules in different prey. Furthermore, because a much bigger proportion of positively selected sites in venomous snakes compared to non-venomous snakes is under positive selection in the purely venomous Viperidae but not in Colubridae which contain only one venomous species the set of sites from venomous snakes likely contributes to the toxin function.

6.2.2 Sites involved in adaptation evolve fast and are located close to the protein surface

Because positively selected sites are those sites that ensure the toxin function and interact with the target molecules in the prey they are assumed to be fast evolving surface residues. In reality, half of the positively selected sites in varanid PLA_2 are buried (Figure 15) according to the threshold of 0.25 that is suggested in the literature [82]. But in CRISPs of vipers and colubrids positively selected sites are more exposed to the protein surface than the average (Figure 23) and can be considered surface residues. The reason why I do not consider positively selected sites in varanid PLA_2 surface proteins although they have similar RSA values is that the average RSA in PLA_2 is with a value of 0.36 higher than the suggested threshold of 0.25 in the literature and only half of the sites under positive selection are more than 36% accessible to the solvent. Surface residues tend to evolve faster than buried residues [85]. However, sites under positive selection in PLA_2 do not evolve faster than the average (Figure 17). In contrast, sites under positive selection in venom CRISPs of Viperidae and Colubridae evolve on average faster than other sites. These high evolutionary rates could be caused by the need to rapidly adapt to changes in the target molecule of the prey since among the fastest evolving positively selected sites are surface residues that most likely interact with other molecules.

6.2.3. Does the functional role of sites under positive selection lead to co-adaptation?

Sites under positive selection do neither in PLA_2 nor in CRISP tend to co-adapt with other residues. Only one residue of each protein appears in a coevolving group of sites. Both of these sites are surface residues. All sites they are co-adapting with have lower evolutionary rates than the average (Figure 19 – yellow group, Figure 26 – orange group). But the coevolving positively selected site in PLA_2 is co-adapting with other surface residues because their RSA values range from 0.64 to 0.72 which is higher than the average. The positively selected site GLN82 in colubrid CRISP appears in a coevolving group that was retrieved with the approach that accounts for the polarity of amino acids, meaning substitutions between sites of different polarity are weighted differently. This makes a polarity change of the residue likely. But why does this residue need to change its polarity? Maybe the residue needs to react to polarity changes in the target molecule of the prey. The reason why no coevolving group was found with the approach that accounts for the charge of amino acids is that charge plays a role in correlated mutations that compensate for deleterious mutations since mutations that change the charge of a site tend to be compensated for by neighboring amino acids [86]. This is why this approach does not find co-adapting groups but usually retrieves groups of sites that preserve deleterious mutations in the population because they coevolve with a compensating mutation. Finally, because positively selected sites do not co-adapt with residues from the same protein, it could be possible that these sites are co-adapting with the target molecules of the prey because of the constant need of venomous species to adapt the venom to changes in the prey.

7. Conclusion

i) The phylogenetic analyses indicate multiple convergent evolutions of venomousness in Squamata because Toxicofera is not monophyletic in the large-scale phylogeny. One of the origins of venomousness could be confidently traced back to the last common ancestor of Colubroidea snakes because all more basal snakes are non-venomous according to all phylogenies regardless of the data source. Further origins of venomousness highly depend on the phylogeny and therefore on the data source, taxon sampling, rooting, and the effects of long-branch-attraction. Additionally, it is important to reliably identify more venomous families in Anguimorpha and Iguania. Anyhow, the convergent evolution of venom in Colubroidea, Anguimorpha, and Iguania appears to be likely, because of the conflict between venom gland morphology and the basal position of snakes in Toxicofera. Furthermore, multiple convergent evolutions of venom with an overall same composition of toxins make convergent evolution more likely than a loss of venomousness in basal snakes followed by reoccurrence in Colubroidea and secondary losses in several Colubridae. Following studies would first need to identify more venomous families in Anguimorpha and Iguania to find the most parsimonious scenario of venom evolution in squamates. An actual reliable phylogeny can probably only be obtained by using full genomes.

ii) Positive selection could be detected in venomous clades for both analyses of venom proteins. In Phospholipase A_2 positive selection was significant only in Varanidae while in cysteine-rich secretory protein positive selection was significant in snakes and Helodermatidae. This indicates these sites are interacting with the target molecules in the prey and are therefore essential for toxin function. Overall no positive selection could be found in Iguania. This could be on one hand due to their mostly insectivore or herbivore diet or because venom in iguanas evolved independently of those of other toxicoferans.

iii) Positively selected sites are only partially fast-evolving surface residues. In PLA_2 half of the sites under positive selection are considered surface residues but in CRISP all except for one site are considered surface residues because their RSA values are higher than the overall very low average. Only positively selected sites in CRISP evolve faster than other sites. In PLA_2 sites under positive selection do not evolve faster than the average. High evolutionary rates combined with locations close to the surface give evidence these sites need to rapidly adapt to changes in the target molecule in the prey.

v) One fast-evolving surface residue under positive selection of each venom protein belongs to a co-adapting group of sites. It could be possible that other positively selected sites are co-adapting instead with the target molecules because of the constant need of venomous species to adapt their venom to changes in the prey.

8. Acknowledgements

I thank Julien Dutheil for his supervision, teaching, and encouragement. Thank you for making this project I am passionate about for already so many years possible. Thank you for teaching me all these methods and for encouraging me in coding.

To Tal Dagan, thank you for walking me through the masters's program, everything I have learned in your courses, and for examining my thesis.

I want to thank the Christian-Albrechts-University of Kiel and the Max Planck Society for cooperating and enabling projects apart from the university.

I would like to thank the whole systems biology lab. Thank you Fernanda Trancoso, Nataša Puzović, and Jinyang Liang for integrating and supporting me.

Special thanks to my parents and fur babies who always supported, encouraged, and listened to me. Thank you for your advice and love!

Finally, I want to thank my husband and his "magical IT superpowers" for providing me with an incredible technical setup in my home office which is essential these days. Thank you for all your love and patience!

9. References

1. Simões TR, Pyron RA. THE SQUAMATE TREE OF LIFE. Bulletin of the Museum of Comparative Zoology. 2021;163. doi:10.3099/0027-4100-163.2.47

2. Pincheira-Donoso D, Bauer AM, Meiri S, Uetz P. Global Taxonomic Diversity of Living Reptiles. Fontaneto D, editor. PLoS ONE. 2013;8: e59741. doi:10.1371/journal.pone.0059741

3. Gemmell NJ, Rutherford K, Prost S, Tollis M, Winter D, Macey JR, et al. The tuatara genome reveals ancient features of amniote evolution. Nature. 2020;584: 403–409. doi:10.1038/s41586-020-2561-9

4. Pasquesi GIM, Adams RH, Card DC, Schield DR, Corbin AB, Perry BW, et al. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. Nat Commun. 2018;9: 2774. doi:10.1038/s41467-018-05279-1

5. Vidal N, Hedges SB. The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes. Comptes Rendus Biologies. 2005;328: 1000–1008. doi:10.1016/j.crvi.2005.10.001

6. Vidal N, Hedges SB. The molecular evolutionary tree of lizards, snakes, and amphisbaenians. Comptes Rendus Biologies. 2009;332: 129–139. doi:10.1016/j.crvi.2008.07.010

7. Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SFR, et al. Early evolution of the venom system in lizards and snakes. Nature. 2006;439: 584–588. doi:10.1038/nature04328

8. Burbrink FT, Grazziotin FG, Pyron RA, Cundall D, Donnellan S, Irish F, et al. Interrogating Genomic-Scale Data for Squamata (Lizards, Snakes, and Amphisbaenians) Shows no Support for Key Traditional Morphological Relationships. Thomson R, editor. Systematic Biology. 2020;69: 502–520. doi:10.1093/sysbio/syz062

9. Pyron R, Burbrink FT, Wiens JJ. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evol Biol. 2013;13: 93. doi:10.1186/1471-2148-13-93

10. Reeder TW, Townsend TM, Mulcahy DG, Noonan BP, Wood PL, Sites JW, et al. Integrated Analyses Resolve Conflicts over Squamate Reptile Phylogeny and Reveal Unexpected Placements for Fossil Taxa. Wilf P, editor. PLoS ONE. 2015;10: e0118199. doi:10.1371/journal.pone.0118199

11. Estes R, de Queiroz K, Gauthier J. Phylogenetic Relationships within Squamata. Phylogenetic relationships of the lizard families. 1988. pp. 119–281.

12. Lee M. Squamate phylogeny, taxon sampling, and data congruence. Organisms Diversity & Evolution. 2005;5: 25–45. doi:10.1016/j.ode.2004.05.003

13. Lee MSY. Convergent evolution and character correlation in burrowing reptiles: towards a resolution of squamate relationships. Biological Journal of the Linnean Society. 1998;65: 369–453. doi:10.1111/j.1095-8312.1998.tb01148.x

14. Böhme MU, Fritzsch G, Tippmann A, Schlegel M, Berendonk TU. The complete mitochondrial genome of the Green Lizard Lacerta viridis viridis (Reptilia: Lacertidae) and

its phylogenetic position within squamate reptiles. Gene. 2007;394: 69–77. doi:10.1016/j.gene.2007.02.006

15. Conrad JL. Phylogeny And Systematics Of Squamata (Reptilia) Based On Morphology. Bulletin of the American Museum of Natural History. 2008;310: 1–182. doi:10.1206/310.1

16. Albert EM, San Mauro D, García-París M, Rüber L, Zardoya R. Effect of taxon sampling on recovering the phylogeny of squamate reptiles based on complete mitochondrial genome and nuclear gene sequence data. Gene. 2009;441: 12–21. doi:10.1016/j.gene.2008.05.014

17. Grechko VV. The problems of molecular phylogenetics with the example of squamate reptiles: Mitochondrial DNA markers. Mol Biol. 2013;47: 55–74. doi:10.1134/S0026893313010056

18. McMahan CD, Freeborn LR, Wheeler WC, Crother BI. Forked Tongues Revisited: Molecular Apomorphies Support Morphological Hypotheses of Squamate Evolution. Copeia. 2015;103: 525–529. doi:10.1643/CH-14-015

19. Lillywhite HB. How snakes work: structure, function and behavior of the world's snakes. Oxford ; New York: Oxford University Press; 2014.

20. Thornton SL. Snakes. Encyclopedia of Toxicology. Elsevier; 2014. pp. 310–312. doi:10.1016/B978-0-12-386454-3.00786-7

21. Fry BG, Undheim EAB, Ali SA, Jackson TNW, Debono J, Scheib H, et al. Squeezers and Leaf-cutters: Differential Diversification and Degeneration of the Venom System in Toxicoferan Reptiles. Molecular & Cellular Proteomics. 2013;12: 1881–1899. doi:10.1074/mcp.M112.023143

22. Mongiardino Koch N, Gauthier JA. Noise and biases in genomic data may underlie radically different hypotheses for the position of Iguania within Squamata. Hejnol A, editor. PLoS ONE. 2018;13: e0202729. doi:10.1371/journal.pone.0202729

23. Simões TR, Caldwell MW, Tałanda M, Bernardi M, Palci A, Vernygora O, et al. The origin of squamates revealed by a Middle Triassic lizard from the Italian Alps. Nature. 2018;557: 706–709. doi:10.1038/s41586-018-0093-3

24. Broschinski A. Ein Lacertilier (Scincomorpha, Paramacellodidae) aus dem Oberen Jura von Tendaguru (Tansania). Foss Rec. 1999;2: 155–158. doi:10.1002/mmng.1999.4860020111

25. John C, Murphy. Secrets of the Snake Charmer. Bertrams Print On Demand; 2010. Available: http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9781450221276

26. Castoe TA, de Koning APJ, Kim H-M, Gu W, Noonan BP, Naylor G, et al. Evidence for an ancient adaptive episode of convergent molecular evolution. Proc Natl Acad Sci USA. 2009;106: 8986–8991. doi:10.1073/pnas.0900233106

27. Vawter L, Brown WM. Nuclear and Mitochondrial DNA Comparisons Reveal Extreme Rate Variation in the Molecular Clock. Science. 1986;234: 194–196. doi:10.1126/science.3018931

28. Moore WS. INFERRING PHYLOGENIES FROM MT DNA VARIATION: MITOCHONDRIAL-GENE TREES VERSUS NUCLEAR-GENE TREES. Evolution. 1995;49: 718–726. doi:10.1111/j.1558-5646.1995.tb02308.x

29. Rokas A, Carroll SB. More Genes or More Taxa? The Relative Contribution of Gene Number and Taxon Number to Phylogenetic Accuracy. Molecular Biology and Evolution. 2005;22: 1337–1344. doi:10.1093/molbev/msi121

30. Bordon K de CF, Cologna CT, Fornari-Baldo EC, Pinheiro-Júnior EL, Cerni FA, Amorim FG, et al. From Animal Poisons and Venoms to Medicines: Achievements, Challenges and Perspectives in Drug Discovery. Front Pharmacol. 2020;11: 1132. doi:10.3389/fphar.2020.01132

31. Koludarov I, Jackson TN, Brouw B op den, Dobson J, Dashevsky D, Arbuckle K, et al. Enter the Dragon: The Dynamic and Multifunctional Evolution of Anguimorpha Lizard Venoms. Toxins. 2017;9: 242. doi:10.3390/toxins9080242

32. Fry BG, Wroe S, Teeuwisse W, van Osch MJP, Moreno K, Ingle J, et al. A central role for venom in predation by *Varanus komodoensis* (Komodo Dragon) and the extinct giant *Varanus* (*Megalania*) *priscus*. Proc Natl Acad Sci USA. 2009;106: 8969–8974. doi:10.1073/pnas.0810883106

33. Fry BG, Casewell NR, Wüster W, Vidal N, Young B, Jackson TNW. The structural and functional diversification of the Toxicofera reptile venom system. Toxicon. 2012;60: 434–448. doi:10.1016/j.toxicon.2012.02.013

34. Barlow A, Pook CE, Harrison RA, Wüster W. Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. Proc R Soc B. 2009;276: 2443–2449. doi:10.1098/rspb.2009.0048

35. Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, et al. The Toxicogenomic Multiverse: Convergent Recruitment of Proteins Into Animal Venoms. Annu Rev Genom Hum Genet. 2009;10: 483–511. doi:10.1146/annurev.genom.9.081307.164356

36. Župunski V, Kordiš D. Strong and widespread action of site-specific positive selection in the snake venom Kunitz/BPTI protein family. Sci Rep. 2016;6: 37054. doi:10.1038/srep37054

37. Sunagar K, Johnson WE, O'Brien SJ, Vasconcelos V, Antunes A. Evolution of CRISPs Associated with Toxicoferan-Reptilian Venom and Mammalian Reproduction. Molecular Biology and Evolution. 2012;29: 1807–1822. doi:10.1093/molbev/mss058

38. Yeang C-H, Haussler D. Detecting Coevolution in and among Protein Domains. Rzhetsky A, editor. PLoS Comput Biol. 2007;3: e211. doi:10.1371/journal.pcbi.0030211

39. Fry BG. 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. 2005;15: 403–420. doi:10.1101/gr.3228405

40. Doley R, Kini RM. Protein complexes in snake venom. Cell Mol Life Sci. 2009;66: 2851–2871. doi:10.1007/s00018-009-0050-2

41. Casewell NR, Huttley GA, Wüster W. Dynamic evolution of venom proteins in squamate reptiles. Nat Commun. 2012;3: 1066. doi:10.1038/ncomms2065

42. Burke JE, Dennis EA. Phospholipase A2 Biochemistry. Cardiovasc Drugs Ther. 2009;23: 49–59. doi:10.1007/s10557-008-6132-9

43. Lynch VJ. Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. BMC Evolutionary Biology. 2007;7: 2. doi:10.1186/1471-2148-7-2

44. Manjunatha Kini R. Excitement ahead: structure, function and mechanism of snake venom phospholipase A2 enzymes. Toxicon. 2003;42: 827–840. doi:10.1016/j.toxicon.2003.11.002

45. Aparna K, Priyakumari CJ, Kezia JG. HOMOLOGY MODELLING, SECONDARY STRUCTURE PREDICTION AND PHYLOGENETIC COMPARATIVE ANALYSIS OF PHOSPHOLIPASE A2FROM DIFFERENT ORGANISMS K. Aparna*, C. Joyce Priyakumari and J Gladies Kezia Bioinformatics Centre of BTIS net, Madras Christian College, Chennai, Tamil Nadu, India ABSTRACT: Phospholipase A2. 2015.

