De novo genome assembly of Vipera ammodytes ammodytes reveals venom gene expansion and ecological adaptations

Andreas Laustsen

aho1a@bio.dtu.dk

Technical University of Denmark https://orcid.org/0000-0001-6918-5574

Weiqiao Rao

Esperanza Rivera-de-Torre

Technical University of Denmark https://orcid.org/0000-0002-0272-6150

Lorenzo Seneci

Technical University of Denmark

Min-hui Shi

BGI Research

Yaolei Zhang

BGI Research, Qingdao

Liang Lin

BGI Research

Tianming Lan

BGI-Shenzhen https://orcid.org/0000-0002-6934-0439

Jože Pungerčar

Jožef Stefan Institute

Siqi LIU

BGI Group https://orcid.org/0000-0001-9744-3681

Biological Sciences - Article

Keywords:

Posted Date: March 20th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-5578199/v1

Additional Declarations: There is **NO** Competing Interest.

Abstract

The Western nose-horned viper (*Vipera ammodytes ammodytes*) is the most venomous snake in Europe, known for its potent venom and the danger it poses to humans. To better understand the genetic architecture behind its venom composition and ecological adaptations, we assembled a high-quality *de novo* chromosomal-level genome for this species. Using a combination of PacBio and Illumina sequencing, we achieved a 131x coverage, surpassing previous snake genome assemblies. Our analysis revealed a substantial expansion of olfactory receptor genes, which may be linked to the species' adaptation to high-altitude and cold environments. Furthermore, we identified 112 venom-related genes across 15 families, with notable tandem duplications in snake venom metalloproteinases (SVMPs), shedding light on the molecular evolution of its venom. Transposable elements, particularly LINEs, were abundant, suggesting ongoing genomic activity. This assembly provides crucial insights into the evolutionary dynamics of venomous snakes and offers a valuable resource for comparative genomics, antivenom research, and development of venom-derived therapeutics.

1. Introduction

Snakes are a highly diverse and widespread lineage of reptiles comprising more than 3,500 species[1]. They are distributed across all continents except Antarctica and are found in a wide variety of ecosystems thanks to their remarkable evolutionary adaptations. These adaptations encompass thermosensitive labial or loreal pits, which aid in predation and thermoregulation. For venomous species, adaptations also involve the evolution of sophisticated venom delivery structures and a biochemical arsenal including peptides and proteins known as toxins that enables the subjugation of prey through molecular mechanisms[1,2]. In particular, the extreme diversification and specialization of snake venoms are a prime example of highly dynamic, adaptive evolution driven by strong selection pressures (e.g., dietary specialization and prey resistance to toxins)[3–6]. These unique traits make snakes ideal model organisms for evolutionary biology studies.

Despite scientific interest in genomics in general, snakes were largely left out of the widespread diffusion of genomics-based evolutionary research until recent years. To date, whole genomes are available for only ~ 60 snake species, most of which belong to the two main venomous families, Elapidae (cobras, mambas, sea snakes, kraits, etc.)[7,8] and Viperidae (vipers)[9–11]. Nonetheless, genomics approaches to venom evolution studies in snakes have already yielded intriguing findings, such as the differences in chromosomal location between toxin gene families and the dynamics of the gene regulatory network influencing the expression of toxin genes. Such discoveries underscore the rich potential of genomics as a tool to shed light on the evolutionary history and functional features of snake venom.

To contribute to the recent surge in snake genomics studies, we present here the reference genome and transcriptome of the Western nose-horned viper (*Vipera ammodytes ammodytes*), found from northeastern Italy to western Bulgaria[12]. *V. ammodytes ammodytes* is considered the most venomous snake in Europe[13,14], mainly because of the presence of potent neurotoxic phospholipases A_2 (PLA_2s), known as ammodytoxins, that are abundant in its venom[15–17]. Venom variation and extensive genetic differentiation are documented in different subspecies of *V. ammodytes* (*V. ammodytes ammodytes* included), with multiple previous studies describing the venom transcriptome and proteome of this viper in detail[17–19]. Our present in-depth study thus allows for a comprehensive comparative analysis of the *V. ammodytes ammodytes* venom profile at the genomic and transcriptomic levels, respectively. Furthermore, this is to our knowledge the best quality genome available for a true viper (subfamily Viperinae), as all other viper genomes reported thus far belong to pit vipers (subfamily Crotalinae)[20,21] or Fea's vipers (subfamily Azemiopinae)[22].or has limited coverage compromising the quality of the annotation (i.e., *V. berus berus* and *V. latastei*)[23,24]. The phylogenetic gap filled by the *V. ammodytes ammodytes* genome will thus enable further investigation into the evolution of key functional traits in vipers, which are at the root of the remarkable diversification and evolutionary success of this snake family.

2. Results

2.1 Genome sequencing, assembly, and annotation

2.1.1 Sequencing

To achieve high-quality sequencing of the whole *V. ammodytes ammodytes* genome, DNA fragments prepared from dissected liver tissue of an adult male specimen collected in the wild in Slovenia were sequenced both on a PacBio Sequel and on an Illumina HiSeq X-Ten platforms. In total, 160.1 giga base pairs (Gb) PacBio long reads (mean length: 16,900 base pairs (bp)) and 119.5 Gb of paired-end Illumina short reads (read1: 150 bp; read2: 150 bp) were generated (Table S1, Table S2).