46. Aird SD, Aggarwal S, Villar-Briones A, Tin MM-Y, Terada K, Mikheyev AS. Snake venoms are integrated systems, but abundant venom proteins evolve more rapidly. BMC Genomics. 2015;16: 647. doi:10.1186/s12864-015-1832-6

47. Hariprasad G, Srinivasan A, Singh R. Structural and phylogenetic basis for the classification of group III phospholipase A2. J Mol Model. 2013;19: 3779–3791. doi:10.1007/s00894-013-1913-x

48. Yamazaki Y, Morita T. Structure and function of snake venom cysteine-rich secretory proteins. Toxicon. 2004;44: 227–231. doi:10.1016/j.toxicon.2004.05.023

49. Tadokoro T, M. Modahl C, Maenaka K, Aoki-Shioi N. Cysteine-Rich Secretory Proteins (CRISPs) from Venomous Snakes: An Overview of the Functional Diversity in a Large and Underappreciated Superfamily. Toxins. 2020;12: 175. doi:10.3390/toxins12030175

50. Wang F, Li H, Liu M, Song H, Han H, Wang Q, et al. Structural and functional analysis of natrin, a venom protein that targets various ion channels. Biochemical and Biophysical Research Communications. 2006;351: 443–448. doi:10.1016/j.bbrc.2006.10.067

51. Yang Z, Nielsen R. Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. Molecular Biology and Evolution. 2000;17: 32–43. doi:10.1093/oxfordjournals.molbev.a026236

52. Mugal CF, Wolf JBW, Kaj I. Why Time Matters: Codon Evolution and the Temporal Dynamics of dN/dS. Molecular Biology and Evolution. 2014;31: 212–231. doi:10.1093/molbev/mst192

53. Miele V, Penel S, Duret L. Ultra-fast sequence clustering from similarity networks with SiLiX. BMC Bioinformatics. 2011;12: 116. doi:10.1186/1471-2105-12-116

54. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004;32: 1792–1797. doi:10.1093/nar/gkh340

55. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7: 539. doi:10.1038/msb.2011.75

56. Guéguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, et al. Bio++: Efficient Extensible Libraries and Tools for Computational Molecular Evolution. Molecular Biology and Evolution. 2013;30: 1745–1750. doi:10.1093/molbev/mst097

57. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30: 1312–1313. doi:10.1093/bioinformatics/btu033

58. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How Many Bootstrap Replicates Are Necessary? Journal of Computational Biology. 2010;17: 337–354. doi:10.1089/cmb.2009.0179

59. Lemoine F, Domelevo Entfellner J-B, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature. 2018;556: 452–456. doi:10.1038/s41586-018-0043-0

60. Felsenstein J. CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP. Evolution. 1985;39: 783–791. doi:10.1111/j.1558-5646.1985.tb00420.x

61. Tria FDK, Landan G, Dagan T. Phylogenetic rooting using minimal ancestor deviation. Nat Ecol Evol. 2017;1: 0193. doi:10.1038/s41559-017-0193

62. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Schwartz R, editor. Bioinformatics. 2019;35: 526–528. doi:10.1093/bioinformatics/bty633

63. Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. 2021. Available: https://CRAN.R-project.org/package=dplyr

64. Henry L, Wickham H. purrr: Functional Programming Tools. 2020. Available: https://CRAN.R-project.org/package=purrr

65. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. Available: https://ggplot2.tidyverse.org

66. Charif D, Lobry JR. SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In: Bastolla U, Porto M, Roman HE, Vendruscolo M, editors. Structural Approaches to Sequence Evolution. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. pp. 207–232. doi:10.1007/978-3-540-35306-5_10

67. Gouy M, Tannier E, Comte N, Parsons DP. Seaview Version 5: A Multiplatform Software for Multiple Sequence Alignment, Molecular Phylogenetic Analyses, and Tree Reconciliation. In: Katoh K, editor. Multiple Sequence Alignment. New York, NY: Springer US; 2021. pp. 241–260. doi:10.1007/978-1-0716-1036-7_15

68. Louca S, Doebeli M. Efficient comparative phylogenetics on large trees. Valencia A, editor. Bioinformatics. 2018;34: 1053–1055. doi:10.1093/bioinformatics/btx701

69. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25: 1422–1423. doi:10.1093/bioinformatics/btp163

70. Löytynoja A. Phylogeny-aware alignment with PRANK. In: Russell DJ, editor. Multiple Sequence Alignment Methods. Totowa, NJ: Humana Press; 2014. pp. 155–170. doi:10.1007/978-1-62703-646-7_10

71. Sela I, Ashkenazy H, Katoh K, Pupko T. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res. 2015;43: W7–W14. doi:10.1093/nar/gkv318

72. Castresana J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Molecular Biology and Evolution. 2000;17: 540–552. doi:10.1093/oxfordjournals.molbev.a026334

73. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution. 2007;24: 1586–1591. doi:10.1093/molbev/msm088

74. Zhang J. Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. Molecular Biology and Evolution. 2005;22: 2472–2479. doi:10.1093/molbev/msi237

75. Yang Z. Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. Molecular Biology and Evolution. 2005;22: 1107–1118. doi:10.1093/molbev/msi097

76. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589. doi:10.1038/s41586-021-03819-2

77. Humphrey W, Dalke A, Schulten K. VMD – Visual Molecular Dynamics. Journal of Molecular Graphics. 1996;14: 33–38.

78. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22: 2577–2637. doi:10.1002/bip.360221211

79. Dutheil J, Galtier N. Detecting groups of co-evolving positions in a molecule: a clustering approach. BMC Evol Biol. 2007;7: 242. doi:10.1186/1471-2148-7-242

80. Comte N, Morel B, Hasić D, Guéguen L, Boussau B, Daubin V, et al. Treerecs: an integrated phylogenetic tool, from sequences to reconciliations. Ponty Y, editor. Bioinformatics. 2020;36: 4822–4824. doi:10.1093/bioinformatics/btaa615

81. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021;373: 871–876. doi:10.1126/science.abj8754

82. Levy ED. A Simple Definition of Structural Regions in Proteins and Its Use in Analyzing Interface Evolution. Journal of Molecular Biology. 2010;403: 660–670. doi:10.1016/j.jmb.2010.09.028

83. Vacher J-P, Manzi S, Rodrigues MT, Fouquet A. The complete mitochondrial genome of *Iphisa elegans* (Reptilia: Squamata: Gymnophthalmidae). Mitochondrial DNA Part B. 2020;5: 3088–3090. doi:10.1080/23802359.2020.1797549

84. Kumazawa Y. Mitochondrial DNA Sequences of Five Squamates: Phylogenetic Affiliation of Snakes. DNA Research. 2004;11: 137–144. doi:10.1093/dnares/11.2.137

85. Goldman N, Thorne JL, Jones DT. Assessing the Impact of Secondary Structure and Solvent Accessibility on Protein Evolution. Genetics. 1998;149: 445–458. doi:10.1093/genetics/149.1.445

86. Neher E. How frequent are correlated changes in families of protein sequences? Proc Natl Acad Sci USA. 1994;91: 98–102. doi:10.1073/pnas.91.1.98

10. Supplementary material

10.1 Supplementary table 1: Dataset evaluation of 81 possible combinations of sequences identity and alignment coverage

Identity	Coverage	nb_families	nb_families ≥ 4seqs	nb_families_nodup	nb_families_nodup ≥ 4seqs	sum_nb_seq ≥ 4seqs	sum_nb_seq_nodup	sum_nb_seq_nodup≥ 4seqs	mean_nb_seq	mean_nb_seq ≥ 4seqs	mean_nb_seq_nodup	mean_nb_seq_nodup ≥ 4seqs	mean_nb_seq_ratio	median_nb_seq_ratio	ratio_means_nb_seq	ratio_medians_nb_seq	mean_nb_seq_ratio ≥ 4seqs	median_nb_seq_ratio ≥ 4seqs	ratio_means_nb_seq ≥ 4seqs	ratio_medians_nb_seq ≥ 4seqs
0.1	0.1	12999	9679	10655	5748	583372	56868	50041	45.16516655	60.27192892	5.337212576	8.705810717	0.476632157	0.318181818	0.118170993	0.129032258	0.329897087	0.25	0.144442212	0.205128205
0.1	0.2	14696	9829	12337	5861	581783	59504	51042	39.94978225	59.19045681	4.82321472	8.708752773	0.540735535	0.423076923	0.12073194	0.107142857	0.33112412	0.25	0.147131028	0.205128205
0.1	0.3	17689	10212	15169	6097	578886	64509	53236	33.19023122	56.68683901	4.2526864	8.731507299	0.611320099	0.777777778	0.128130665	0.1	0.335297946	0.25	0.154030591	0.210526316
0.1	0.4	21547	10815	18774	6502	575151	71668	56841	27.24750545	53.18085992	3.817407052	8.74207936	0.668349142	1	0.140101157	0.25	0.342316624	0.259259259	0.164383941	0.210526316
0.1	0.5	26044	11506	23004	7003	570694	80505	61514	22.54269697	49.59968712	3.499608764	8.783949736	0.711899294	1	0.15524357	0.5	0.352575918	0.269230769	0.177096878	0.222222222
0.1	0.6	31402	12305	28106	7632	565250	91300	67233	18.69632507	45.93661113	3.248416708	8.809355346	0.7461169	1	0.173746268	1	0.365183568	0.28125	0.19177199	0.228571429
0.1	0.7	38712	13177	35114	8355	557627	104961	73951	15.16589171	42.31820596	2.989149627	8.851107121	0.7800559	1	0.197096859	1	0.381828114	0.296296296	0.20915601	0.242424242
0.1	0.8	49520	14148	45443	9080	546218	121798	80071	11.85585622	38.60743568	2.68023678	8.81839207	0.816398153	1	0.226068597	1	0.404929825	0.32	0.228411753	0.258064516
0.1	0.9	70280	16029	64997	10369	524121	151969	89841	8.353756403	32.69829684	2.338092527	8.664384222	0.856036984	1	0.27988517	1	0.447233776	0.363636364	0.26497968	0.307692308
0.2	0.1	17893	11271	15204	6877	579683	71369	60711	32.81182585	51.43137255	4.69409366	8.828122728	0.574750668	0.55	0.143061032	0.130434783	0.350757542	0.269230769	0.171648593	0.216216216
0.2	0.2	17893	11271	15204	6877	579683	71369	60711	32.81182585	51.43137255	4.69409366	8.828122728	0.574750668	0.55	0.143061032	0.130434783	0.350757542	0.269230769	0.171648593	0.216216216
0.2	0.3	19689	11392	16996	6975	577888	74166	61697	29.81878206	50.72752809	4.363732643	8.845448029	0.614843302	0.777777778	0.146341746	0.105263158	0.352817722	0.272727273	0.174371754	0.216216216
0.2	0.4	22804	11649	19975	7176	574620	79007	63477	25.74557095	49.3278393	3.955294118	8.845735786	0.664502086	1	0.15363008	0.2	0.356214355	0.275862069	0.179325426	0.222222222
0.2	0.5	26697	11994	23656	7427	570469	85178	65832	21.99130989	47.56286477	3.60069327	8.86387505	0.707985735	1	0.163732551	0.5	0.361986848	0.28	0.186361253	0.222222222
0.2	0.6	31728	12572	28466	7884	565177	94111	69873	18.5042234	44.95521794	3.306084452	8.862633181	0.743586682	1	0.17866648	1	0.370875493	0.285714286	0.197143593	0.228571429
0.2	0.7	38818	13261	35230	8437	557594	105894	74825	15.12447833	42.04765855	3.005790519	8.868673699	0.779450134	1	0.198736806	1	0.383828848	0.296296296	0.210919561	0.242424242
0.2	0.8	49551	14176	45479	9107	546213	122077	80336	11.84843898	38.53082675	2.684249874	8.821346217	0.816129486	1	0.22654882	1	0.405163516	0.32	0.228942563	0.258064516
0.2	0.9	70285	16033	65002	10373	524120	152000	89870	8.353162126	32.69007672	2.338389588	8.663838812	0.856016252	1	0.279940644	1	0.447241217	0.363636364	0.265029626	0.307692308
0.3	0.1	23752	12985	20861	8391	574883	91206	75856	24.71800269	44.27285329	4.372081875	9.040162078	0.644632918	1	0.176878445	0.066666667	0.382509373	0.3	0.204191991	0.235294118
0.3	0.2	23752	12985	20861	8391	574883	91206	75856	24.71800269	44.27285329	4.372081875	9.040162078	0.644632918	1	0.176878445	0.066666667	0.382509373	0.3	0.204191991	0.235294118
0.3	0.3	23752	12985	20861	8391	574883	91206	75856	24.71800269	44.27285329	4.372081875	9.040162078	0.644632918	1	0.176878445	0.066666667	0.382509373	0.3	0.204191991	0.235294118
0.3	0.4	25782	13123	22865	8522	572758	94464	77051	22.77177876	43.64535548	4.131379838	9.041422201	0.672080555	1	0.181425434	0.2	0.384801169	0.3	0.207156571	0.235294118
0.3	0.5	29124	13352	26050	8728	569119	99774	79071	20.15870073	42.62425105	3.830095969	9.059463795	0.707185791	1	0.189997164	0.5	0.389242654	0.307692308	0.212542475	0.242424242
0.3	0.6	33686	13726	30429	9048	564163	107087	81868	17.42866473	41.10177765	3.519241513	9.048187445	0.741079404	1	0.201922612	1	0.395952306	0.3125	0.220141025	0.25
0.3	0.7	40304	14190	36757	9408	556893	116647	84879	14.566842	39.24545455	3.173463558	9.022002551	0.776154138	1	0.217855288	1	0.404271513	0.321428571	0.22988656	0.258064516
0.3	0.8	50572	14830	46536	9781	545744	129523	87280	11.6092304	36.8	2.783286058	8.923422963	0.813424374	1	0.23974768	1	0.417915764	0.333333333	0.24248432	0.266666667
0.3	0.9	70811	16354	65526	10724	523851	155865	93531	8.291112963	32.03197994	2.378674114	8.721652369	0.855429717	1	0.286894428	1	0.453915717	0.375	0.272279528	0.307692308
0.4	0.1	30944	14750	27735	10051	568309	113670	92387	18.97304809	38.52942373	4.098431585	9.191821709	0.698840635	1	0.216013345	0.5	0.412968419	0.333333333	0.238566291	0.290322581
0.4	0.2	30944	14750	27735	10051	568309	113670	92387	18.97304809	38.52942373	4.098431585	9.191821709	0.698840635	1	0.216013345	0.5	0.412968419	0.333333333	0.238566291	0.290322581
0.4	0.3	30944	14750	27735	10051	568309	113670	92387	18.97304809	38.52942373	4.098431585	9.191821709	0.698840635	1	0.216013345	0.5	0.412968419	0.333333333	0.238566291	0.290322581
0.4	0.4	30944	14750	27735	10051	568309	113670	92387	18.97304809	38.52942373	4.098431585	9.191821709	0.698840635	1	0.216013345	0.5	0.412968419	0.333333333	0.238566291	0.290322581
0.4	0.5	33088	14863	29863	10176	566005	117244	93744	17.74365329	38.08147749	3.926062351	9.212264151	0.718020061	1	0.221265727	0.5	0.415719265	0.333333333	0.241909315	0.3
0.4	0.6	37049	15098	33684	10421	561566	123424	95921	15.84663554	37.19472778	3.664172901	9.204586892	0.745493813	1	0.231227183	1	0.421295656	0.346153846	0.247470205	0.3
0.4	0.7	43234	15420	39620	10679	554709	131774	98082	13.5796364	35.9733463	3.325946492	9.184567843	0.7776794	1	0.244921616	1	0.428323144	0.352941176	0.255315915	0.310344828
0.4	0.8	52967	15878	48865	10875	544037	142428	98691	11.08429777	34.26357224	2.91472424	9.075034483	0.813241174	1	0.262959756	1	0.437977539	0.36	0.264859555	0.285714286
0.4	0.9	72597	17090	67258	11509	522520	165116	101678	8.087138587	30.57460503	2.454964465	8.834651143	0.855006375	1	0.30356404	1	0.468082059	0.391304348	0.288953893	0.32
0.5	0.1	40223	16549	36459	11664	559053	138193	108763	14.59617632	33.78167865	3.790367262	9.324674211	0.745603213	1	0.259682206	1	0.44224413	0.375	0.276027556	0.333333333
0.5	0.2	40223	16549	36459	11664	559053	138193	108763	14.59617632	33.78167865	3.790367262	9.324674211	0.745603213	1	0.259682206	1	0.44224413	0.375	0.276027556	0.333333333
0.5	0.3	40223	16549	36459	11664	559053	138193	108763	14.59617632	33.78167865	3.790367262	9.324674211	0.745603213	1	0.259682206	1	0.44224413	0.375	0.276027556	0.333333333
0.5	0.4	40223	16549	36459	11664	559053	138193	108763	14.59617632	33.78167865	3.790367262	9.324674211	0.745603213	1	0.259682206	1	0.44224413	0.375	0.276027556	0.333333333
0.5	0.5	40223	16549	36459	11664	559053	138193	108763	14.59617632	33.78167865	3.790367262	9.324674211	0.745603213	1	0.259682206	1	0.44224413	0.375	0.276027556	0.333333333
0.5	0.6	42750	16679	38975	11818	556268	142210	110129	13.73338012	33.35139996	3.648749198	9.318751058	0.760032514	1	0.265684716	1	0.44589836	0.379310345	0.279411091	0.333333333
0.5	0.7	48240	16940	44246	12028	550243	149416	111842	12.17043947	32.48187721	3.376938028	9.298470236	0.785574374	1	0.277470509	1	0.452059309	0.384615385	0.286266406	0.346153846
0.5	0.8	57408	17309	52949	12141	540218	158600	111569	10.2268325	31.21023745	2.995335134	9.189440738	0.816750679	1	0.29288982	1	0.461417809	0.4	0.294436745	0.36
0.5	0.9	76361	18234	70688	12545	519122	178376	112143	7.688505913	28.4700011	2.52342689	8.939258669	0.85638648	1	0.328207706	1	0.486322527	0.423076923	0.313988701	0.363636364
0.6	0.1	54031	18839	49024	13293	544544	164766	122717	10.86602136	28.90514359	3.360925261	9.231700895	0.789353215	1	0.309305969	1	0.471978191	0.416666667	0.319379174	0.391304348
0.6	0.2	54031	18839	49024	13293	544544	164766	122717	10.86602136	28.90514359	3.360925261	9.231700895	0.789353215	1	0.309305969	1	0.471978191	0.416666667	0.319379174	0.391304348
0.6	0.3	54031	18839	49024	13293	544544	164766	122717	10.86602136	28.90514359	3.360925261	9.231700895	0.789353215	1	0.309305969	1	0.471978191	0.416666667	0.319379174	0.391304348
0.6	0.4	54031	18839	49024	13293	544544	164766	122717	10.86602136	28.90514359	3.360925261	9.231700895	0.789353215	1	0.309305969	1	0.471978191	0.416666667	0.319379174	0.391304348
0.6	0.5	54031	18839	49024	13293	544544	164766	122717	10.86602136	28.90514359	3.360925261	9.231700895	0.789353215	1	0.309305969	1	0.471978191	0.416666667	0.319379174	0.391304348
0.6	0.6	54031	18839	49024	13293	544544	164766	122717	10.86602136	28.90514359	3.360925261	9.231700895	0.789353215	1	0.309305969	1	0.471978191	0.416666667	0.319379174	0.391304348
0.6	0.7	57652	18988	52559	13412	540609	169409	123589	10.18354957	28.471087	3.223215815	9.214807635	0.802472099	1	0.316512017	1	0.476043503	0.416666667	0.323654929	0.409090909
0.6	0.8	65701	19262	60192	13453	531774	177016	122834	8.935967489	27.60741356	2.940855928	9.13060284	0.825648162	1	0.329103248	1	0.485081619	0.428571429	0.330730107	0.409090909
0.6	0.9	83873	20022	77175	13707	511556	194735	122124	6.999892695	25.54969534	2.523291221	8.909608229	0.860920782	1	0.3604757	1	0.509650003	0.458333333	0.348716809	0.4
0.7	0.1	76793	21858	68552	14081	519592	187968	124215	7.645254125	23.7712508	2.741976893	8.821461544	0.833729329	1	0.35865085	1	0.502768704	0.458333333	0.371097912	0.444444444
0.7	0.2	76793	21858	68552	14081	519592	187968	124215	7.645254125	23.7712508	2.741976893	8.821461544	0.833729329	1	0.35865085	1	0.502768704	0.458333333	0.371097912	0.444444444
0.7	0.3	76793	21858	68552	14081	519592	187968	124215	7.645254125	23.7712508	2.741976893	8.821461544	0.833729329	1	0.35865085	1	0.502768704	0.458333333	0.371097912	0.444444444
0.7	0.4	76793	21858	68552	14081	519592	187968	124215	7.645254125	23.7712508	2.741976893	8.821461544	0.833729329	1	0.35865085	1	0.502768704	0.458333333	0.371097912	0.444444444
0.7	0.5	76793	21858	68552	14081	519592	187968	124215	7.645254125	23.7712508	2.741976893	8.821461544	0.833729329	1	0.35865085	1	0.502768704	0.458333333	0.371097912	0.444444444
0.7	0.6	76793	21858	68552	14081	519592	187968	124215	7.645254125	23.7712508	2.741976893	8.821461544	0.833729329	1	0.35865085	1	0.502768704	0.458333333	0.371097912	0.444444444
0.7	0.7	76793	21858	68552	14081	519592	187968	124215	7.645254125	23.7712508	2.741976893	8.821461544	0.833729329	1	0.35865085	1	0.502768704	0.458333333	0.371097912	0.444444444
0.7	0.8	82251	22019	73794	14150	513582	193687	124217	7.137931454	23.32449248	2.624698485	8.778586572	0.845346198	1	0.36771136	1	0.51084577	0.473684211	0.376367742	0.444444444
0.7	**0.9**	**98991**	**22616**	**89554**	**14359**	**494880**	**209997**	**123424**	**5.930862402**	**21.88185356**	**2.344920383**	**8.595584651**	**0.871528694**	**1**	**0.395375955**	**1**	**0.536727049**	**0.5**	**0.392817941**	**0.470588235**
0.8	0.1	115787	26105	100251	13267	474621	207848	106436	5.070534689	18.18122965	2.073276077	8.022612497	0.875145709	1	0.408887071	1	0.547877963	0.523809524	0.441257971	0.583333333
0.8	0.2	115787	26105	100251	13267	474621	207848	106436	5.070534689	18.18122965	2.073276077	8.022612497	0.875145709	1	0.408887071	1	0.547877963	0.523809524	0.441257971	0.583333333
0.8	0.3	115787	26105	100251	13267	474621	207848	106436	5.070534689	18.18122965	2.073276077	8.022612497	0.875145709	1	0.408887071	1	0.547877963	0.523809524	0.441257971	0.583333333
0.8	0.4	115787	26105	100251	13267	474621	207848	106436	5.070534689	18.18122965	2.073276077	8.022612497	0.875145709	1	0.408887071	1	0.547877963	0.523809524	0.441257971	0.583333333
0.8	0.5	115787	26105	100251	13267	474621	207848	106436	5.070534689	18.18122965	2.073276077	8.022612497	0.875145709	1	0.408887071	1	0.547877963	0.523809524	0.441257971	0.583333333
0.8	0.6	115787	26105	100251	13267	474621	207848	106436	5.070534689	18.18122965	2.073276077	8.022612497	0.875145709	1	0.408887071	1	0.547877963	0.523809524	0.441257971	0.583333333
0.8	0.7	115787	26105	100251	13267	474621	207848	106436	5.070534689	18.18122965	2.073276077	8.022612497	0.875145709	1	0.408887071	1	0.547877963	0.523809524	0.441257971	0.583333333
0.8	0.8	115787	26105	100251	13267	474621	207848	106436	5.070534689	18.18122965	2.073276077	8.022612497	0.875145709	1	0.408887071	1	0.547877963	0.523809524	0.441257971	0.583333333
0.8	0.9	129219	26336	113227	13511	459207	223302	107638	4.54346497	17.43647479	1.972162117	7.966693805	0.890342264	1	0.434065659	1	0.573896513	0.555555556	0.456898192	0.636363636
0.9	0.1	190675	30440	158868	9593	381748	239365	68322	3.079071719	12.54099869	1.506691089	7.122068175	0.921971512	1	0.489332899	1	0.630699114	0.666666667	0.567902793	0.857142857
0.9	0.2	190675	30440	158868	9593	381748	239365	68322	3.079071719	12.54099869	1.506691089	7.122068175	0.921971512	1	0.489332899	1	0.630699114	0.666666667	0.567902793	0.857142857
0.9	0.3	190675	30440	158868	9593	381748	239365	68322	3.079071719	12.54099869	1.506691089	7.122068175	0.921971512	1	0.489332899	1	0.630699114	0.666666667	0.567902793	0.857142857
0.9	0.4	190675	30440	158868	9593	381748	239365	68322	3.079071719	12.54099869	1.506691089	7.122068175	0.921971512	1	0.489332899	1	0.630699114	0.666666667	0.567902793	0.857142857
0.9	0.5	190675	30440	158868	9593	381748	239365	68322	3.079071719	12.54099869	1.506691089	7.122068175	0.921971512	1	0.489332899	1	0.630699114	0.666666667	0.567902793	0.857142857
0.9	0.6	190675	30440	158868	9593	381748	239365	68322	3.079071719	12.54099869	1.506691089	7.122068175	0.921971512	1	0.489332899	1	0.630699114	0.666666667	0.567902793	0.857142857
0.9	0.7	190675	30440	158868	9593	381748	239365	68322	3.079071719	12.54099869	1.506691089	7.122068175	0.921971512	1	0.489332899	1	0.630699114	0.666666667	0.567902793	0.857142857
0.9	0.8	190675	30440	158868	9593	381748	239365	68322	3.079071719	12.54099869	1.506691089	7.122068175	0.921971512	1	0.489332899	1	0.630699114	0.666666667	0.567902793	0.857142857
0.9	0.9	190675	30440	158868	9593	381748	239365	68322	3.079071719	12.54099869	1.506691089	7.122068175	0.921971512	1	0.489332899	1	0.630699114	0.666666667	0.567902793	0.857142857

10.2 Supplementary table 2: List of species taxonomy

Species	Family	Superfamily	Infraorder
Abronia graminea	Anguidae	Anguioidea	Anguimorpha
Achalinus meiguensis	Xenodermatidae	Colubroidea	Serpentes
Achalinus rufescens	Xenodermatidae	Colubroidea	Serpentes
Achalinus spinalis	Xenodermatidae	Colubroidea	Serpentes
Acrochordus granulatus	Acrochordidae	Acrochordoidea	Serpentes
Aeluroscalabotes felinus	Gekkonidae		Gekkota
Agkistrodon contortrix	Viperidae	Colubroidea	Serpentes
Agkistrodon piscivorus	Viperidae	Colubroidea	Serpentes
Amblyrhynchus cristatus	Iguanidae	Iguanidae	Iguania
Amerotyphlops reticulatus	Typhlopidae	Solecophidia	Serpentes
Amphisbaena schmidti	Amphisbaenidae	Amphisbaenia	Laterata
Anguis cephallonica	Anguidae	Anguioidea	Anguimorpha
Anguis colchica	Anguidae	Anguioidea	Anguimorpha
Anguis fragilis	Anguidae	Anguioidea	Anguimorpha
Anguis graeca	Anguidae	Anguioidea	Anguimorpha
Anguis veronensis	Anguidae	Anguioidea	Anguimorpha
Anilius scytale	Aniliidae	Henophidia	Serpentes
Anolis carolinensis	Dactyloidae	Iguania	Iguania
Anolis punctatus	Dactyloidae	Iguania	Iguania
Aprasia parapulchella	Pygopodidae	Pygopoidea	Gekkota
Asymblepharus himalayanus	Scincidae	Scinciformata	Scinciformata
Ateuchosaurus chinensis	Scincidae	Scinciformata	Scinciformata
Azemiops feae	Viperidae	Colubroidea	Serpentes
Basiliscus vittatus	Iguanidae	Iguanidae	Iguania
Bavayia robusta	Diplodactylidae	Pygopoidea	Gekkota
Bipes biporus	Bipdedidae	Amphisbaenia	Laterata
Bipes canaliculatus	Bipdedidae	Amphisbaenia	Laterata
Bipes tridactylus	Bipdedidae	Amphisbaenia	Laterata
Blanus cinereus	Blanidae	Amphisbaenia	Laterata
Boa constrictor	Boidae	Henophidia	Serpentes
Bothrops diporus	Viperidae	Colubroidea	Serpentes
Bothrops jararaca	Viperidae	Colubroidea	Serpentes
Bothrops pubescens	Viperidae	Colubroidea	Serpentes
Brookesia decaryi	Chamaeleonidae	Acrodonta	Iguania
Bungarus fasciatus	Elapidae	Colubroidea	Serpentes
Calotes mystaceus	Agamidae	Acrodonta	Iguania
Calotes versicolor	Agamidae	Acrodonta	Iguania
Causus defilippii	Viperidae	Colubroidea	Serpentes
Chalarodon madagascariensis	Iguanidae	Iguanidae	Iguania
Chamaeleo africanus	Chamaeleonidae	Acrodonta	Iguania
Chamaeleo arabicus	Chamaeleonidae	Acrodonta	Iguania
Chamaeleo calcaricarens	Chamaeleonidae	Acrodonta	Iguania
Chamaeleo calyptratus	Chamaeleonidae	Acrodonta	Iguania
Chamaeleo chamaeleon	Chamaeleonidae	Acrodonta	Iguania
Chamaeleo dilepis	Chamaeleonidae	Acrodonta	Iguania
Chamaeleo monachus	Chamaeleonidae	Acrodonta	Iguania
Chamaeleo zeylanicus	Chamaeleonidae	Acrodonta	Iguania
Chlamydosaurus kingii	Agamidae	Acrodonta	Iguania
Cnemaspis limi	Gekkonidae		Gekkota
Coleonyx variegatus	Eublepharidae		Gekkota
Conolophus subcristatus	Iguanidae	Iguanidae	Iguania
Crotalus adamanteus	Viperidae	Colubroidea	Serpentes
Crotalus tigris	Viperidae	Colubroidea	Serpentes
Cyclura pinguis	Iguanidae	Iguanidae	Iguania
Cylindrophis ruffus	Cylindrophiidae	Henophidia	Serpentes
Cyrtopodion scabrum	Gekkonidae		Gekkota
Daboia russelii	Viperidae	Colubroidea	Serpentes
Darevskia armeniaca	Lacertidae	Lacertabaenia	Laterata
Darevskia brauneri	Lacertidae	Lacertabaenia	Laterata
Darevskia caucasica	Lacertidae	Lacertabaenia	Laterata
Darevskia chlorogaster	Lacertidae	Lacertabaenia	Laterata
Darevskia clarkorum	Lacertidae	Lacertabaenia	Laterata
Darevskia daghestanica	Lacertidae	Lacertabaenia	Laterata
Darevskia dahli	Lacertidae	Lacertabaenia	Laterata
Darevskia derjugini	Lacertidae	Lacertabaenia	Laterata
Darevskia mixta	Lacertidae	Lacertabaenia	Laterata
Darevskia parvula	Lacertidae	Lacertabaenia	Laterata
Darevskia portschinskii	Lacertidae	Lacertabaenia	Laterata
Darevskia praticola	Lacertidae	Lacertabaenia	Laterata
Darevskia raddei	Lacertidae	Lacertabaenia	Laterata
Darevskia rudis	Lacertidae	Lacertabaenia	Laterata
Darevskia saxicola	Lacertidae	Lacertabaenia	Laterata
Darevskia valentini	Lacertidae	Lacertabaenia	Laterata
Deinagkistrodon acutus	Viperidae	Colubroidea	Serpentes
Diploderma flaviceps	Agamidae	Acrodonta	Iguania
Diploderma micangshanensis	Agamidae	Acrodonta	Iguania