2.1.2 Assembly and scaffolding

The estimated size of the *V. ammodytes ammodytes* genome based on k-mer analysis is 1.61 Gb. The initial assembly was carried out with Nextdenovo (v2.5.0) [25] assembler on the PacBio reads, followed by NextPolish (v1.4.0)[26] to increase the accuracy of single bases from Illumina data. Subsequently, Purge_dups (v1.2.5)[27] was used to purge the draft genome from redundancy and yielded a *de novo* assembled genome at the scaffold level with 262 contigs weighing 1.55 Gb and a contig N50 of 45.9 Mb (Table 1 and Table S3). To annotate the assembled genome at the chromosomal level, we used RaGOO[28] to anchor our *de novo* assembly to the Lataste's viper (*Vipera latastei*) genome (NCBI accession: GCA_02494585.1) as a reference. The *V. latastei* genome was chosen due to its phylogenetic relation and the similar genome size to that of *V. ammodytes ammodytes* (Table S4). The pairing produced a predicted chromosomal-level assembly comprising 18 pseudo-chromosomes (due to the primary comparison with *V. latastei* as reference mapping species which are in the same family and possess 18 chromosomes) ranging from 6.79 Mb to 352.46 Mb (Fig. 1 and Table S5). They correspond to the actual 21 chromosomes composed of 10 macrochromosomes (Chr1(1a + 1b), Chr2(2a + 2b), Chr3(3a + 3b), Chr4, Chr5, Chr6, and Chr7)), one sex chromosome (Chr18), and ten microchromosomes, the latter also denoted below as (pseudo-)chromosomes (Chr) 8 to 17 (Fig. 1). The presumed real number of chromosomes of *V. ammodytes ammodytes* would therefore be 21 pairs (2n = 42) based on its previously determined diploid karyotype of *V. ammodytes* is [29]. The chromosome mapping covered 99.96% of the scaffolds, (see Table 1 and Table S3 for a full report on the quality and size of the genome). Likewise, the predicted chromosome lengths are generally between 0.58 and 8.24% shorter in *V. ammodytes ammodytes* than in *V. latastei* reference (Table S6).

Table 1
V. ammodytes genome assembly statistics and comparisons to other snakes.

	V. ammodytes ammodytes	***C. tigris***	***C. viridis***	***O. hannah [O. bungarus]****	***N. naja***
Assembly size, Gb	1.55/**1.55**	1.61/1.59	1.34	1.66	1.79
Number of scaffolds	261/**24**	4228/160	7,043	296,399	1,897
Scaffold N50, Mb	45.9/**210.48**	2.11/207.72	179.89	0.23	223.35
Number of contigs	261	4,228	166,667	816,633	13,066
Contig N50, Mb	45.90	2.110	0.015	0.004	0.300
GC content, %	40.7	39.9/39.8	-	40.6	40.5
Protein-coding genes	20,775	18,240	17,352	18,506	23,248
venom protein-coding genes	158	51	92	232	139
BUSCO	97.1%	98.45%	75.40%	67.17%	76.41%

For *V. ammodytes*, *de novo* assembly statistics are shown on the left, and RaGOO assembly statistics are shown on the right with bold font (where appropriate). * Due to the latest taxonomic revision *O. hannah* was split in two different species, being *O. bungarus* the species used for this analysis[94].

To assess the quality and completeness of the assembly, we evaluated four indicators addressing the conservation of ancestral sequences and the GC content and distribution: (1) 98% of the sequenced short paired-end reads can be mapped back to the assembled genome, indicating that most of the repetitive sequences were assembled; (2) The single-peak distribution of sequencing depth as a function of GC content (Fig S1) suggests that this assembly has no contamination and no redundancy; (3) GC content distribution is similar to those of the *V. lastastei* and *N. naja* genomes (Fig S2); and (4) Genome completeness analysis with Benchmarking Universal Single-Copy Orthologs (BUSCO) showed that 97.1% of the *de novo* assembled genome was matched onto the Vertebrata gene set (n = 3,354, v5.2.2) (Table S7), similar to other recent squamate genomes[30–34]. These objective findings underline the high quality of the whole genome assembly for *V. ammodytes*.

2.1.3 Genome annotation

After masking repeat sequences with the RepeatMasker program[35], we predicted 20,775 protein-coding genes. We performed a BUSCO completeness study to check whether the predicted ORFs correspond to a complete assembly. Briefly, the BUSCO completeness score for the ORFs of *V. ammodytes ammodytes* was 86.5% using the Vertebrata gene set (n = 3,354, v5.2.2), with a duplication score of 4.9% (Table S8). The *V. ammodytes ammodytes* BUSCO score is comparable to other published snake genomes allowing further comparative evolutionary analysis. To analyze the potential function of the proteins encoded by the predicted ORFs, we mapped protein sequences of the predicted genes to KEGG[36], Swiss-Prot[37], InterPro[38], and TrEMBL[37] databases. This preliminary annotation assigned functions to these genes, generating a function annotation percentage of 96.59% (Table S9).

2.2 Transposable elements expansion

The newly sequenced *V. ammodytes ammodytes* genome was annotated using *de novo* prediction and homology-based methods. The results show that 53.75% of the *V. ammodytes ammodytes* genome is annotated as transposable elements (TEs) consisting of 4.29% DNA transposons, 41.87% Long Interspersed Nuclear Elements (LINEs), 1.93% Short

Interspersed Nuclear Elements (SINEs), and 14.35% Long Terminal Repeats (LTR). Some of the TEs were classified within two or more categories (Fig. 1 and Table 2).

Table 2
Transposable-element sequences annotated in the *V. ammodytes* genome.

Items	Repbase TEs		TE proteins		*De novo*		Combined TEs	
Type	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome	Length (Bp)	% in genome
DNA	44,089,502	2.85	4,422,643	0.29	24,907,933	1.61	66,352,486	4.29
LINE	219,270,988	14.19	188,554,801	12.20	486,278,036	31.46	647,118,082	41.87
SINE	26,678,643	1.73	0	0.00	3,580,206	0.23	29,861,221	1.93
LTR	33,823,512	2.19	52,335,264	3.39	181,716,713	11.76	221,841,537	14.35
Other	34,439	0.00	0	0.00	0	0.00	34,439	0.00
Unknown	0	0.00	0	0.00	3,493,092	0.23	3,493,092	0.23
Total	310,971,951	20.12	245,044,378	15.85	660,002,210	42.70	830,756,940	53.75

The TE content in the genome of *V. ammodytes* is higher than other snake species for which genome annotation data are publicly available, ranging between 33 and 48% (Table S10). In particular, a large proportion of LINEs in the TE landscape of *V. ammodytes ammodytes* distinguishes it from other venomous snakes. (Fig. 1, Fig. 2 and Table S10). Even though *V. ammodytes ammodytes* genome shares a wave of TE expansion with other venomous snake genomes (*N. naja* and *V. latastei*) (Fig. 2), the more recent expansion of TEs seen in venomous snakes appears to have started in the lineage leading to *V. ammodytes ammodytes* and *V. latastei*. The peak observed among the most recent elements implies that this expansion is probably ongoing. Detailed studies of TE expression in *V. ammodytes ammodytes* would provide more insights into how TEs have shaped the genome evolution.