Diplometopon zarudnyi	Trogonophidae	Amphisbaenia	Laterata
Dopasia gracilis	Anguidae	Anguioidea	Anguimorpha
Dopasia hainanensis	Anguidae	Anguioidea	Anguimorpha
Dopasia harti	Anguidae	Anguioidea	Anguimorpha
Elaphe anomala	Colubridae	Colubroidea	Serpentes
Elaphe bimaculata	Colubridae	Colubroidea	Serpentes
Elaphe davidi	Colubridae	Colubroidea	Serpentes
Elaphe dione	Colubridae	Colubroidea	Serpentes
Elaphe schrenckii	Colubridae	Colubroidea	Serpentes
Elaphe taeniura	Colubridae	Colubroidea	Serpentes
Eremias brenchleyi	Lacertidae	Lacertabaenia	Laterata
Eremias dzungarica	Lacertidae	Lacertabaenia	Laterata
Eremias multiocellata	Lacertidae	Lacertabaenia	Laterata
Eremias przewalskii	Lacertidae	Lacertabaenia	Laterata
Eremias stummeri	Lacertidae	Lacertabaenia	Laterata
Eremias vermiculata	Lacertidae	Lacertabaenia	Laterata
Eublepharis macularius	Eublepharidae		Gekkota
Euprepiophis perlacea	Colubridae	Colubroidea	Serpentes
Furcifer oustaleti	Chamaeleonidae	Acrodonta	Iguania
Gambelia wislizenii	Iguanidae	Iguanidae	Iguania
Gekko chinensis	Gekkonidae		Gekkota
Gekko gecko	Gekkonidae		Gekkota
Gekko hokouensis	Gekkonidae		Gekkota
Gekko japonicus	Gekkonidae		Gekkota
Gekko subpalmatus	Gekkonidae		Gekkota
Gekko swinhonis	Gekkonidae		Gekkota
Gekko vittatus	Gekkonidae		Gekkota
Geocalamus acutus	Amphisbaenidae	Amphisbaenia	Laterata
Gloydius brevicaudus	Viperidae	Colubroidea	Serpentes
Gloydius intermedius	Viperidae	Colubroidea	Serpentes
Gloydius saxatilis	Viperidae	Colubroidea	Serpentes
Gloydius shedaoensis	Viperidae	Colubroidea	Serpentes
Gloydius strauchi	Viperidae	Colubroidea	Serpentes
Gloydius ussuriensis	Viperidae	Colubroidea	Serpentes
Gonatodes albogularis	Sphaerodactylidae		Gekkota
Goniurosaurus luii	Eublepharidae		Gekkota
Gonyosoma frenatum	Colubridae	Colubroidea	Serpentes
Heloderma suspectum	Helodermatidae		Anguimorpha
Hemidactylus bowringii	Gekkonidae		Gekkota
Hemidactylus frenatus	Gekkonidae		Gekkota
Hemitheconyx caudicinctus	Eublepharidae		Gekkota
Heteronotia binoei	Gekkonidae		Gekkota
Holbrookia lacerata	Phrynosomatidae	Iguanidae	Iguania
Hydrophis curtus	Hydrophiidae	Colubroidea	Serpentes
Hydrophis cyanocinctus	Hydrophiidae	Colubroidea	Serpentes
Hydrosaurus amboinensis	Agamidae	Acrodonta	Iguania
Hypsiglena chlorophaea chlorophaea	Dipsadidae	Colubroidea	Serpentes
Hypsiglena chlorophaea deserticola	Dipsadidae	Colubroidea	Serpentes
Hypsiglena jani texana	Dipsadidae	Colubroidea	Serpentes
Hypsiglena ochrorhyncha klauberi	Dipsadidae	Colubroidea	Serpentes
Hypsiglena ochrorhyncha nuchalata	Dipsadidae	Colubroidea	Serpentes
Hypsiglena ochrorhyncha ochrorhyncha	Dipsadidae	Colubroidea	Serpentes
Hypsiglena slevini	Dipsadidae	Colubroidea	Serpentes
Hypsiglena sp. DGM-2008	Dipsadidae	Colubroidea	Serpentes
Hypsiglena torquata	Dipsadidae	Colubroidea	Serpentes
Hypsiglena unaocularus	Dipsadidae	Colubroidea	Serpentes
Hypsiscopus plumbea	Homalopsidae	Colubroidea	Serpentes
Iguana delicatissima	Iguanidae	Iguanidae	Iguania
Iguana iguana	Iguanidae	Iguanidae	Iguania
Imantodes cenchoa	Dipsadidae	Colubroidea	Serpentes
Indotyphlops braminus	Typhlopidae	Solecophidia	Serpentes
Iphisa elegans	Gymnophthalmidae	Teiioidea	Laterata
Isopachys gyldenstolpei	Scincidae	Scinciformata	Scinciformata
Kinyongia fischeri	Chamaeleonidae	Acrodonta	Iguania
Lacerta agilis	Lacertidae	Lacertabaenia	Laterata
Lacerta bilineata	Lacertidae	Lacertabaenia	Laterata
Lacerta viridis viridis	Lacertidae	Lacertabaenia	Laterata
Laticauda colubrina	Elapidae	Colubroidea	Serpentes
Laticauda laticaudata	Elapidae	Colubroidea	Serpentes
Laticauda semifasciata	Elapidae	Colubroidea	Serpentes
Leiocephalus personatus	Iguanidae	Iguanidae	Iguania
Leiolepis guttata	Agamidae	Acrodonta	Iguania
Leiolepis reevesii	Agamidae	Acrodonta	Iguania
Lepidodactylus lugubris	Gekkonidae		Gekkota
Lepidophyma flavimaculatum	Xantusiidae	Scinciformata	Scinciformata
Leptodeira polysticta	Dipsadidae	Colubroidea	Serpentes

Liolaemus darwinii	Iguanidae	Iguanidae	Iguania
Liolaemus millcayac	Iguanidae	Iguanidae	Iguania
Liolaemus parthenos	Iguanidae	Iguanidae	Iguania
Lycodon flavozonatus	Colubridae	Colubroidea	Serpentes
Lycodon rufozonatus	Colubridae	Colubroidea	Serpentes
Lycodon ruhstrati	Colubridae	Colubroidea	Serpentes
Lycodon semicarinatus	Colubridae	Colubroidea	Serpentes
Macrovipera schweizeri	Viperidae	Colubroidea	Serpentes
Malayopython reticulatus	Pythonidae	Henophidia	Serpentes
Micrurus fulvius	Elapidae	Colubroidea	Serpentes
Naja atra	Elapidae	Colubroidea	Serpentes
Naja naja	Elapidae	Colubroidea	Serpentes
Nerodia sipedon	Colubridae	Colubroidea	Serpentes
Notechis scutatus	Elapidae	Colubroidea	Serpentes
Oligodon chinensis	Colubridae	Colubroidea	Serpentes
Oocatochus rufodorsatus	Colubridae	Colubroidea	Serpentes
Ophiophagus hannah	Elapidae	Colubroidea	Serpentes
Opisthotropis guangxiensis	Colubridae	Colubroidea	Serpentes
Opisthotropis latouchii	Colubridae	Colubroidea	Serpentes
Oplurus grandidieri	Iguanidae	Iguanidae	Iguania
Oreocryptophis porphyraceus	Colubridae	Colubroidea	Serpentes
Orientocoluber spinalis	Colubridae	Colubroidea	Serpentes
Ovophis okinavensis	Viperidae	Colubroidea	Serpentes
Pantherophis guttatus	Colubridae	Colubroidea	Serpentes
Pantherophis slowinskii	Colubridae	Colubroidea	Serpentes
Pareas boulengeri	Pareatidae	Colubroidea	Serpentes
Pareas formosensis	Pareatidae	Colubroidea	Serpentes
Pareas stanleyi	Pareatidae	Colubroidea	Serpentes
Paroedura picta	Gekkonidae		Gekkota
Phoenicolacerta kulzeri	Lacertidae	Lacertabaenia	Laterata
Phrynocephalus albolineatus	Agamidae	Acrodonta	Iguania
Phrynocephalus axillaris	Agamidae	Acrodonta	Iguania
Phrynocephalus forsythii	Agamidae	Acrodonta	Iguania
Phrynocephalus grumgrzimailoi	Agamidae	Acrodonta	Iguania
Phrynocephalus guinanensis	Agamidae	Acrodonta	Iguania
Phrynocephalus helioscopus	Agamidae	Acrodonta	Iguania
Phrynocephalus maculatus	Agamidae	Acrodonta	Iguania
Phrynocephalus nasatus	Agamidae	Acrodonta	Iguania
Phrynocephalus przewalskii	Agamidae	Acrodonta	Iguania
Phrynocephalus putjatai	Agamidae	Acrodonta	Iguania
Phrynocephalus versicolor	Agamidae	Acrodonta	Iguania
Phrynosoma blainvillii	Phrynosomatidae	Iguanidae	Iguania
Phyllodactylus unctus	Phyllodactylidae		Gekkota
Plestiodon chinensis	Scincidae	Scinciformata	Scinciformata
Plestiodon egregius	Scincidae	Scinciformata	Scinciformata
Plestiodon elegans	Scincidae	Scinciformata	Scinciformata
Plestiodon tunganus	Scincidae	Scinciformata	Scinciformata
Podarcis muralis	Lacertidae	Lacertabaenia	Laterata
Podarcis siculus	Lacertidae	Lacertabaenia	Laterata
Pogona vitticeps	Agamidae	Acrodonta	Iguania
Polychrus marmoratus	Polycrotidae		Iguania
Protobothrops cornutus	Viperidae	Colubroidea	Serpentes
Protobothrops dabieshanensis	Viperidae	Colubroidea	Serpentes
Protobothrops flavoviridis	Viperidae	Colubroidea	Serpentes
Protobothrops himalayanus	Viperidae	Colubroidea	Serpentes
Protobothrops jerdonii	Viperidae	Colubroidea	Serpentes
Protobothrops kaulbacki	Viperidae	Colubroidea	Serpentes
Protobothrops mangshanensis	Viperidae	Colubroidea	Serpentes
Protobothrops maolanensis	Viperidae	Colubroidea	Serpentes
Protobothrops mucrosquamatus	Viperidae	Colubroidea	Serpentes
Protobothrops tokarensis	Viperidae	Colubroidea	Serpentes
Pseudagkistrodon rudis	Colubridae	Colubroidea	Serpentes
Pseudocalotes microlepis	Agamidae	Acrodonta	Iguania
Pseudoleptodeira latifasciata	Dipsadidae	Colubroidea	Serpentes
Pseudonaja textilis	Elapidae	Colubroidea	Serpentes
Pseudotrapelus sinaitus	Agamidae	Acrodonta	Iguania
Pseudoxenodon stejnegeri	Colubridae	Colubroidea	Serpentes
Ptyas dhumnades	Colubridae	Colubroidea	Serpentes
Ptyas major	Colubridae	Colubroidea	Serpentes
Ptyas mucosa	Colubridae	Colubroidea	Serpentes
Ptyodactylus guttatus	Phyllodactylidae		Gekkota
Python bivittatus	Pythonidae	Henophidia	Serpentes
Python molurus molurus	Pythonidae	Henophidia	Serpentes
Python regius	Pythonidae	Henophidia	Serpentes
Rena humilis	Leptotyphlopidae	Solecophidia	Serpentes
Rhabdophis tigrinus	Colubridae	Colubroidea	Serpentes
Rhineura floridana	Rhineuridae	Amphisbaenia	Laterata
Sceloporus occidentalis	Phrynosomatidae	Iguanidae	Iguania
Sceloporus undulatus	Phrynosomatidae	Iguania	Iguania
Scincella huanrenensis	Scincidae	Scinciformata	Scinciformata

Scincella modesta	Scincidae	Scinciformata	Scinciformata
Scincella reevesii	Scincidae	Scinciformata	Scinciformata
Scincella vandenburghi	Scincidae	Scinciformata	Scinciformata
Shinisaurus crocodilurus	Shinisauridae		Anguimorpha
Sibon nebulatus	Dipsadidae	Colubroidea	Serpentes
Sibynophis chinensis	Colubridae	Colubroidea	Serpentes
Sibynophis collaris	Colubridae	Colubroidea	Serpentes
Sinomicrurus macclellandi	Elapidae	Colubroidea	Serpentes
Smaug warreni	Cordylidae	Cordyloidea	Scinciformata
Sphenomorphus incognitus	Scincidae	Scinciformata	Scinciformata
Sphenomorphus indicus	Scincidae	Scinciformata	Scinciformata
Stichophanes ningshaanensis	Colubridae	Colubroidea	Scinciformata
Takydromus amurensis	Lacertidae	Lacertabaenia	Laterata
Takydromus sexlineatus	Lacertidae	Lacertabaenia	Laterata
Takydromus wolteri	Lacertidae	Lacertabaenia	Laterata
Tarentola mauritanica	Phyllodactylidae		Gekkota
Teratoscincus roborowskii	Sphaerodactylidae		Gekkota
Thamnophis elegans	Colubridae	Colubroidea	Serpentes
Thamnophis sirtalis	Colubridae	Colubroidea	Serpentes
Thermophis baileyi	Dipsadidae	Colubroidea	Serpentes
Thermophis shangrila	Dipsadidae	Colubroidea	Serpentes
Thermophis zhaoermii	Dipsadidae	Colubroidea	Serpentes
Trimeresurus albolabris	Viperidae	Colubroidea	Serpentes
Trimeresurus sichuanensis	Viperidae	Colubroidea	Serpentes
Trimeresurus stejnegeri stejnegeri	Viperidae	Colubroidea	Serpentes
Trioceros melleri	Chamaeleonidae	Acrodonta	Iguania
Tropidophis haetianus	Tropidophiidae	Henophidia	Serpentes
Tropidophorus hangnam	Scincidae	Scinciformata	Scinciformata
Uromastyx benti	Agamidae	Acrodonta	Iguania
Uroplatus ebenaui	Gekkonidae		Gekkota
Uroplatus fimbriatus	Gekkonidae		Gekkota
Urosaurus nigricaudus	Phrynosomatidae	Iguanidae	Iguania
Uta stansburiana	Phrynosomatidae	Iguanidae	Iguania
Varanus komodoensis	Varanidae	Varanoidea	Anguimorpha
Varanus salvator	Varanidae	Varanoidea	Anguimorpha
Vipera berus	Viperidae	Colubroidea	Serpentes
Xenagama taylori	Agamidae	Acrodonta	Iguania
Xenopeltis unicolor	Xenopeltidae	Henophidia	Serpentes
Xerotyphlops vermicularis	Typhlopidae	Solecophidia	Serpentes
Zootoca vivipara	Lacertidae	Lacertabaenia	Laterata

10.3 Supplementary table 3: List of protein families

Protein Family	Protein Name red - mitochondrial tree
refseqs_squamata_FAM000001.fasta	cytochrome b (mitochondrion)
refseqs_squamata_FAM000004.fasta	NADH dehydrogenase subunit 4 (mitochondrion)
refseqs_squamata_FAM000005.fasta	NADH dehydrogenase subunit 4L (mitochondrion)
refseqs_squamata_FAM000006.fasta	NADH dehydrogenase subunit 3 (mitochondrion)
refseqs_squamata_FAM000007.fasta	cytochrome c oxidase subunit III (mitochondrion)
refseqs_squamata_FAM000008.fasta	ATP synthase F0 subunit 6 (mitochondrion)
refseqs_squamata_FAM000010.fasta	cytochrome c oxidase subunit II (mitochondrion)
refseqs_squamata_FAM000011.fasta	cytochrome c oxidase subunit I (mitochondrion)
refseqs_squamata_FAM000012.fasta	NADH dehydrogenase subunit 2 (mitochondrion)
refseqs_squamata_FAM000013.fasta	NADH dehydrogenase subunit 1 (mitochondrion)
refseqs_squamata_FAM000014.fasta	cytochrome b (mitochondrion)
refseqs_squamata_FAM000015.fasta	NADH dehydrogenase subunit 6 (mitochondrion)
refseqs_squamata_FAM000017.fasta	NADH dehydrogenase subunit 4 (mitochondrion)
refseqs_squamata_FAM000018.fasta	NADH dehydrogenase subunit 4L (mitochondrion)
refseqs_squamata_FAM000019.fasta	NADH dehydrogenase subunit 3 (mitochondrion)
refseqs_squamata_FAM000020.fasta	cytochrome c oxidase subunit III (mitochondrion)
refseqs_squamata_FAM000021.fasta	ATP synthase F0 subunit 6 (mitochondrion)
refseqs_squamata_FAM000022.fasta	cytochrome c oxidase subunit II (mitochondrion)
refseqs_squamata_FAM000023.fasta	NADH dehydrogenase subunit 2 (mitochondrion)
refseqs_squamata_FAM000062.fasta	DNA replication licensing factor MCM3
refseqs_squamata_FAM000063.fasta	transcription factor AP-2-delta
refseqs_squamata_FAM000084.fasta	dynactin subunit 6
refseqs_squamata_FAM000087.fasta	dual specificity protein phosphatase 4
refseqs_squamata_FAM000088.fasta	leucine-rich repeat-containing protein 14B
refseqs_squamata_FAM000094.fasta	signal recognition particle subunit SRP72
refseqs_squamata_FAM000098.fasta	actin-like protein 7A
refseqs_squamata_FAM000101.fasta	DNA-directed RNA polymerase II subunit RPB2
refseqs_squamata_FAM000112.fasta	ADP-ribosylation factor-like protein 9
refseqs_squamata_FAM000145.fasta	G-protein coupled receptor 42-like
refseqs_squamata_FAM000153.fasta	rho-related GTP-binding protein RhoN
refseqs_squamata_FAM000154.fasta	uncharacterized protein C17orf98 homolog
refseqs_squamata_FAM000156.fasta	vacuolar protein-sorting-associated protein 25
refseqs_squamata_FAM000157.fasta	60S ribosomal protein L23
refseqs_squamata_FAM000191.fasta	Ig-like V-type domain-containing protein FAM187A