2.3 Gene family expansion

2.3.1 Gene family analysis among reptiles

Taking advantage of the genomic information available for other snake species, we performed a comparative genomic analysis to identify expansions and contractions of gene families in the newly assembled genome of *V. ammodytes ammodytes*. We compared the gene families of *V. ammodytes* to those of seven other venomous snake species (*Deinagkistrodon acutus*, *Protobothrops mucrosquamatus*, *Crotalus tigris*, *Bungarus multicinctus*, *Naja naja*, *Notechis scutatus*, and *Pseudonaja textilis*) and three non-venomous snake species (*Thermophis baileyi*, *Thamnophis elegans*, and *Python bivittatus*), with the last non-venomous (*P. bivittatus*) as an outgroup. Statistical cluster analysis of homologous genes showed that the distribution of single and multicopy orthologs in *V. ammodytes ammodytes* was comparable to other species. However, the number of genes annotated as unique paralogs were higher in *V. ammodytes ammodytes* than in any other species, except for the *T. baileyi* (Fig S2). The CAFE method[39] was used to determine which gene families expanded or contracted across the species' phylogeny, with 311 gene families displaying significant expansion ($p < 0.05$), while 1,395 gene families displayed significant contraction (Fig. 3). As expected, due to their shared evolutionary history, venom-related gene family expansions, and ecological adaptations, *V. ammodytes ammodytes*, *D. acutus*, *P. mucrosquamatus*, and *C. tigris* (all vipers) were placed in a separate branch from other species (Fig. 3). Based on eight reference time points from the TimeTree database[40], we estimated *V. ammodytes ammodytes* divergence time to be 37.4 (33.8–42.8) million years ago.

Functional enrichment analysis of the 311 expanded gene families were conducted with functional pathway (also known as Kyoto Encyclopedia of Genes and Genomes (KEGG)) and specific biochemical functions (also known as Gene Ontology terms (GO)) enrichment analyses revealed that 207 gene families with 71 significant GO terms and 23 KEGG pathways were significantly enriched (adjusted $p < 0.01$), respectively (Fig S5-S8). The results indicate that the enriched GO terms of biological processes (BPs) include regulation of RNA metabolic process (GO:0051252), regulation of biosynthetic process (GO:0009889), regulation of gene expression (GO:0010468), among others (Fig S6). GO enrichment and KEGG pathway analysis were performed on 1,395 contracted genes in *V. ammodytes ammodytes* (Fig S7 and Fig S8) and the enriched GO terms of BPs related to G-protein coupled receptors (GPCRs) signaling pathway (GO:0007186), signal transduction (GO:0007165), and single organism signaling (GO:0044700) (Fig S8).

2.3.2 Olfactory receptors

An analysis of gene family expansion and construction in the *V. ammodytes ammodytes* genome revealed that the most notable changes in terms of GO terms occurred in chemosensory GPCRs. Vertebrates primarily use four prominent multigene families of GPCRs to detect odors: olfactory receptors (ORs), trace amine-associated receptors (TAARs), vomeronasal type-1 receptors (V1Rs), and vomeronasal type-2 receptors (V2Rs)[41].

To assess the proportion of each category within venomous snakes, we compared the newly sequenced data from *V. ammodytes ammodytes* (an ambush predator with poor eyesight and no thermal sensing) with *Crotalus tigris* (it marks its territory by leaving scent trails and using pheromones), *Naja naja* (an elapid that hunts actively rather than by ambush and relies more on eyesight than any viper),*Python bivittatus* (pythons can use the odor cues, as well as the visual and thermal cues provided by the infrared-sensitive pits on their snout scales and labial scales, to detect approaching prey) and *Thermophis baileyi* (also known as the hot spring snake, is currently the snake known to survive at the highest altitudes; it skillfully uses its excellent sense of smell and vision to locate prey). The number of TAAR genes in the five snake species *V. ammodytes ammodytes*, *C. tigris*, *P. bivittatus*, *N. naja*, and *T. baileyi* are 3, 6, 7, 3, and 6, respectively; whereas the V1R genes are absent in *V. ammodytes ammodytes*, present in two copies in *C. tigris*, and occur in one copy each in *P. bivittatus*, *N. naja*, and *T. baileyi*[42,43]. Olfactory-related GPCR genes were identified in the genome of *V. ammodytes ammodytes* with assemblies at the chromosome level. However, *V. ammodytes ammodytes* possessed a contracted family of V2Rs in comparison to the other four snakes (374 copies in *C. tigris*, 261 copies in *P. bivittatus*, 203 copies in *N. naja*, 124 copies in *T. baileyi* and only 73 copies in *V. ammodytes*). Although an expansion occurred in the repertoire of ORs genes in *V. ammodytes ammodytes* compared to *N. naja* (151 vs. 30 copies), it still shows a significant contraction compared to the other three snake species (374 copies in *C. tigris*, 616 copies in *P. bivittatus*, and 326 copies in *T. baileyi*) (Fig. 4).

2.3.3 Evolutionary stability of TRP channel genes involved in thermal sensing

The transient receptor potential (TRP) channel family is a multimodal signal detector that has been shown to respond to physical and chemical stimuli such as temperature, pressure, and toxic substances[44–46]. TRP genes were identified in the five snake genomes assembled at the chromosomic level (26 genes in *V. ammodytes ammodytes*, 28 in *C. tigris*, 44 in *P. bivittatus*, 38 in *N. naja* and 24 in *T. baileyi*) (Fig. 4,). One notable member of the TRP family is TRPA1 (transient receptor potential ankyrin 1), a calcium ion channel protein reported as the infrared receptor that responds to changes in temperature in pit vipers. Each of the five snake species have one TRPA1 gene in their genome, and no expansion or contraction in this gene family was detected in the genomes of the five snake species (Fig. 4).

2.4 Venom gene family analysis

Venom is a remarkable evolutionary adaptation of venomous snakes, allowing them to subdue prey chemically rather than physically. The toxin genes of *V. ammodytes* have been widely studied from the proteomics perspective[14,15]. However,

previous research on *V. ammodytes* venom was carried out without a high-quality complete genome for this species. Therefore, our goal was to characterize all venom genes in *V. ammodytes* and compare the entire genome with a venom gland transcriptome to provide an essential background for the evolution of venom in *V. ammodytes ammodytes* and the Viperidae snake family.

Our results reveal 112 toxin gene homologs from 15 gene families in the genome of *V. ammodytes ammodytes*, including 20 snake venom metalloproteinases (SVMPs), 19 snake C-type lectin-like proteins (snaclecs), 11 secreted phospholipases A_2 (sPLA$_2$s), 11 serine protease inhibitors (Kunitz-type) (SPi), nine snake venom serine proteases (SVSPs), seven 5'-nucleotidases (5Nase), seven *L*-amino acid oxidases (LAAOs), six metalloproteinase inhibitors (MPi), five hyaluronidases (Hyal), four cystatins, four cysteine-rich secretory protein (CRISP), three natriuretic peptides (NPs), three venom nerve growth factors (VNGFs), two vascular endothelial growth factors (VEGFs) and one veficolin (Fig. 5 and Table S11).

To determine the chromosomal location of all toxin genes in the *V. ammodytes ammodytes* genome, we analyzed the mapping results of the scaffolded assembly and compared the position of *V. ammodytes ammodytes* toxin genes to the orthologous toxin genes in *V. latastei*. Toxin genes are dispersed throughout 15 of the 18 pseudo-chromosomes, spanning across macrochromosomes (Chr1(1a + 1b), Chr2(2a + 2b), Chr3(3a + 3b), Chr4, Chr5, Chr6, and Chr7), sex chromosome Z, and microchromosomes (Chr8, Chr9, Chr10, Chr13, Chr14, Chr16, and Chr17).

2.4.1 Genomic landscape of sPLA$_2$s

sPLA$_2$s are an important component in the venom of *V. ammodytes ammodytes*[15,47]. In general, these highly diverse toxins are relatively small (13–19 kDa), Ca^{2+}-dependent, and highly stable enzymes with 5–8 disulfide bonds[48–50]. Although sPLA$_2$s are very similar in structure, they show various pharmacological effects, including myotoxicity, neurotoxicity, cardiotoxicity, and anticoagulant activity[48,51].

In terms of chromosomal location, 11 sPLA$_2$s genes were found on six chromosomes. More specifically, microchromosome 10 (Chr17) contained 5 genes; Chr 4 contained 2 genes; Chr (1a + 1b), 5, 13 and 14 (the latter two being michrochromosomes) contained one gene each (Fig. 6).

2.4.12 Genomic landscape of SVMPs

The SVMPs are a diverse group of enzymes divided into the three subclasses P-III (comprising a metalloproteinase domain, a disintegrin domain, and a cysteine-rich domain), P-II (metalloproteinase and disintegrin domain), and P-I (metalloproteinase domain only)[52]. SVMPs are known to display a wide range of toxic activities in the cardiovascular system, including hemorrhage, platelet aggregation/disruption, fibrinolysis, and prothrombin activation[19,52,53].

Interestingly, two main distinct groups of SVMPs gene tandem duplication events were found in *V. ammodytes ammodytes*. More precisely, 10 out of 20 of the SVMP genes identified in the genome of *V. ammodytes ammodytes* are found in the same orientation located on contig 132 of Chr8, nine SVMP genes in the same orientation on contig 63 and one SVMP gene (VAMM20782) in the reverse orientation are located on Chr1, whereas the remaining one gene (VAMM20790) is present on Chr14 (Fig. 7).

2.4.3 Divergence of SVMPs

The SVMP family is one of the most diverse and important toxin families in snakes and also constitutes a relatively large part of the *V. ammodytes* genome (Fig. 5). To explore the evolutionary relationships within the SVMP family, we constructed an evolutionary tree using the protein sequences of the SVMPs genes from several species including *Deinagkistrodon acutus, Protobothrops mucrosquamatus, Crotalus tigris, Bungarus multicinctus, Naja naja, Notechis*

scutatus, Pseudonaja textilis, Thermophis baileyi, Thamnophis elegans, Python bivittatus and *V. ammodytes*. 78 SVMPs genes were identified across the genomes of the 11 snakes and classified into four clades. At this point, we should mention that it is still to be confirmed in the future if all these proteins are expressed in snake venom glands or (also) elsewhere.

It has been observed that *V. ammodytes* has more SVMP gene copies than the other ten snake species in the assembly. The numbers of SVMP genes identified were as follows: *D. acutus* (four), *P. mucrosquamatus* (six), *C. tigris* (four), *B. multicinctus* (five), *N. naja* (six), *N. scutatus* (two), *P. textilis* (twelve), *T. baileyi* (three), *T. elegans* (nine), *P. bivittatus* (nine) and *V. ammodytes ammodytes* (twenty). According to the phylogenetic tree grouping results of these 11 snakes (Fig. 6). It is noteworthy that clade I contains only one gene from *P. bivittatus*; clade II comprises three genes, one each from *P. bivittatus*, *N. naja*, and *V. ammodytes ammodytes*; clade III consist of 17 genes, with five from *P. bivittatus*, eight from *V. ammodytes ammodytes*, three from *D. acutus*, and one from *N. scutatus*. The remaining 57 SVMP genes are part of clade IV. These findings highlight substantial amplification and diversity in SVMP gene counts across different snake species.

SVMPs from various snake species are grouped into distinct clades, suggesting that different SVMP genes may fulfill unique biological functions, with notable differences in branch numbers. It may also be hypothesized that these branches are associated with distinct functions. Genomic analysis of the *V. ammodytes ammodytes* SVMP family revealed that all nine SVMP genes belonging to clade II and III are clustered on Chr1 (Fig. 6), but their expression in venom glands is still to be confirmed. In contrast, ten of the remaining 11 SVMP genes, which belong to clade IV, appear capable of being expressed in venom glands as precursors of different mature toxins. This distribution suggests evolutionary and functional diversification within the *V. ammodytes ammodytes* SVMP gene family, offering valuable insights into their potential roles in the snake's defense mechanisms.