refseqs_squamata_FAM000202.fasta	ribonucleoside-diphosphate reductase large subunit
refseqs_squamata_FAM000265.fasta	lipoyl synthase, mitochondrial
refseqs_squamata_FAM000352.fasta	proton channel OTOP1
refseqs_squamata_FAM000388.fasta	kelch-like protein 41
refseqs_squamata_FAM000412.fasta	Golgi reassembly-stacking protein 2
refseqs_squamata_FAM000428.fasta	zinc transporter 9
refseqs_squamata_FAM000601.fasta	proliferating cell nuclear antigen
refseqs_squamata_FAM000628.fasta	U5 small nuclear ribonucleoprotein 200 kDa helicase
refseqs_squamata_FAM000744.fasta	V-type proton ATPase subunit E 1
refseqs_squamata_FAM000755.fasta	BTB/POZ domain-containing protein 17
refseqs_squamata_FAM000770.fasta	E3 SUMO-protein ligase CBX4
refseqs_squamata_FAM000774.fasta	chromobox protein homolog 2
refseqs_squamata_FAM000816.fasta	5-hydroxytryptamine receptor 3A
refseqs_squamata_FAM000890.fasta	malate dehydrogenase, mitochondrial
refseqs_squamata_FAM000898.fasta	claudin-3-like
refseqs_squamata_FAM000962.fasta	ATP synthase subunit delta, mitochondrial
refseqs_squamata_FAM000989.fasta	glutamate receptor ionotropic, NMDA 3B
refseqs_squamata_FAM001034.fasta	PHD finger-like domain-containing protein 5A
refseqs_squamata_FAM001036.fasta	protein Tob2
refseqs_squamata_FAM001108.fasta	histidine ammonia-lyase
refseqs_squamata_FAM001117.fasta	protein phosphatase 1 regulatory subunit 3D
refseqs_squamata_FAM001120.fasta	leukotriene A-4 hydrolase
refseqs_squamata_FAM001151.fasta	protein ABHD16B-like
refseqs_squamata_FAM001152.fasta	protein ABHD16B-like
refseqs_squamata_FAM001155.fasta	histamine H3 receptor
refseqs_squamata_FAM001170.fasta	adenylyltransferase and sulfurtransferase MOCS3
refseqs_squamata_FAM001199.fasta	proteasome subunit alpha type-7
refseqs_squamata_FAM001206.fasta	probable ATP-dependent RNA helicase DDX47
refseqs_squamata_FAM001227.fasta	sodium-coupled monocarboxylate transporter 1-like
refseqs_squamata_FAM001266.fasta	transcription factor SOX-12
refseqs_squamata_FAM001275.fasta	chorion-specific transcription factor GCMb
refseqs_squamata_FAM001278.fasta	cyclic AMP-dependent transcription factor ATF-4
refseqs_squamata_FAM001299.fasta	partitioning defective 6 homolog beta
refseqs_squamata_FAM001329.fasta	thioredoxin domain-containing protein 5
refseqs_squamata_FAM001343.fasta	E3 ubiquitin-protein ligase RBX1
refseqs_squamata_FAM001347.fasta	hsc70-interacting protein
refseqs_squamata_FAM001360.fasta	protein BTG1
refseqs_squamata_FAM001363.fasta	protein APCDD1-like
refseqs_squamata_FAM001387.fasta	fibronectin type III domain-containing protein 11
refseqs_squamata_FAM001427.fasta	transmembrane protein 17
refseqs_squamata_FAM001431.fasta	homeobox protein MSX-2
refseqs_squamata_FAM001441.fasta	neurogenin-1
refseqs_squamata_FAM001450.fasta	motilin receptor
refseqs_squamata_FAM001464.fasta	G-protein coupled receptor 151
refseqs_squamata_FAM001468.fasta	EF-hand calcium-binding domain-containing protein 9
refseqs_squamata_FAM001469.fasta	sulfate transporter-like
refseqs_squamata_FAM001483.fasta	60S ribosomal protein L26-like 1
refseqs_squamata_FAM001515.fasta	pre-mRNA-splicing factor RBM22
refseqs_squamata_FAM001525.fasta	dual specificity protein phosphatase 1
refseqs_squamata_FAM001537.fasta	stanniocalcin-2
refseqs_squamata_FAM001616.fasta	T-complex protein 1 subunit delta
refseqs_squamata_FAM001641.fasta	coiled-coil domain-containing protein 42
refseqs_squamata_FAM001664.fasta	histone chaperone ASF1B
refseqs_squamata_FAM001673.fasta	protein CYR61-like
refseqs_squamata_FAM001684.fasta	peptidyl-tRNA hydrolase ICT1, mitochondrial
refseqs_squamata_FAM001792.fasta	pre-mRNA-splicing factor SYF1
refseqs_squamata_FAM001826.fasta	28S ribosomal protein S7, mitochondrial
refseqs_squamata_FAM001885.fasta	pleckstrin homology domain-containing family J member 1
refseqs_squamata_FAM001897.fasta	activity-regulated cytoskeleton-associated protein
refseqs_squamata_FAM001956.fasta	divergent protein kinase domain 1B
refseqs_squamata_FAM001987.fasta	dopamine beta-hydroxylase
refseqs_squamata_FAM002103.fasta	claudin-2
refseqs_squamata_FAM002123.fasta	transcription factor SOX-8
refseqs_squamata_FAM002164.fasta	probable glutamate--tRNA ligase, mitochondrial
refseqs_squamata_FAM002276.fasta	vesicle transport protein GOT1B
refseqs_squamata_FAM002301.fasta	mitochondrial carnitine/acylcarnitine carrier protein
refseqs_squamata_FAM002302.fasta	RING1 and YY1-binding protein
refseqs_squamata_FAM002324.fasta	26S proteasome non-ATPase regulatory subunit 6
refseqs_squamata_FAM002352.fasta	THO complex subunit 7 homolog
refseqs_squamata_FAM002403.fasta	pyruvate dehydrogenase E1 component subunit beta, mitochondrial
refseqs_squamata_FAM002426.fasta	probable G-protein coupled receptor 27
refseqs_squamata_FAM002429.fasta	cAMP-dependent protein kinase type II-alpha regulatory subunit
refseqs_squamata_FAM002446.fasta	troponin C, slow skeletal and cardiac muscles
refseqs_squamata_FAM002457.fasta	mesencephalic astrocyte-derived neurotrophic factor
refseqs_squamata_FAM002477.fasta	DNA replication licensing factor MCM2
refseqs_squamata_FAM002478.fasta	homeobox protein BarH-like 1
refseqs_squamata_FAM002499.fasta	actin-3-like
refseqs_squamata_FAM002503.fasta	glutathione peroxidase 1
refseqs_squamata_FAM002509.fasta	protein SEC13 homolog
refseqs_squamata_FAM002527.fasta	pre-mRNA-splicing factor ISY1 homolog

refseqs_squamata_FAM002529.fasta	dnaJ homolog subfamily B member 8
refseqs_squamata_FAM002536.fasta	netrin-4-like
refseqs_squamata_FAM002537.fasta	caM kinase-like vesicle-associated protein
refseqs_squamata_FAM002547.fasta	transmembrane protein 115
refseqs_squamata_FAM002573.fasta	actin-related protein 8
refseqs_squamata_FAM002575.fasta	ruvB-like 1
refseqs_squamata_FAM002592.fasta	ras-related protein Rab-43
refseqs_squamata_FAM002599.fasta	2',5'-phosphodiesterase 12
refseqs_squamata_FAM002617.fasta	ras association domain-containing protein 1 isoform X2
refseqs_squamata_FAM002644.fasta	urocanate hydratase
refseqs_squamata_FAM002646.fasta	T-cell leukemia translocation-altered gene protein
refseqs_squamata_FAM002677.fasta	alpha-2Da adrenergic receptor-like
refseqs_squamata_FAM002689.fasta	PRELI domain-containing protein 1, mitochondrial
refseqs_squamata_FAM002696.fasta	magnesium transporter NIPA4
refseqs_squamata_FAM002720.fasta	U7 snRNA-associated Sm-like protein LSm11
refseqs_squamata_FAM002752.fasta	malonyl-CoA-acyl carrier protein transacylase, mitochondrial
refseqs_squamata_FAM002774.fasta	homeobox protein Hox-A4
refseqs_squamata_FAM002777.fasta	homeobox protein Hox-A5
refseqs_squamata_FAM002778.fasta	homeobox protein Hox-A1
refseqs_squamata_FAM002782.fasta	homeobox protein Hox-A13
refseqs_squamata_FAM002791.fasta	mitochondrial import receptor subunit TOM7 homolog
refseqs_squamata_FAM002798.fasta	homeobox protein Hox-A10
refseqs_squamata_FAM002801.fasta	TLR4 interactor with leucine rich repeats
refseqs_squamata_FAM002825.fasta	complexin-3
refseqs_squamata_FAM002849.fasta	LOW QUALITY PROTEIN: ornithine decarboxylase antizyme 2
refseqs_squamata_FAM002895.fasta	WNT1-inducible-signaling pathway protein 1
refseqs_squamata_FAM002915.fasta	SLIT and NTRK-like protein 1
refseqs_squamata_FAM002918.fasta	solute carrier family 35 member F6
refseqs_squamata_FAM002937.fasta	coiled-coil domain-containing protein 25
refseqs_squamata_FAM002992.fasta	olfactory marker protein
refseqs_squamata_FAM003003.fasta	cyclic AMP-dependent transcription factor ATF-5
refseqs_squamata_FAM003024.fasta	olfactory receptor 5F1-like
refseqs_squamata_FAM003037.fasta	olfactory receptor 11L1
refseqs_squamata_FAM003084.fasta	calcium-activated potassium channel subunit beta-4
refseqs_squamata_FAM003095.fasta	claudin-22-like
refseqs_squamata_FAM003103.fasta	T-complex protein 1 subunit beta
refseqs_squamata_FAM003105.fasta	ras-related protein Rab-21
refseqs_squamata_FAM003106.fasta	leucine-rich repeat-containing protein 10
refseqs_squamata_FAM003118.fasta	YEATS domain-containing protein 4
refseqs_squamata_FAM003149.fasta	homogentisate 1,2-dioxygenase
refseqs_squamata_FAM003153.fasta	triosephosphate isomerase
refseqs_squamata_FAM003177.fasta	cystatin-B-like
refseqs_squamata_FAM003181.fasta	probable 28S rRNA (cytosine(4447)-C(5))-methyltransferase
refseqs_squamata_FAM003183.fasta	ribosomal RNA small subunit methyltransferase NEP1
refseqs_squamata_FAM003201.fasta	voltage-dependent calcium channel gamma-1 subunit
refseqs_squamata_FAM003211.fasta	matrix metalloproteinase-19
refseqs_squamata_FAM003212.fasta	voltage-dependent calcium channel gamma-4 subunit
refseqs_squamata_FAM003218.fasta	nucleolar protein 11
refseqs_squamata_FAM003225.fasta	receptor tyrosine-protein kinase erbB-3
refseqs_squamata_FAM003226.fasta	growth/differentiation factor 11
refseqs_squamata_FAM003239.fasta	guanine nucleotide-binding protein subunit alpha-13
refseqs_squamata_FAM003279.fasta	splicing factor 3B subunit 4
refseqs_squamata_FAM003303.fasta	peroxisomal membrane protein 11B
refseqs_squamata_FAM003341.fasta	T-complex protein 1 subunit gamma
refseqs_squamata_FAM003370.fasta	proteasome subunit beta type-4
refseqs_squamata_FAM003384.fasta	heat shock 70 kDa protein II-like
refseqs_squamata_FAM003410.fasta	ATP-dependent RNA helicase DDX54
refseqs_squamata_FAM003417.fasta	60S acidic ribosomal protein P0
refseqs_squamata_FAM003510.fasta	stress-induced-phosphoprotein 1
refseqs_squamata_FAM003682.fasta	retinol-binding protein 3
refseqs_squamata_FAM003692.fasta	BTB/POZ domain-containing protein KCTD4
refseqs_squamata_FAM003716.fasta	crk-like protein
refseqs_squamata_FAM003726.fasta	glutamate-rich WD repeat-containing protein 1
refseqs_squamata_FAM003745.fasta	apoptosis regulator BAX
refseqs_squamata_FAM003825.fasta	epidermal growth factor receptor kinase substrate 8-like protein 1
refseqs_squamata_FAM003845.fasta	bolA-like protein 2
refseqs_squamata_FAM003874.fasta	trans-1,2-dihydrobenzene-1,2-diol dehydrogenase-like
refseqs_squamata_FAM003886.fasta	urotensin-2 receptor-like
refseqs_squamata_FAM003903.fasta	RNA-binding protein 12-like isoform X2
refseqs_squamata_FAM003921.fasta	guanine nucleotide exchange factor MSS4
refseqs_squamata_FAM003933.fasta	myogenin
refseqs_squamata_FAM003935.fasta	growth/differentiation factor 5
refseqs_squamata_FAM003943.fasta	short transient receptor potential channel 4-associated protein
refseqs_squamata_FAM003949.fasta	charged multivesicular body protein 4b
refseqs_squamata_FAM003976.fasta	eukaryotic translation initiation factor 2 subunit 2
refseqs_squamata_FAM003977.fasta	eukaryotic translation initiation factor 6
refseqs_squamata_FAM003988.fasta	DNA-directed RNA polymerase I subunit RPA1
refseqs_squamata_FAM004031.fasta	C-C chemokine receptor type 10
refseqs_squamata_FAM004035.fasta	angiotensin-converting enzyme
refseqs_squamata_FAM004046.fasta	C-C chemokine receptor type 7

refseqs_squamata_FAM004047.fasta	telethonin
refseqs_squamata_FAM004057.fasta	insulin-like growth factor-binding protein 4
refseqs_squamata_FAM004076.fasta	60S ribosomal protein L19
refseqs_squamata_FAM004079.fasta	protein Wnt-9b
refseqs_squamata_FAM004161.fasta	glycine dehydrogenase (decarboxylating), mitochondrial
refseqs_squamata_FAM004167.fasta	zinc finger protein 367
refseqs_squamata_FAM004189.fasta	cocaine- and amphetamine-regulated transcript protein
refseqs_squamata_FAM004209.fasta	homeobox protein orthopedia
refseqs_squamata_FAM004232.fasta	serine/threonine-protein kinase PLK2
refseqs_squamata_FAM004247.fasta	39S ribosomal protein S30, mitochondrial
refseqs_squamata_FAM004251.fasta	guanine deaminase
refseqs_squamata_FAM004273.fasta	protein kish-A
refseqs_squamata_FAM004296.fasta	5,6-dihydroxyindole-2-carboxylic acid oxidase
refseqs_squamata_FAM004304.fasta	peptidylprolyl isomerase domain and WD repeat-containing protein 1
refseqs_squamata_FAM004316.fasta	solute carrier family 25 member 46
refseqs_squamata_FAM004340.fasta	uncharacterized protein C9orf85 homolog
refseqs_squamata_FAM004342.fasta	iron-sulfur cluster assembly 1 homolog, mitochondrial
refseqs_squamata_FAM004361.fasta	heat shock protein beta-3
refseqs_squamata_FAM004421.fasta	microtubule-associated protein 1B
refseqs_squamata_FAM004433.fasta	5-hydroxytryptamine receptor 1A
refseqs_squamata_FAM004447.fasta	transmembrane emp24 domain-containing protein 7-like
refseqs_squamata_FAM004459.fasta	adenylate kinase isoenzyme 6
refseqs_squamata_FAM004471.fasta	ubiquitin-like protein ATG12
refseqs_squamata_FAM004472.fasta	cysteine dioxygenase type 1
refseqs_squamata_FAM004486.fasta	proteinase-activated receptor 3
refseqs_squamata_FAM004491.fasta	osteoclast-stimulating factor 1
refseqs_squamata_FAM004534.fasta	protein limb expression 1 homolog
refseqs_squamata_FAM004543.fasta	G-protein coupled receptor 26
refseqs_squamata_FAM004717.fasta	zinc finger CCCH domain-containing protein 15
refseqs_squamata_FAM004723.fasta	mesogenin-1
refseqs_squamata_FAM004755.fasta	60S ribosomal protein L4
refseqs_squamata_FAM004761.fasta	U2 small nuclear ribonucleoprotein A'
refseqs_squamata_FAM004771.fasta	chondroitin sulfate synthase 1
refseqs_squamata_FAM004793.fasta	kelch repeat and BTB domain-containing protein 13
refseqs_squamata_FAM004807.fasta	microfibrillar-associated protein 1
refseqs_squamata_FAM004848.fasta	alpha-N-acetylglucosaminidase
refseqs_squamata_FAM004860.fasta	sclerostin
refseqs_squamata_FAM004865.fasta	neurexophilin-3
refseqs_squamata_FAM004878.fasta	homeobox protein Hox-B6
refseqs_squamata_FAM004879.fasta	homeobox protein Hox-B5
refseqs_squamata_FAM004882.fasta	dual specificity protein phosphatase 3
refseqs_squamata_FAM004895.fasta	homeobox protein DLX-3
refseqs_squamata_FAM004904.fasta	pyridoxine-5'-phosphate oxidase
refseqs_squamata_FAM004907.fasta	homeobox protein Hox-B2
refseqs_squamata_FAM004914.fasta	potassium voltage-gated channel subfamily A member 7
refseqs_squamata_FAM004933.fasta	homeobox protein Hox-B4
refseqs_squamata_FAM004970.fasta	lutropin subunit beta-like
refseqs_squamata_FAM005005.fasta	developmentally-regulated GTP-binding protein 1
refseqs_squamata_FAM005032.fasta	splicing factor 3B subunit 6
refseqs_squamata_FAM005057.fasta	cationic amino acid transporter 4
refseqs_squamata_FAM005063.fasta	forkhead box protein O1
refseqs_squamata_FAM005090.fasta	reticulon-4 receptor
refseqs_squamata_FAM005140.fasta	calicin
refseqs_squamata_FAM005180.fasta	aquaporin-3
refseqs_squamata_FAM005183.fasta	protein farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha
refseqs_squamata_FAM005208.fasta	protein NipSnap homolog 3B-like
refseqs_squamata_FAM005239.fasta	ATP-binding cassette sub-family D member 1
refseqs_squamata_FAM005264.fasta	caveolin-3-like
refseqs_squamata_FAM005268.fasta	red-sensitive opsin
refseqs_squamata_FAM005287.fasta	hsp90 co-chaperone Cdc37
refseqs_squamata_FAM005350.fasta	synaptophysin
refseqs_squamata_FAM005372.fasta	UPF0160 protein MYG1, mitochondrial
refseqs_squamata_FAM005388.fasta	homeobox protein Hox-C12
refseqs_squamata_FAM005412.fasta	transcription factor Sp5-like
refseqs_squamata_FAM005413.fasta	G-protein coupled receptor 182
refseqs_squamata_FAM005414.fasta	homeobox protein Hox-C13
refseqs_squamata_FAM005420.fasta	25-hydroxyvitamin D-1 alpha hydroxylase, mitochondrial
refseqs_squamata_FAM005421.fasta	homeobox protein Hox-C11
refseqs_squamata_FAM005426.fasta	inhibin beta E chain
refseqs_squamata_FAM005438.fasta	homeobox protein Hox-C9
refseqs_squamata_FAM005440.fasta	homeobox protein Hox-C8
refseqs_squamata_FAM005441.fasta	aquaporin-2
refseqs_squamata_FAM005443.fasta	mitogen-activated protein kinase kinase kinase 12
refseqs_squamata_FAM005444.fasta	homeobox protein Hox-C4
refseqs_squamata_FAM005453.fasta	keratin, type I cytoskeletal 18
refseqs_squamata_FAM005497.fasta	caveolae-associated protein 4
refseqs_squamata_FAM005503.fasta	mediator of RNA polymerase II transcription subunit 10
refseqs_squamata_FAM005509.fasta	NADH dehydrogenase [ubiquinone] iron-sulfur protein 6, mitochondrial
refseqs_squamata_FAM005511.fasta	T-complex protein 1 subunit epsilon
refseqs_squamata_FAM005512.fasta	homeobox protein Nkx-2.2

refseqs_squamata_FAM005518.fasta	grpE protein homolog 1, mitochondrial
refseqs_squamata_FAM005519.fasta	transcriptional adapter 2-beta
refseqs_squamata_FAM005527.fasta	cleft lip and palate transmembrane protein 1-like protein
refseqs_squamata_FAM005532.fasta	RNA cytosine C(5)-methyltransferase NSUN2
refseqs_squamata_FAM005570.fasta	death-associated protein 1
refseqs_squamata_FAM005575.fasta	5'-3' exoribonuclease 2
refseqs_squamata_FAM005576.fasta	uncharacterized protein C5orf49 homolog
refseqs_squamata_FAM005608.fasta	regulatory factor X-associated protein
refseqs_squamata_FAM005640.fasta	charged multivesicular body protein 2b
refseqs_squamata_FAM005641.fasta	nuclear receptor subfamily 0 group B member 1
refseqs_squamata_FAM005643.fasta	superoxide dismutase [Cu-Zn]
refseqs_squamata_FAM005674.fasta	claudin-14
refseqs_squamata_FAM005690.fasta	mycophenolic acid acyl-glucuronide esterase, mitochondrial
refseqs_squamata_FAM005703.fasta	frizzled-4
refseqs_squamata_FAM005704.fasta	protein POLR1D
refseqs_squamata_FAM005706.fasta	dermatopontin
refseqs_squamata_FAM005719.fasta	trifunctional purine biosynthetic protein adenosine-3-like
refseqs_squamata_FAM005724.fasta	G-protein coupled receptor 12
refseqs_squamata_FAM005749.fasta	beta-secretase 2
refseqs_squamata_FAM005765.fasta	mesenteric estrogen-dependent adipogenesis protein
refseqs_squamata_FAM005793.fasta	A disintegrin and metalloproteinase with thrombospondin motifs 1
refseqs_squamata_FAM005846.fasta	60S ribosomal protein L24
refseqs_squamata_FAM005860.fasta	collagenase 3-like
refseqs_squamata_FAM005875.fasta	renin receptor
refseqs_squamata_FAM005885.fasta	membrane transport protein XK
refseqs_squamata_FAM005903.fasta	mastermind-like protein 2
refseqs_squamata_FAM005912.fasta	guanylate cyclase soluble subunit alpha-2
refseqs_squamata_FAM005917.fasta	protein shisa-2 homolog
refseqs_squamata_FAM005922.fasta	immunoglobulin-like domain-containing receptor 1
refseqs_squamata_FAM005984.fasta	thioredoxin-like protein 4B
refseqs_squamata_FAM005991.fasta	A disintegrin and metalloproteinase with thrombospondin motifs 5
refseqs_squamata_FAM005996.fasta	sodium/potassium-transporting ATPase subunit beta-1
refseqs_squamata_FAM006002.fasta	cilia- and flagella-associated protein 300
refseqs_squamata_FAM006006.fasta	protein FAM243A
refseqs_squamata_FAM006009.fasta	eukaryotic translation initiation factor 2 subunit 3
refseqs_squamata_FAM006035.fasta	acetyl-CoA acetyltransferase, mitochondrial
refseqs_squamata_FAM006060.fasta	mitochondrial pyruvate carrier 2
refseqs_squamata_FAM006084.fasta	neutrophil cytosol factor 4
refseqs_squamata_FAM006092.fasta	heme oxygenase 1
refseqs_squamata_FAM006093.fasta	tRNA-splicing ligase RtcB homolog
refseqs_squamata_FAM006118.fasta	myoglobin
refseqs_squamata_FAM006124.fasta	monocarboxylate transporter 3
refseqs_squamata_FAM006132.fasta	endoplasmin
refseqs_squamata_FAM006167.fasta	metalloproteinase inhibitor 3
refseqs_squamata_FAM006200.fasta	probable ATP-dependent RNA helicase DDX17
refseqs_squamata_FAM006208.fasta	aldehyde dehydrogenase, mitochondrial
refseqs_squamata_FAM006224.fasta	stromal cell-derived factor 2-like protein 1
refseqs_squamata_FAM006239.fasta	claudin-5
refseqs_squamata_FAM006268.fasta	testis-specific serine/threonine-protein kinase 1-like
refseqs_squamata_FAM006387.fasta	zinc finger protein SNAI1
refseqs_squamata_FAM006417.fasta	E3 ubiquitin-protein ligase RNF114
refseqs_squamata_FAM006442.fasta	dysbindin domain-containing protein 2
refseqs_squamata_FAM006540.fasta	UPF0184 protein C9orf16 homolog
refseqs_squamata_FAM006556.fasta	ras-related protein Rab-14
refseqs_squamata_FAM006604.fasta	protein Niban 2
refseqs_squamata_FAM006618.fasta	prostaglandin E synthase 2
refseqs_squamata_FAM006622.fasta	putative UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase LOC100288842
refseqs_squamata_FAM006709.fasta	homeobox protein BarH-like 2
refseqs_squamata_FAM006711.fasta	glutamate receptor ionotropic, kainate 4
refseqs_squamata_FAM006767.fasta	laminin subunit gamma-1
refseqs_squamata_FAM006780.fasta	regulator of G-protein signaling 1
refseqs_squamata_FAM006783.fasta	regulator of G-protein signaling 5
refseqs_squamata_FAM006787.fasta	regulator of G-protein signaling 2
refseqs_squamata_FAM006797.fasta	nicotinamide/nicotinic acid mononucleotide adenylyltransferase 2
refseqs_squamata_FAM006807.fasta	regulator of G-protein signaling 21
refseqs_squamata_FAM006846.fasta	G-protein coupled receptor 52
refseqs_squamata_FAM006860.fasta	E3 ubiquitin-protein ligase RNF139
refseqs_squamata_FAM006885.fasta	beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase 3
refseqs_squamata_FAM006916.fasta	RNA polymerase-associated protein LEO1
refseqs_squamata_FAM006918.fasta	photoreceptor-specific nuclear receptor
refseqs_squamata_FAM006994.fasta	F-box and leucine-rich protein 22
refseqs_squamata_FAM007023.fasta	transmembrane channel-like protein 3
refseqs_squamata_FAM007036.fasta	solute carrier family 49 member 4
refseqs_squamata_FAM007039.fasta	protein CNPPD1
refseqs_squamata_FAM007047.fasta	lactase-phlorizin hydrolase
refseqs_squamata_FAM007052.fasta	inhibin beta B chain
refseqs_squamata_FAM007056.fasta	cyclin-dependent kinase 5 activator 2
refseqs_squamata_FAM007067.fasta	DNA replication licensing factor MCM6
refseqs_squamata_FAM007086.fasta	ankyrin repeat domain-containing protein 13D

refseqs_squamata_FAM007105.fasta	inhibin alpha chain
refseqs_squamata_FAM007112.fasta	tumor necrosis factor-inducible gene 6 protein
refseqs_squamata_FAM007120.fasta	angio-associated migratory cell protein
refseqs_squamata_FAM007138.fasta	desmin
refseqs_squamata_FAM007140.fasta	rho-related GTP-binding protein RhoE
refseqs_squamata_FAM007152.fasta	glutathione peroxidase 2
refseqs_squamata_FAM007160.fasta	sorting nexin-4
refseqs_squamata_FAM007269.fasta	phosphoglycerate kinase 1
refseqs_squamata_FAM007301.fasta	pannexin-3
refseqs_squamata_FAM007309.fasta	mitochondrial RNA pseudouridine synthase RPUSD4
refseqs_squamata_FAM007355.fasta	transcriptional and immune response regulator
refseqs_squamata_FAM007359.fasta	steroidogenic acute regulatory protein, mitochondrial
refseqs_squamata_FAM007400.fasta	adhesion G protein-coupled receptor A2
refseqs_squamata_FAM007409.fasta	transcription factor 7-like 1
refseqs_squamata_FAM007414.fasta	putative aspartate aminotransferase, cytoplasmic 2
refseqs_squamata_FAM007417.fasta	D(1) dopamine receptor-like
refseqs_squamata_FAM007421.fasta	nudC domain-containing protein 3
refseqs_squamata_FAM007424.fasta	neuropeptide Y receptor type 1-like
refseqs_squamata_FAM007461.fasta	trafficking protein particle complex subunit 6B
refseqs_squamata_FAM007487.fasta	dehydrogenase/reductase SDR family member 7
refseqs_squamata_FAM007493.fasta	G-protein coupled receptor 176
refseqs_squamata_FAM007559.fasta	serine palmitoyltransferase small subunit A
refseqs_squamata_FAM007575.fasta	probable E3 ubiquitin-protein ligase IRF2BPL
refseqs_squamata_FAM007581.fasta	cytosolic phospholipase A2 zeta
refseqs_squamata_FAM007589.fasta	proto-oncogene c-Fos
refseqs_squamata_FAM007594.fasta	calcineurin B homologous protein 1
refseqs_squamata_FAM007606.fasta	actin-related protein 10
refseqs_squamata_FAM007615.fasta	sprouty-related, EVH1 domain-containing protein 1
refseqs_squamata_FAM007622.fasta	26S proteasome regulatory subunit 6A
refseqs_squamata_FAM007633.fasta	protein RD3-like
refseqs_squamata_FAM007656.fasta	ER membrane protein complex subunit 7
refseqs_squamata_FAM007665.fasta	26S proteasome regulatory subunit 10B
refseqs_squamata_FAM007673.fasta	V(D)J recombination-activating protein 2
refseqs_squamata_FAM007675.fasta	V(D)J recombination-activating protein 1
refseqs_squamata_FAM007693.fasta	pinin
refseqs_squamata_FAM007696.fasta	homeobox protein SIX6
refseqs_squamata_FAM007701.fasta	RNA transcription, translation and transport factor protein
refseqs_squamata_FAM007713.fasta	vacuolar protein sorting-associated protein 18 homolog
refseqs_squamata_FAM007723.fasta	apoptosis inhibitor 5
refseqs_squamata_FAM007738.fasta	neuroglobin
refseqs_squamata_FAM007739.fasta	basic leucine zipper transcriptional factor ATF-like
refseqs_squamata_FAM007781.fasta	zinc transporter ZIP9
refseqs_squamata_FAM007819.fasta	peroxisomal biogenesis factor 16
refseqs_squamata_FAM007838.fasta	acyl-coenzyme A thioesterase 1-like
refseqs_squamata_FAM007845.fasta	proteasome subunit alpha type-6
refseqs_squamata_FAM007851.fasta	activator of 90 kDa heat shock protein ATPase homolog 1 isoform X1
refseqs_squamata_FAM007853.fasta	phosphatidylglycerophosphatase and protein-tyrosine phosphatase 1
refseqs_squamata_FAM007898.fasta	26S proteasome regulatory subunit 4
refseqs_squamata_FAM007902.fasta	SNW domain-containing protein 1
refseqs_squamata_FAM007912.fasta	NPC intracellular cholesterol transporter 2
refseqs_squamata_FAM007939.fasta	HHIP-like protein 1
refseqs_squamata_FAM007948.fasta	proteasome subunit alpha type-3
refseqs_squamata_FAM007968.fasta	transmembrane protein 151B-like
refseqs_squamata_FAM007972.fasta	signal recognition particle 14 kDa protein
refseqs_squamata_FAM007982.fasta	synaptojanin-2-binding protein
refseqs_squamata_FAM007990.fasta	fibroblast growth factor 3
refseqs_squamata_FAM007992.fasta	fibroblast growth factor 19
refseqs_squamata_FAM007998.fasta	G1/S-specific cyclin-D1
refseqs_squamata_FAM008049.fasta	hepatocyte nuclear factor 3-gamma
refseqs_squamata_FAM008100.fasta	nuclear receptor subfamily 0 group B member 2
refseqs_squamata_FAM008129.fasta	zinc finger protein 593
refseqs_squamata_FAM008132.fasta	digestive cysteine proteinase 2-like
refseqs_squamata_FAM008155.fasta	lariat debranching enzyme
refseqs_squamata_FAM008160.fasta	nuclear migration protein nudC
refseqs_squamata_FAM008161.fasta	NF-kappa-B inhibitor beta
refseqs_squamata_FAM008180.fasta	protein lin-28 homolog A
refseqs_squamata_FAM008184.fasta	transmembrane protein 222
refseqs_squamata_FAM008212.fasta	exostosin-like 1
refseqs_squamata_FAM008227.fasta	forkhead box protein O6
refseqs_squamata_FAM008243.fasta	selenoprotein N
refseqs_squamata_FAM008271.fasta	dnaJ homolog subfamily C member 8
refseqs_squamata_FAM008289.fasta	U5 small nuclear ribonucleoprotein 40 kDa protein
refseqs_squamata_FAM008326.fasta	pentraxin-related protein PTX3
refseqs_squamata_FAM008336.fasta	growth hormone secretagogue receptor type 1
refseqs_squamata_FAM008337.fasta	claudin-11
refseqs_squamata_FAM008356.fasta	probable cationic amino acid transporter
refseqs_squamata_FAM008364.fasta	A-kinase anchor protein 4-like
refseqs_squamata_FAM008376.fasta	26S proteasome non-ATPase regulatory subunit 2
refseqs_squamata_FAM008454.fasta	C-type lectin domain family 3 member A
refseqs_squamata_FAM008463.fasta	ubiquinone biosynthesis protein COQ9, mitochondrial