In the non-venomous colubrid snake *T. elegans*, nine SVMP genes were identified, all within clade IV. In contrast, in the elapid species *P. textilis*, 11 SVMP genes were observed, with all the genes in clade IV. These findings do not align with the evolutionary relationships among the snake species (Fig. 3), suggesting that SVMP genes may play a key role in snake evolution. Although the number of SVMP genes is relatively low in the remaining venomous snakes, their persistence throughout these species' evolutionary history, and their potential roles in development and response to environmental stress are noteworthy.

3. Discussion

V. ammodytes ammodytes is one of the most widespread and dangerous snakes within Europe, producing a venom rich in both proteolytic and neurotoxic components[54], as well as in compounds that interfere with hemostasis[14].

In this study, we utilize a high-quality, chromosome-level *de novo* genomic assembly of *Vipera ammodytes ammodytes* to address critical gaps in our understanding of the genetic architecture underlying its venom composition and its evolutionary dynamics. While previous research has extensively documented the proteomic and transcriptomic profiles of *V. ammodytes ammodytes* venom, our work provides a comprehensive genomic perspective that enables a deeper exploration into the genetic variations and evolutionary pressures shaping venom genes. This genomic approach not only enhances our understanding of venom evolution but also paves the way for the discovery of bioactive compounds and the development of more effective antivenoms. By integrating genomic data with existing proteomic and transcriptomic insights, we can begin to unravel the complex interplay between gene regulation, venom composition, and functional adaptation in this species.

3.1 Chemosensory gene regulation

Snakes have evolved an intricate chemosensory system consisting of both the main olfactory bulb and the accessory vomeronasal organs, which together facilitate the detection of a wide range of odorants[55,56]. This sophisticated system relies on the coordinated action of various GPCR families, including ORs, TAARs, V1Rs, and V2Rs, which are expressed in the olfactory epithelium and serve as principal chemosensors. In our study, by using comparative genomics and bioinformatics, we identified significant genetic contraction within *V. ammodytes ammodytes* that likely extend to true vipers more broadly. Notably, we observed that the expansion of SINEs and LTRs correlates with the diversification of ORs and V2Rs gene families, suggesting that these transposable elements have played a crucial role in shaping the chemosensory capabilities of vipers (Fig. 4). This genetic expansion likely contributes to the ecological adaptability and predator-prey interactions of *V. ammodytes ammodytes*, providing new insights into the molecular mechanisms underlying sensory evolution in snakes.

3.1.1. Vomeronasal receptor (VR) genes

Compared to snakes such as the *C. tigris* and *N. naja*, the significant contraction of the V2R genes in true vipers suggests that this contraction may correspond to their phenotype of poor vision and lack of heat-sensing capabilities. This finding aligns with the preponderant role of the vomeronasal system in snake physiology, particularly its use in processing chemical cues to relocate incapacitated prey after envenomation, a behavior known as strike-induced chemosensory searching (SICS)[57].

There is a notable divergence in V2R gene copy numbers between *V. ammodytes ammodytes*, *C. tigris*, *P. bivittatus*, *N. naja*, and *T. baileyi.* Although *C. tigris* and *V. ammodytes ammodytes* are both members of the Viperidae family, *C. tigris* possesses loreal thermoreceptors—a trait characteristic of pit vipers but absent in true vipers like *V. ammodytes*. The dramatic disparity in V2R gene copy numbers between these two species (73 in *V. ammodytes ammodytes* vs. 374 in *C. tigris*) suggests that pit vipers may highly rely on vomeronasal sensing and thermoreception to prey targets[3,58]. This hypothesis is further supported by the observation that the elapid *N. naja*, which also lacks loreal pits, has 203 V2R genes and relies on visual cues to actively hunt. Additionally, high V2R gene counts have been identified in species such as the Burmese python (*Python bivittatus*, which has labial thermoreceptors and 261 V2R genes)[59] the common garter snake (*Thamnophis sirtalis*, a colubrid)[60], and entirely aquatic sea snakes (*Hydrophis curtus* and *H. cyanocinctus*)[42], supports the idea that V2R gene expansion occurred early in the evolution of modern snakes. This pattern aligns with previous studies and suggests a possible secondary contraction of this gene family in pit vipers[57].

However, this genetic evidence contrast with observations in live animals, as *C. tigris* has been as a model for most studies on SICS in snakes to date and clearly relies extensively on the vomeronasal organ for predation similar to other pit vipers assessed thus far[57]. Therefore, it appears that expansions and contractions in gene copy numbers do not necessarily result in enhanced or reduced functionality or relevance of sensory organs at the organismal level. Further research that integrates field observations, laboratory experiments with live animals, and genomic analysis across a broader range of species and populations will be crucial in uncovering the genetic and ecological evolution of vomeronasal sensing in snakes.

3.1.2. Olfactory receptor (OR) genes

The distribution of OR genes in *V. ammodytes ammodytes*, *C. tigris*, *P. bivittatus*, *N. naja*, and *T. baileyi* differs from the typical pattern observed in OR gene copy numbers. This difference may be attributed to the fact that *V. ammodytes ammodytes* is an ambush predator with poor eyesight, unlike *N. naja*. Interestingly, the colubrid *T. baileyi*, which is visually oriented active forager like *N. naja*, also presents remarkably higher OR genes. This suggests that multiple episodes of OR gene family contraction may have occurred in colubrids, elapids, and pit vipers, possibly driven by the increasing importance of vision and thermosensation..

However, the presence of fully functional olfactory systems in *N. naja* and the high number of OR genes in *P. bivittatus*[59,61] (which has thermosensitive organs, although not homologous) compared to *V. ammodytes ammodytes* suggests caution in making broad statements about gene expansion or contraction across entire lineages. Instead, OR genes may have undergone clade-specific episodes of expansion or contraction in response to particular selection pressures, reflecting a global trend observed in vertebrates[61]. This is further supported by the extensive loss of OR genes in sea snakes, whose heavily aquatic lifestyle largely excludes them from exposure to volatile odorants, relying instead on chemical stimuli detected by the accessory olfactory system (i.e., tongue and vomeronasal organs)[42,62].

3.2. Venom genes

Although snake venom composition varies geographically and even between individuals, due to a complex interplay of environmental factors, dietary habits, and biogeographic history, the mechanisms underlying this variation at the genetic level remain largely unknown for many species[63–66]. In the present study, the genetic architecture and differential selection regimes underlying the evolution of *V. ammodytes* venom were elucidated. Among the 112 toxin genes that we identified in *V. ammodytes*, particularly SVMPs, snaclecs, SVSPs, sPLA$_2$s, and LAAOs (Fig. 4) were found to exhibit widespread duplication (Fig. 5). The SVMPs, in particular, have undergone substantial expansion within the Viperidae family and have been incorporated into the venom as a primary class of effector toxins.

Our phylogenetic reconstruction reveals the significant diversification and distinct evolutionary paths these toxins have taken (Fig. 7). The distribution and diversity of SVMP genes across different snake species highlight the complexity of venom evolution and suggests the potential functional specialization within SVMP gene families. The unique genomic distribution patterns observed in *V. ammodytes ammodytes* along with variations in SVMP gene numbers and types between venomous and non-venomous snakes underscore the need for further research into the functional roles of SVMPs in venom and their broader biological significance. Among the SVMPs genes identified, the gene with the ID VAMM20789 can be considered as a newly discovered gene. Actually, it has been confirmed at the protein level to lack the typical metalloproteinase domain, but possessing a truncated disintegrin-like/cysteine-rich domains. Its complete gene coding sequence could not be previously confirmed due to the lack of a reference genome, which prevented accurate classification. Utilizing incomplete genomic data, we have recently confirmed that this gene (tentatively named VaaMPIII-3) is unique due to the absence of the metalloproteinase domain in its coding sequence and should be classified into a new P-IIIe subclass[67]. sPLA$_2$s are another crucial toxin family of toxins in viperid snake venoms and also show an expansion in *V. ammodytes ammodytes* (Fig. 4 **and** Fig. 5).

It may be possible to expand the repertoire of effective chemical weapons to kill and capture prey by having an abundance of different toxin gene copies within a gene family[68,69]. Thus, rapid diversification of toxin genes can provide genetic resources in response to variations in prey availability for *V. ammodytes ammodytes*. Therefore, natural selection seems to be a major driving force behind toxin gene evolution to maintain, enhance, and generate a wide variety of venom components. Overall, the pattern described herein for *V. ammodytes ammodytes* is consistent with the widely accepted birth-and-death dynamic of venom gene evolution in snakes, with extensive duplication and neo- or subfunctionalization of the resulting new gene copies following recruitment in the venom apparatus[70–75].

Our chromosome-level genome assembly data for *V. ammodytes ammodytes* obtained (Fig. 1) using next-generation sequencing technologies also provides a valuable resource for exploring the evolution of snakes, as well as enabling venom-driven drug discovery.

3.3 Limitations of the study

It must be noted that only the genome and transcriptome of a single *V. ammodytes ammodytes* specimen were sequenced for this study, which precludes any inference of intraspecific variation, sexual dimorphism, and/or ontogeny in gene regulation and expression. Furthermore, the chromosome mapping described herein was not performed de novo but

instead over the existing scaffold of the *V. latastei* genome, which could result in certain mismatches and inaccurate alignments. Lastly, although we have identified members of the toxin gene family in the genome of *V. ammodytes*, we have not explored their expression through proteomic analysis for the venom of the particular individual used for this study, forcing us to rely on previous proteomic characterizations of *V. ammodytes ammodytes*. Since post-transcriptional modifications are common in the process of toxin expression, some toxin genes may not be translated into proteins and therefore do not function as toxins when secreting venom. Therefore, our toxin gene expression profile may differ from the actual mixture of *V. ammodytes ammodytes* venom. Moreover, the study used only one venom gland that had been milked for venom. Thus, the current data are not suitable for time-series gene expression analysis, and more replicate experiments are needed for future research of this kind.

Nevertheless, we believe that our *V. ammodytes ammodytes* genome will provide a valuable resource also for comparative studies on the evolution of viperid snakes and snakes in general. Moreover, we are confident that the comprehensive characterization of *V. ammodytes ammodytes* based on a sequencing approach can play an essential role in promoting antivenom development and contributing to the establishment and enrichment of databases used for the high-throughput discovery of bioactive molecules. Ultimately, the new insights revealed by the *V. ammodytes ammodytes* genome highlight how further decoding other venomous snake genomes will be critical to understanding the evolution of highly adaptive functional traits such as the venom apparatus, the olfactory system, and thermosensitive organs.

4. Materials and Methods

4.1 Data and code availability

The raw sequencing data and assembled genomes that support the findings of this study have been deposited into CNGB Sequence Archive (CNSA)[76] of China National GenBank Database (CNGBdb)[77], https://db.cngb.org/cnsa/ with accession number CNP0002260 and NCBI in BioProject ID is PRJNA1177245. This paper does not report the original code of the bioinformatic analysis.

4.2 Sample collection

After obtaining permission from the Slovenian Ministry of Environment and Spatial Planning, an adult male specimen of *V. ammodytes* was captured in the wild in the northwestern region of Slovenia, near the town of Jesenice. Two days after milking, the snake was dissected, some of its organs (liver, heart, venom gland, and skeletal muscles) immediately frozen in liquid nitrogen and stored at − 80 ℃ until use.

4.3 DNA extraction and whole-genome sequencing.

High molecular mass (> 100 Kb) genomic DNA of *V. ammodytes* was isolated from frozen liver tissue using MegaLong™ purification kit (G-Biosciences, Saint Louis, USA) according to a modified procedure. In short, the deeply frozen liver was carefully ground in a mortar in the presence of liquid nitrogen and cell nuclei isolated using the manufacturer's protocol. After incubation of the liver cell nuclei with proteinase K, instead of using the provided dialysis cups, DNA sample was dialyzed overnight using Spectra/Por 2 dialysis tubing (with a molecular weight cut-off of 12–14 kDa; Spectrum Laboratories, San Francisco, USA) against 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA with three replacements of the buffer. Isolated genomic DNA was finally stored at 4 °C. The purified DNA concentration was measured using NanoDrop (Thermo Fisher Scientific) and Qubit 2.0 (Invitrogen) instruments.