refseqs_squamata_FAM008499.fasta	SS18-like protein 2
refseqs_squamata_FAM008508.fasta	group XV phospholipase A2
refseqs_squamata_FAM008542.fasta	tyrosine--tRNA ligase, mitochondrial
refseqs_squamata_FAM008571.fasta	dexamethasone-induced Ras-related protein 1
refseqs_squamata_FAM008603.fasta	suppressor of cytokine signaling 1
refseqs_squamata_FAM008608.fasta	probable G-protein coupled receptor 139
refseqs_squamata_FAM008609.fasta	noggin-2-like
refseqs_squamata_FAM008628.fasta	class A basic helix-loop-helix protein 15
refseqs_squamata_FAM008631.fasta	H(+)/Cl(-) exchange transporter 7
refseqs_squamata_FAM008639.fasta	insulin receptor
refseqs_squamata_FAM008641.fasta	WAP, Kazal, immunoglobulin, Kunitz and NTR domain-containing protein 1
refseqs_squamata_FAM008650.fasta	parvalbumin, thymic CPV3-like
refseqs_squamata_FAM008671.fasta	ADP-ribosylation factor-like protein 6-interacting protein 1
refseqs_squamata_FAM008725.fasta	guanine nucleotide exchange protein SMCR8
refseqs_squamata_FAM008780.fasta	tektin-5
refseqs_squamata_FAM008789.fasta	V-type proton ATPase 16 kDa proteolipid subunit
refseqs_squamata_FAM008827.fasta	chemokine-like receptor 1
refseqs_squamata_FAM008837.fasta	transmembrane protein 204
refseqs_squamata_FAM008852.fasta	vomeronasal type-2 receptor 1-like
refseqs_squamata_FAM008857.fasta	thymidylate kinase
refseqs_squamata_FAM008875.fasta	clarin-1
refseqs_squamata_FAM008878.fasta	RING-box protein 2 isoform X1
refseqs_squamata_FAM008881.fasta	protein mab-21-like 4
refseqs_squamata_FAM008885.fasta	translocon-associated protein subunit gamma
refseqs_squamata_FAM008889.fasta	probable G-protein coupled receptor 149
refseqs_squamata_FAM008890.fasta	secretogranin-2
refseqs_squamata_FAM008894.fasta	acetyl-coenzyme A transporter 1
refseqs_squamata_FAM008912.fasta	carbohydrate sulfotransferase 2
refseqs_squamata_FAM008921.fasta	insulin receptor substrate 1
refseqs_squamata_FAM008922.fasta	transmembrane 4 L6 family member 4
refseqs_squamata_FAM008923.fasta	mast cell carboxypeptidase A-like
refseqs_squamata_FAM008925.fasta	phenylalanine--tRNA ligase beta subunit
refseqs_squamata_FAM008992.fasta	39S ribosomal protein L44, mitochondrial
refseqs_squamata_FAM009003.fasta	sodium/potassium-transporting ATPase subunit beta-3
refseqs_squamata_FAM009025.fasta	60S ribosomal protein L5
refseqs_squamata_FAM009028.fasta	transmembrane emp24 domain-containing protein 5
refseqs_squamata_FAM009081.fasta	selenoprotein F
refseqs_squamata_FAM009134.fasta	probable G-protein coupled receptor 88
refseqs_squamata_FAM009170.fasta	ADP-ribosylation factor-like protein 4C
refseqs_squamata_FAM009179.fasta	elongation factor 1-beta
refseqs_squamata_FAM009209.fasta	COP9 signalosome complex subunit 8
refseqs_squamata_FAM009224.fasta	integral membrane protein 2B
refseqs_squamata_FAM009225.fasta	collagen alpha-1(VI) chain
refseqs_squamata_FAM009227.fasta	myosin light chain 1/3, skeletal muscle isoform isoform X1
refseqs_squamata_FAM009240.fasta	caveolae-associated protein 2
refseqs_squamata_FAM009253.fasta	60 kDa heat shock protein, mitochondrial
refseqs_squamata_FAM009336.fasta	sia-alpha-2,3-Gal-beta-1,4-GlcNAc-R:alpha 2,8-sialyltransferase
refseqs_squamata_FAM009391.fasta	ATP synthase subunit alpha, mitochondrial
refseqs_squamata_FAM009407.fasta	prostaglandin E2 receptor EP4 subtype
refseqs_squamata_FAM009427.fasta	Golgi phosphoprotein 3
refseqs_squamata_FAM009438.fasta	G-patch domain and KOW motifs-containing protein
refseqs_squamata_FAM009460.fasta	relaxin-3 receptor 1
refseqs_squamata_FAM009480.fasta	protocadherin-18
refseqs_squamata_FAM009499.fasta	fibrinogen gamma chain
refseqs_squamata_FAM009535.fasta	fibrinogen beta chain
refseqs_squamata_FAM009543.fasta	carboxypeptidase E
refseqs_squamata_FAM009574.fasta	40S ribosomal protein S3a
refseqs_squamata_FAM009587.fasta	FRAS1-related extracellular matrix protein 3
refseqs_squamata_FAM009646.fasta	cyclin-dependent kinase 4 inhibitor C
refseqs_squamata_FAM009653.fasta	neutrophil cytosol factor 1
refseqs_squamata_FAM009663.fasta	protein SMG8
refseqs_squamata_FAM009680.fasta	amyloid protein-binding protein 2
refseqs_squamata_FAM009693.fasta	proprotein convertase subtilisin/kexin type 9
refseqs_squamata_FAM009702.fasta	transcription factor AP-1
refseqs_squamata_FAM009763.fasta	phospholipid phosphatase 3
refseqs_squamata_FAM009777.fasta	nardilysin
refseqs_squamata_FAM009791.fasta	WD repeat and SOCS box-containing protein 1
refseqs_squamata_FAM009872.fasta	protein canopy homolog 3
refseqs_squamata_FAM009873.fasta	glycine N-methyltransferase
refseqs_squamata_FAM009882.fasta	protein phosphatase 1B isoform X2
refseqs_squamata_FAM009911.fasta	microsomal triglyceride transfer protein large subunit
refseqs_squamata_FAM009922.fasta	oligosaccharyltransferase complex subunit OSTC
refseqs_squamata_FAM009941.fasta	cysteine-rich PDZ-binding protein
refseqs_squamata_FAM009951.fasta	hydroxyacyl-coenzyme A dehydrogenase, mitochondrial
refseqs_squamata_FAM009956.fasta	homeobox protein MOX-2
refseqs_squamata_FAM009986.fasta	mRNA decay activator protein ZFP36L2
refseqs_squamata_FAM010003.fasta	myelin and lymphocyte protein
refseqs_squamata_FAM010006.fasta	dickkopf-related protein 2
refseqs_squamata_FAM010011.fasta	cytochrome c oxidase subunit 7A-related protein, mitochondrial
refseqs_squamata_FAM010017.fasta	sclerostin domain-containing protein 1

refseqs_squamata_FAM010022.fasta	T-complex protein 1 subunit eta
refseqs_squamata_FAM010030.fasta	peripherin-2
refseqs_squamata_FAM010045.fasta	zinc transporter 1
refseqs_squamata_FAM010066.fasta	extracellular tyrosine-protein kinase PKDCC
refseqs_squamata_FAM010087.fasta	leucine-rich repeat transmembrane neuronal protein 1
refseqs_squamata_FAM010088.fasta	visual pigment-like receptor peropsin
refseqs_squamata_FAM010093.fasta	stearoyl-CoA desaturase 5
refseqs_squamata_FAM010096.fasta	neutral and basic amino acid transport protein rBAT
refseqs_squamata_FAM010101.fasta	pituitary homeobox 2 isoform X1
refseqs_squamata_FAM010107.fasta	mitotic spindle assembly checkpoint protein MAD2A
refseqs_squamata_FAM010127.fasta	inactive tyrosine-protein kinase 7
refseqs_squamata_FAM010216.fasta	replication protein A 14 kDa subunit
refseqs_squamata_FAM010268.fasta	N-arachidonyl glycine receptor
refseqs_squamata_FAM010276.fasta	alcohol dehydrogenase class-3
refseqs_squamata_FAM010281.fasta	potassium-transporting ATPase subunit beta
refseqs_squamata_FAM010286.fasta	pyroglutamylated RF-amide peptide receptor
refseqs_squamata_FAM010362.fasta	mannosyl-oligosaccharide glucosidase
refseqs_squamata_FAM010379.fasta	eukaryotic initiation factor 4A-III
refseqs_squamata_FAM010526.fasta	dnaJ homolog subfamily A member 2
refseqs_squamata_FAM010530.fasta	DNA replication complex GINS protein PSF2
refseqs_squamata_FAM010603.fasta	cytochrome c oxidase copper chaperone
refseqs_squamata_FAM010629.fasta	probable G-protein coupled receptor 174
refseqs_squamata_FAM010662.fasta	alpha-galactosidase A
refseqs_squamata_FAM010678.fasta	heat shock transcription factor, Y-linked-like
refseqs_squamata_FAM010696.fasta	SH3 domain-binding glutamic acid-rich-like protein
refseqs_squamata_FAM010720.fasta	integral membrane protein 2A
refseqs_squamata_FAM010762.fasta	syntaxin-3-like
refseqs_squamata_FAM010769.fasta	methionine--tRNA ligase, mitochondrial
refseqs_squamata_FAM010770.fasta	glucose-dependent insulinotropic receptor
refseqs_squamata_FAM010784.fasta	protein phosphatase 1 regulatory subunit 1C isoform X1
refseqs_squamata_FAM010785.fasta	homeobox protein Hox-D4
refseqs_squamata_FAM010792.fasta	solute carrier family 40 member 1
refseqs_squamata_FAM010814.fasta	tetratricopeptide repeat protein 30B
refseqs_squamata_FAM010818.fasta	growth/differentiation factor 8
refseqs_squamata_FAM010899.fasta	tyrosyl-DNA phosphodiesterase 2
refseqs_squamata_FAM010937.fasta	melanocortin receptor 5
refseqs_squamata_FAM010948.fasta	E3 ubiquitin-protein ligase NHLRC1
refseqs_squamata_FAM010949.fasta	melanocortin receptor 4
refseqs_squamata_FAM010959.fasta	uncharacterized protein C6orf62 homolog
refseqs_squamata_FAM010971.fasta	ras-related protein Rab-12
refseqs_squamata_FAM010992.fasta	protein APCDD1
refseqs_squamata_FAM011037.fasta	parathyroid hormone
refseqs_squamata_FAM011042.fasta	RNA polymerase-associated protein CTR9 homolog
refseqs_squamata_FAM011064.fasta	cathepsin D
refseqs_squamata_FAM011070.fasta	fin bud initiation factor homolog
refseqs_squamata_FAM011084.fasta	solute carrier family 15 member 3
refseqs_squamata_FAM011087.fasta	26S proteasome non-ATPase regulatory subunit 13
refseqs_squamata_FAM011096.fasta	reticulocalbin-1
refseqs_squamata_FAM011105.fasta	follitropin subunit beta
refseqs_squamata_FAM011163.fasta	ATP-sensitive inward rectifier potassium channel 11
refseqs_squamata_FAM011200.fasta	tumor susceptibility gene 101 protein isoform X3
refseqs_squamata_FAM011232.fasta	uncharacterized protein C11orf91 homolog
refseqs_squamata_FAM011254.fasta	CD81 antigen
refseqs_squamata_FAM011262.fasta	leucine-rich repeat-containing protein 10B
refseqs_squamata_FAM011300.fasta	3-ketoacyl-CoA thiolase, peroxisomal
refseqs_squamata_FAM011354.fasta	eukaryotic translation initiation factor 3 subunit I
refseqs_squamata_FAM011381.fasta	proteasome subunit beta type-2
refseqs_squamata_FAM011413.fasta	probable G-protein coupled receptor 158
refseqs_squamata_FAM011416.fasta	frizzled-8
refseqs_squamata_FAM011423.fasta	sonic hedgehog protein
refseqs_squamata_FAM011440.fasta	tissue alpha-L-fucosidase
refseqs_squamata_FAM011453.fasta	vasoactive intestinal polypeptide receptor 2
refseqs_squamata_FAM011475.fasta	vimentin
refseqs_squamata_FAM011482.fasta	ras-related protein Rab-18
refseqs_squamata_FAM011559.fasta	PX domain-containing protein 1
refseqs_squamata_FAM011560.fasta	forkhead box protein F2
refseqs_squamata_FAM011601.fasta	class E basic helix-loop-helix protein 22
refseqs_squamata_FAM011602.fasta	charged multivesicular body protein 4c
refseqs_squamata_FAM011618.fasta	gamma-glutamyl hydrolase
refseqs_squamata_FAM011620.fasta	desmoglein-1-beta-like
refseqs_squamata_FAM011638.fasta	neuropeptides B/W receptor type 1
refseqs_squamata_FAM011639.fasta	COP9 signalosome complex subunit 5
refseqs_squamata_FAM011648.fasta	transthyretin
refseqs_squamata_FAM011687.fasta	zinc finger protein SNAI2
refseqs_squamata_FAM011688.fasta	CCAAT/enhancer-binding protein delta
refseqs_squamata_FAM011740.fasta	minichromosome maintenance domain-containing protein 2
refseqs_squamata_FAM011755.fasta	60S ribosomal protein L7
refseqs_squamata_FAM011761.fasta	carbonic anhydrase 2
refseqs_squamata_FAM011765.fasta	deubiquitinating protein VCIP135
refseqs_squamata_FAM011790.fasta	zinc finger and BTB domain-containing protein 10