Whole genome sequencing for de novo assembly was performed using 18-Cell single molecule, real-time (SMRT) sequencing and Sequencing Chemistry 2.0 kit (PacBio Sequel System, RRID: SCR 017989) by BGI-Wuhan Sequencing Center, generating 19.7 million reads with a total length of 199 Gb. DNA was also sequenced using paired end 150 bp (PE150) reads utilizing an Illumina HiSeq X-TEN (Illumina HiSeq X Ten, RRID: SCR 016385) at the BGI-Wuhan Sequencing

Center. Paired-end libraries with insert sizes of 410 and 670 bp and mate-pair libraries with insert sizes of 2, 5, 10, and 20 Kb were constructed, following a standard protocol provided by Illumina, achieving ~131× coverage.

4.4 RNA isolation and transcriptome sequencing

Total RNAs from the VG (venom gland), LH (liver and heart mix), and MB (muscle and bone mix) tissues of *V. ammodytes* were extracted using TRIzol reagent (Invitrogen). With an Agilent 4200 Bioanalyzer (Agilent) and RNA 6000 Nano LabChip Kit (Agilent), total RNA contents were determined using electrophoresis and quantification. Then, using the MGIEasy RNA Directional Library Prep Kit (MGI Tech), 1 μg of polyA-enriched RNA was used to construct the sequencing library for MGIseq-2000. Finally, paired-end sequencing was performed on the MGISEQ-2000 sequencing platform, following the manufacturer's instructions, yielding PE150 reads.

4.5 Genome assembly and annotation

Genome assembly. The draft genomes of *V. ammodytes ammodytes* were assembled by Nextdenovo (v2.5.0)[25] software with PacBio sequencing data. In the genome polishing stage, NextPolish (v1.4.0)[26] was mainly used to improve the accuracy of single bases by two rounds of short reads. Redundancy removal of genomes was performed by Purge_dups (v1.2.5)[27], and genome assembly improvement, resulted in a de novo assembled genome at the scaffold level. The de novo assembled genome was provided as input, with other parameters left at their default setting. Genome assembly completeness was assessed by using BUSCO[78] (v5.2.2) with vertebrata_odb10 data set. The genome assembly has been submitted to NCBI under the BioProject PRNJA1177245.

Repeat Regions Prediction. Transposable elements (TEs) and other repetitive elements were identified using a combination of homology-based and de novo approaches. For the homology-based approach at both the DNA and protein levels, the genome assembly was aligned to the known repeat database REPBASE (v21.01) using RepeatMasker[79] (v4.0.5), RepeatProteinMask and Tandem Repeats Finder (TRF)[80] (v4.07b). For the de novo-based approach, RepeatModeler[35] (v2.0) and LTR_retriever were used to construct a de novo repeat library. We found that the *V. ammodytes* genome contained 53.75% TEs. Long Interspersed Nuclear Elements (LINEs) accounted for most TEs, occupying about ~ 42% of the genome. All repetitive elements were masked for gene annotation.

Annotation of protein-encoding genes. We combined homology-based, de novo, and transcriptome-based methods to predict assembled gene content. In a homology-based approach, GeneWise[81] (v2.4.1) was used to map 4 closely related or high-quality protein sequences, including *Protobothrops mucrosquamatus, Deinagkistrodon acutus, Gloydius shedaoensis*, and *Crotalus tigris*, available in the NCBI database, to the assembled genome with an E-value cutoff of 1e-5. In the de novo method, we ran the repeat-masked genome using Augustus[82] (v3.0.3). In the transcriptome-based method, transcripts were assembled using StringTie[83] (v1.3.3b) based on clean RNA-seq data. The final protein-coding gene set was generated using the MAKER (v3.01.03)[84] pipeline by combining high-quality homology-based, de novo, and RNA-seq supported genes. Based on the above methods, 20,775 genes were annotated in the genome.

Annotation of gene function. Functional annotations of protein-encoding genes were carried out using BLAST (e-value cut-off of 1e-5) against publicly available databases, including the Swiss-Prot[37], TrEMBL[37], Gene ontology (GO)[85] terms, and the KEGG[36] database. InterProScan (v5.52-86.0)[86] was used to predict domains and motifs. 96.21% of the genes in the gene sets of this species were fully annotated in the five databases as mentioned above. In addition, noncoding RNA (ncRNA) genes, including miRNA, tRNA, snRNA, and rRNA, were predicted in the assembled genome. tRNA genes were identified using tRNAscan-SE[87] (v1.3.1). snRNA and miRNA genes were detected by searching the reference genome sequences against the content of the Rfam database (Release 12.0) using BLAST.

4.6 Phylogenetic analyses

Homologous genes of 11 species (*Python bivittatus*, *Vipera ammodytes ammodytes*, *Deinagkistrodon acutus*, *Protobothrops mucrosquamatus*, *Crotalus tigris*, *Thermophis baileyi*, *Thamnophis elegans*, *Bungarus multicinctus*, *Naja naja*, *Notechis scutatus*, and *Pseudonaja textilis*) were determined by performing all-to-all BLASTP analysis by using amino acid sequences with the parameter of "e-value 1e-5". We identified 1,084 single-copy genes shared by these 11 species for constructing the maximum-likelihood phylogenetic tree by IQTREE[88] (v1.6.12). For gene family expansion analysis, we identified 12,200 gene families in the above-mentioned 11 genomes and used CAFE[39] (v4.2.1) to detect gene family expansion and contraction. KEGG and GO enrichment analysis on expanded gene families were performed by R script with the cut-off adjusted p value < 0.05.