refseqs_squamata_FAM011813.fasta	coatomer subunit delta
refseqs_squamata_FAM011848.fasta	coiled-coil domain-containing protein 3
refseqs_squamata_FAM011857.fasta	protein adenylyltransferase SelO, mitochondrial
refseqs_squamata_FAM011932.fasta	EF-hand calcium-binding domain-containing protein 10
refseqs_squamata_FAM011945.fasta	26S proteasome regulatory subunit 7
refseqs_squamata_FAM011948.fasta	fibroleukin
refseqs_squamata_FAM011970.fasta	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5
refseqs_squamata_FAM011975.fasta	leiomodin-2
refseqs_squamata_FAM011976.fasta	caveolin-2
refseqs_squamata_FAM011985.fasta	protein lifeguard 4
refseqs_squamata_FAM012047.fasta	vasopressin V1a receptor
refseqs_squamata_FAM012088.fasta	NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial
refseqs_squamata_FAM012128.fasta	syntaxin-4
refseqs_squamata_FAM012178.fasta	tetratricopeptide repeat protein 5
refseqs_squamata_FAM012181.fasta	type 1 phosphatidylinositol 4,5-bisphosphate 4-phosphatase
refseqs_squamata_FAM012240.fasta	angiopoietin-2
refseqs_squamata_FAM012241.fasta	SPRY domain-containing protein 4
refseqs_squamata_FAM012262.fasta	noggin
refseqs_squamata_FAM012278.fasta	cytokine receptor-like factor 3
refseqs_squamata_FAM012283.fasta	retinoid-inducible serine carboxypeptidase
refseqs_squamata_FAM012324.fasta	myosin light polypeptide 6-like
refseqs_squamata_FAM012335.fasta	rho-related GTP-binding protein Rho6
refseqs_squamata_FAM012376.fasta	eukaryotic translation initiation factor 3 subunit E
refseqs_squamata_FAM012378.fasta	5-hydroxytryptamine receptor 1B
refseqs_squamata_FAM012389.fasta	outer dense fiber protein 1
refseqs_squamata_FAM012417.fasta	elongation of very long chain fatty acids protein 4
refseqs_squamata_FAM012426.fasta	cell cycle control protein 50A
refseqs_squamata_FAM012439.fasta	orexin receptor type 2
refseqs_squamata_FAM012440.fasta	BAG family molecular chaperone regulator 2
refseqs_squamata_FAM012453.fasta	protein MAL2
refseqs_squamata_FAM012454.fasta	sister chromatid cohesion protein DCC1
refseqs_squamata_FAM012458.fasta	lens fiber major intrinsic protein
refseqs_squamata_FAM012460.fasta	actin-binding Rho-activating protein
refseqs_squamata_FAM012462.fasta	proto-oncogene Wnt-1
refseqs_squamata_FAM012516.fasta	splicing factor 3A subunit 3
refseqs_squamata_FAM012577.fasta	succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial
refseqs_squamata_FAM012675.fasta	fat storage-inducing transmembrane protein 1
refseqs_squamata_FAM012692.fasta	protein phosphatase 1 regulatory subunit 3E
refseqs_squamata_FAM012702.fasta	twisted gastrulation protein homolog 1-like
refseqs_squamata_FAM012727.fasta	mediator of RNA polymerase II transcription subunit 19
refseqs_squamata_FAM012732.fasta	macrophage-expressed gene 1 protein-like
refseqs_squamata_FAM012738.fasta	39S ribosomal protein L16, mitochondrial
refseqs_squamata_FAM012741.fasta	apelin receptor
refseqs_squamata_FAM012758.fasta	BCL2/adenovirus E1B 19 kDa protein-interacting protein 3
refseqs_squamata_FAM012804.fasta	dnaJ homolog subfamily C member 9
refseqs_squamata_FAM012885.fasta	exosome complex component RRP40
refseqs_squamata_FAM012928.fasta	transmembrane 9 superfamily member 3
refseqs_squamata_FAM012938.fasta	vesicular acetylcholine transporter
refseqs_squamata_FAM012944.fasta	protein phosphatase 1 regulatory subunit 27
refseqs_squamata_FAM012957.fasta	neuropeptide FF receptor 1
refseqs_squamata_FAM012959.fasta	myeloid-associated differentiation marker-like protein 2
refseqs_squamata_FAM013042.fasta	protein disulfide-isomerase
refseqs_squamata_FAM013047.fasta	synaptogyrin-2
refseqs_squamata_FAM013048.fasta	thymidine kinase, cytosolic
refseqs_squamata_FAM013051.fasta	suppressor of cytokine signaling 3
refseqs_squamata_FAM013054.fasta	metalloproteinase inhibitor 2
refseqs_squamata_FAM013088.fasta	NEDD8-activating enzyme E1 regulatory subunit
refseqs_squamata_FAM013185.fasta	U7 snRNA-associated Sm-like protein LSm10
refseqs_squamata_FAM013239.fasta	potassium channel regulatory protein
refseqs_squamata_FAM013247.fasta	cell division cycle 5-like protein
refseqs_squamata_FAM013256.fasta	transcription factor SOX-11
refseqs_squamata_FAM013270.fasta	thyroid peroxidase
refseqs_squamata_FAM013337.fasta	potassium voltage-gated channel subfamily F member 1
refseqs_squamata_FAM013367.fasta	60S ribosomal protein L14
refseqs_squamata_FAM013484.fasta	alpha-1,3/1,6-mannosyltransferase ALG2
refseqs_squamata_FAM013500.fasta	charged multivesicular body protein 5
refseqs_squamata_FAM013506.fasta	guanylyl cyclase-activating protein 1
refseqs_squamata_FAM013510.fasta	interleukin-10
refseqs_squamata_FAM013512.fasta	vasopressin V1b receptor
refseqs_squamata_FAM013515.fasta	probable methyltransferase TARBP1
refseqs_squamata_FAM013552.fasta	transmembrane protein 151B
refseqs_squamata_FAM013573.fasta	transcription factor LBX1
refseqs_squamata_FAM013585.fasta	transcription factor SOX-14
refseqs_squamata_FAM013610.fasta	carboxypeptidase A1-like
refseqs_squamata_FAM013613.fasta	LYR motif-containing protein 2
refseqs_squamata_FAM013625.fasta	tyrosine-protein kinase FRK
refseqs_squamata_FAM013627.fasta	G-protein coupled receptor family C group 6 member A
refseqs_squamata_FAM013644.fasta	iodotyrosine deiodinase 1
refseqs_squamata_FAM013653.fasta	probable tubulin polyglutamylase TTLL2
refseqs_squamata_FAM013654.fasta	delta-like protein 1

refseqs_squamata_FAM013660.fasta	microtubule-associated proteins 1A/1B light chain 3C
refseqs_squamata_FAM013668.fasta	left-right determination factor 2-like
refseqs_squamata_FAM013675.fasta	denticleless protein homolog
refseqs_squamata_FAM013682.fasta	tetraspanin-2
refseqs_squamata_FAM013710.fasta	ATP synthase F(0) complex subunit B1, mitochondrial
refseqs_squamata_FAM013711.fasta	methylosome protein 50
refseqs_squamata_FAM013725.fasta	ubiquitin thioesterase OTU1
refseqs_squamata_FAM013754.fasta	renin
refseqs_squamata_FAM013762.fasta	protein BTG2
refseqs_squamata_FAM013765.fasta	mitochondrial import receptor subunit TOM6 homolog
refseqs_squamata_FAM013769.fasta	G1/S-specific cyclin-D3
refseqs_squamata_FAM013770.fasta	bystin
refseqs_squamata_FAM013779.fasta	probable E3 ubiquitin-protein ligase makorin-1
refseqs_squamata_FAM013833.fasta	ragulator complex protein LAMTOR5
refseqs_squamata_FAM013855.fasta	von Willebrand factor C domain-containing protein 2-like
refseqs_squamata_FAM013875.fasta	E3 ubiquitin-protein ligase HECTD3
refseqs_squamata_FAM013889.fasta	5-methylcytosine rRNA methyltransferase NSUN4
refseqs_squamata_FAM013976.fasta	alpha-2A adrenergic receptor
refseqs_squamata_FAM013981.fasta	dual specificity protein phosphatase 5
refseqs_squamata_FAM013995.fasta	calcium homeostasis modulator protein 1
refseqs_squamata_FAM013998.fasta	transcription initiation factor TFIID subunit 5
refseqs_squamata_FAM014012.fasta	39S ribosomal protein L43, mitochondrial
refseqs_squamata_FAM014014.fasta	twinkle protein, mitochondrial
refseqs_squamata_FAM014027.fasta	Hermansky-Pudlak syndrome 6 protein
refseqs_squamata_FAM014032.fasta	RRP12-like protein
refseqs_squamata_FAM014043.fasta	MARVEL domain-containing protein 1
refseqs_squamata_FAM014049.fasta	aspartate aminotransferase, cytoplasmic
refseqs_squamata_FAM014063.fasta	presenilins-associated rhomboid-like protein, mitochondrial
refseqs_squamata_FAM014075.fasta	DNA-directed RNA polymerase II subunit RPB4
refseqs_squamata_FAM014076.fasta	vesicle transport protein SFT2C
refseqs_squamata_FAM014118.fasta	dnaJ homolog subfamily B member 11
refseqs_squamata_FAM014160.fasta	magnesium transporter NIPA1
refseqs_squamata_FAM014200.fasta	acidic mammalian chitinase-like
refseqs_squamata_FAM014213.fasta	leucine rich adaptor protein 1
refseqs_squamata_FAM014218.fasta	prolactin-releasing peptide receptor-like
refseqs_squamata_FAM014219.fasta	potassium channel subfamily K member 18
refseqs_squamata_FAM014223.fasta	guanylate cyclase 2G-like
refseqs_squamata_FAM014225.fasta	cilia- and flagella-associated protein 58
refseqs_squamata_FAM014227.fasta	calcium homeostasis modulator protein 3
refseqs_squamata_FAM014336.fasta	proline-rich nuclear receptor coactivator 1
refseqs_squamata_FAM014360.fasta	prolyl endopeptidase
refseqs_squamata_FAM014409.fasta	cardiac phospholamban
refseqs_squamata_FAM014410.fasta	histone chaperone ASF1A
refseqs_squamata_FAM014415.fasta	gap junction alpha-1 protein
refseqs_squamata_FAM014425.fasta	hairy/enhancer-of-split related with YRPW motif protein 2
refseqs_squamata_FAM014464.fasta	p53 apoptosis effector related to PMP-22-like
refseqs_squamata_FAM014465.fasta	p53 apoptosis effector related to PMP-22-like
refseqs_squamata_FAM014474.fasta	vacuolar protein sorting-associated protein VTA1 homolog
refseqs_squamata_FAM014489.fasta	ras-related protein Rab-32
refseqs_squamata_FAM014536.fasta	T-complex protein 1 subunit alpha
refseqs_squamata_FAM014558.fasta	proteasome subunit beta type-1
refseqs_squamata_FAM014565.fasta	uncharacterized protein C1orf198 homolog
refseqs_squamata_FAM014568.fasta	piggyBac transposable element-derived protein 5
refseqs_squamata_FAM014569.fasta	polypeptide N-acetylgalactosaminyltransferase 2
refseqs_squamata_FAM014575.fasta	exocyst complex component 8
refseqs_squamata_FAM014583.fasta	potassium channel subfamily K member 1
refseqs_squamata_FAM014639.fasta	tetratricopeptide repeat protein 27
refseqs_squamata_FAM014663.fasta	NF-kappa-B inhibitor epsilon
refseqs_squamata_FAM014667.fasta	lactoylglutathione lyase
refseqs_squamata_FAM014671.fasta	potassium channel subfamily K member 5
refseqs_squamata_FAM014686.fasta	WD repeat-containing protein 43
refseqs_squamata_FAM014689.fasta	protein LBH
refseqs_squamata_FAM014697.fasta	poly [ADP-ribose] polymerase 1
refseqs_squamata_FAM014699.fasta	Golgi resident protein GCP60
refseqs_squamata_FAM014702.fasta	signal recognition particle 9 kDa protein
refseqs_squamata_FAM014737.fasta	protein KTI12 homolog
refseqs_squamata_FAM014773.fasta	dapper homolog 2
refseqs_squamata_FAM014775.fasta	olfactomedin-like protein 3
refseqs_squamata_FAM014779.fasta	NADH dehydrogenase subunit 6 (mitochondrion)
refseqs_squamata_FAM014780.fasta	NADH dehydrogenase subunit 5 (mitochondrion)
refseqs_squamata_FAM014785.fasta	NADH dehydrogenase subunit 5 (mitochondrion)
refseqs_squamata_FAM014795.fasta	cytochrome c oxidase subunit II (mitochondrion)
refseqs_squamata_FAM014797.fasta	cytochrome c oxidase subunit I (mitochondrion)
refseqs_squamata_FAM014805.fasta	NADH dehydrogenase subunit 2 (mitochondrion)
refseqs_squamata_FAM014806.fasta	NADH dehydrogenase subunit 6 (mitochondrion)
refseqs_squamata_FAM014807.fasta	NADH dehydrogenase subunit 5 (mitochondrion)
refseqs_squamata_FAM014808.fasta	NADH dehydrogenase subunit 4 (mitochondrion)
refseqs_squamata_FAM014809.fasta	NADH dehydrogenase subunit 4L (mitochondrion)
refseqs_squamata_FAM014812.fasta	ATP synthase F0 subunit 8 (mitochondrion)
refseqs_squamata_FAM015125.fasta	transcription factor 25

refseqs_squamata_FAM015269.fasta	39S ribosomal protein L20, mitochondrial
refseqs_squamata_FAM015558.fasta	protein disulfide-isomerase A3
refseqs_squamata_FAM016364.fasta	type I iodothyronine deiodinase
refseqs_squamata_FAM016561.fasta	D-3-phosphoglycerate dehydrogenase
refseqs_squamata_FAM017219.fasta	dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1
refseqs_squamata_FAM017336.fasta	von Hippel-Lindau disease tumor suppressor
refseqs_squamata_FAM017728.fasta	iron-sulfur cluster assembly enzyme ISCU, mitochondrial
refseqs_squamata_FAM017974.fasta	centriole, cilia and spindle-associated protein
refseqs_squamata_FAM018031.fasta	dynein regulatory complex protein 1
refseqs_squamata_FAM018303.fasta	NADH dehydrogenase subunit 4 (mitochondrion)
refseqs_squamata_FAM018309.fasta	ATP synthase F0 subunit 8 (mitochondrion)
refseqs_squamata_FAM028590.fasta	NADH dehydrogenase subunit 6 (mitochondrion)
refseqs_squamata_FAM028591.fasta	NADH dehydrogenase subunit 2 (mitochondrion)
refseqs_squamata_FAM029153.fasta	ATP synthase F0 subunit 8 (mitochondrion)
refseqs_squamata_FAM031882.fasta	ATP synthase F0 subunit 8 (mitochondrion)

10.4 Supplementary table 4: Taxonomy for PLA_2 and CRISP

Species	Family	Infraorder	Clade	Order	venomousness
Gekko japonicus	Gekkonidae	Gekkota		Squamata	non-venomous
Lacerta agilis	Lacertidae	Lacertibaenia	Laterata	Squamata	non-venomous
Podarcis muralis	Lacertidae	Lacertibaenia	Laterata	Squamata	non-venomous
Zootoca vivipara	Lacertidae	Lacertibaenia	Laterata	Squamata	non-venomous
Anolis carolinensis	Dactyloidae	Iguania	Toxicofera	Squamata	non-venomous
Pogona vitticeps	Agamidae	Iguania	Toxicofera	Squamata	venomous
Sceloporus undulatus	Phrynosomatidae	Iguania	Toxicofera	Squamata	non-venomous
Python bivittatus	Pythonidae	Serpentes	Toxicofera	Squamata	non-venomous
Pantherophis guttatus	Colubroidae	Serpentes	Toxicofera	Squamata	non-venomous
Rhabdophis tigrinus	Colubroidae	Serpentes	Toxicofera	Squamata	venomous
Thamnophis elegans	Colubroidae	Serpentes	Toxicofera	Squamata	non-venomous
Thamnophis sirtalis	Colubroidae	Serpentes	Toxicofera	Squamata	non-venomous
Laticauda semifasciata	Elapidae	Serpentes	Toxicofera	Squamata	venomous
Naja atra	Elapidae	Serpentes	Toxicofera	Squamata	venomous
Notechis scutatus	Elapidae	Serpentes	Toxicofera	Squamata	venomous
Ophiophagus hannah	Elapidae	Serpentes	Toxicofera	Squamata	venomous
Pseudonaja textilis	Elapidae	Serpentes	Toxicofera	Squamata	venomous
Agkistrodon piscivorus	Viperidae	Serpentes	Toxicofera	Squamata	venomous
Azemiops feae	Viperidae	Serpentes	Toxicofera	Squamata	venomous
Crotalus adamanteus	Viperidae	Serpentes	Toxicofera	Squamata	venomous
Crotalus tigris	Viperidae	Serpentes	Toxicofera	Squamata	venomous
Daboia russelii	Viperidae	Serpentes	Toxicofera	Squamata	venomous
Deinagkistrodon acutus	Viperidae	Serpentes	Toxicofera	Squamata	venomous
Gloydius intermedius	Viperidae	Serpentes	Toxicofera	Squamata	venomous
Ovophis okinavensis	Viperidae	Serpentes	Toxicofera	Squamata	venomous
Protobothrops flavoviridis	Viperidae	Serpentes	Toxicofera	Squamata	venomous
Protobothrops jerdonii	Viperidae	Serpentes	Toxicofera	Squamata	venomous
Protobothrops mucrosquamatus	Viperidae	Serpentes	Toxicofera	Squamata	venomous
Trimeresurus stejnegeri	Viperidae	Serpentes	Toxicofera	Squamata	venomous
Vipera berus	Viperidae	Serpentes	Toxicofera	Squamata	venomous
Heloderma suspectum	Helodermatidae	Anguimorpha	Toxicofera	Squamata	venomous
Varanus komodoensis	Varanidae	Anguimorpha	Toxicofera	Squamata	venomous
PLA_2 and CRISP - black					
additionally in CRISP - blue					

Declaration

I hereby declare that I have prepared this thesis independently and without outside assistance. I have not used any sources or aids other than those indicated.

The submitted written version of the thesis corresponds to the one on the electronic storage medium.

Furthermore, I certify that this work has not been submitted as a thesis elsewhere.