For the construction of phylogenetic trees, assembly of ten other species (*B. multicinctus, C. tigris, D. acutus, N. naja, N. scutatus, P. mucrosquamatus, P. textilis, P. bivittatus, T. elegans*, and *T. baileyi*) downloaded from a previously published paper (Table S12, along with the genome of *V. ammodytes*, were used to identify single copy orthologous genes obtained from the TreeFam[89] pipeline. Then, alignment was performed with default parameters by using software MUSCLE. A super-gene was assigned to each species based on the gene alignment results. In the third step, both maximum likelihood with PhyML[90] program and Bayesian methods were used with a GTR + gamma model with MrBayes[91] program to construct phylogenetic trees, which displayed identical topology structures. Finally, divergence times between species were estimated using the MCMCTREE program in the PAML[92] (v4.7) package and using fossil time constraints from TimeTree database (http://www.timetree.org) to calibrate a Bayesian relaxed-molecular clock model. *B. multicinctus – C. tigris* (~ 35.3–76.9 million years ago), *B. multicinctus – N. naja* (~ 19.8–36.2 million years ago), *B. multicinctus – N. scutatus* (~ 18.4–34.6 million years ago), *B. multicinctus – P. bivittatus* (~ 72.8–93.5 million years ago), *B. multicinctus – T. elegans* (~ 26.9–88.9 million years ago), *C. tigris – D. acutus* (~ 18.4–64.0 million years ago), *C. tigris – P. mucrosquamatus* (~ 11.7–26.5 million years ago) and *C. tigris – V. ammodytes* (~ 33.6–45.5 million years ago).

4.7 Gene family analysis

Genes were considered species-specific when there were more than two genes within an orthogroup defined by OrthoFinder, and all of them originated from one species[93]. The specific genes of *V. ammodytes* were identified by a hypergeometric analysis involving GO and KEGG annotations as genes that were over-represented (adjusted p < 0.05). The gene family expansion analysis was performed using CAFE[39] (v4.0) with p-values less than 0.05 when FDR correction was used, as well as the phylogenetic tree and divergence time analysis for the eleven species. On each branch of the tree, expanded and contracted gene families were detected by comparing the cluster size of each branch to the maximum-likelihood cluster size of its ancestral node. Gene family expansion is indicated by a smaller ancestral node, while a larger ancestral node indicates family contraction. In addition to calculating the overall p-value (p- value of family-wide in CAFE[39] (v4.0) based on Monte Carlo resampling) of each branch and node, exact p-values (Viterbi method in CAFE[39] (v4.0)) of each significant overall p-value (< 0.01) were also calculated. A significant expansion or contraction of a gene family was defined as a p-value of less than 0.01.

4.8 Accelerated evolution and positive selection analysis

From the OrthoFinder[93] (v2.2.7) results described above, single copy gene families were selected for identifying positively selected genes and genes with high rates of molecular evolution. We first identified the quickly evolving genes using CODEML's branch model in PAML[92] (v4.7), with *V. ammodytes ammodytes* gene set as the foreground branch and the other species as background branches. Assuming the null hypothesis, the u value of each branch was equal, and assuming the alternative hypothesis, the foreground branch did not have the same u value as the background branch. Following, the Bonferroni correction was applied to the orthologous groups, and the likelihood ratio test (LRT) was performed. Genes were considered with quick evolution if their u value on the foreground branch was more significant than their background branches and when the corrected p-value was less than 0.05.

Using CODEML's branch-site model in PAML[92] (v4.7) to test for potentially positively selected genes, *V. ammodytes ammodytes* was used as foreground branches and other species as background branches. It was hypothesized that the ω values for each site on each branch were ≤ 1, whereas it was alternatively hypothesized that the ω values for *V. ammodytes ammodytes* on the foreground branch were > 1. Next, a likelihood ratio test was conducted: the null distribution was a mixture of c2 distributions with 1 degree of freedom and a point mass at zero. The p-values calculated based on this mixture distribution were further corrected for multiple testing through an FDR test with Bonferroni correction. As a result of the Bayes empirical Bayes inference, the positively selected genes should meet the requirements of a corrected p-value < 0.05 and contain at least one positively selected site with a posterior probability (> 0.95).

4.9 Evolutionary analysis of SVMPs family

Based on 11 snake genomes listed in Table S12, a phylogenetic analysis of SVMPs was conducted as one of the most abundant and diverse toxins in snakes. To investigate the genetic diversity and evolutionary patterns of the SVMPs, we identified 78 SVMPs genes from the 11 snakes. Using the Bayesian information criterion, phylogenetic trees were constructed from these alignments using the maximum likelihood method with JTT + I + G4 as a best-fit protein substitution model in IqTREE (v1.6.6) with the SVMPs full proteins. All phylogenetic trees were tested with 1,000 bootstrapping replicates. Genome pairwise similarity sequences were calculated using a maximum composite likelihood model.

Declarations

Author contribution

Conceptualization: AHL, WR, LS, ERT

Methodology: ER, MS

Investigation: WR, LS, ERT

Visualisation: WR, ERT

Funding acquisition: SL, AHL

Project administration: ERT, AHL, JP, SL

Resources: AHL, SL, JP

Supervision: ERT, AHL, JP, SL

Writing – original draft: WR, ERT, JP, AHL

Writing – review & editing: all authors

Funding

A.H.L. acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant number 850974). JP acknowledges support from the Slovenian Research Agency (grant number P1-0207).

Conflict of interest

The authors declare no conflict of interest.

References

1. Wallach, V., Williams, K., Boundy, J.: *Snakes of the World. A Catalogue of Living and Extinct Species. Snakes of the World: A Catalogue of Living and Extinct Species* (2014). 10.1201/b16901
2. JACKSON, K.: The evolution of venom-delivery systems in snakes. Zool. J. Linn. Soc. **137**, 337–354 (2003)
3. DE COCK BUNING, T.: Thermal Sensitivity as a Specialization for Prey Capture and Feeding in Snakes1. Am. Zool. **23**, 363–375 (1983)
4. Jackson, T.N.W., Jouanne, H., Vidal, N.: Snake Venom in Context: Neglected Clades and Concepts. Front. Ecol. Evol. **7**, (2019)
5. Barlow, A., Pook, C.E., Harrison, R.A., Wüster, W.: Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. *Proceedings of the Royal Society B: Biological Sciences* 276, 2443–2449 (2009)
6. Lyons, K., Dugon, M.M., Healy, K.: Diet Breadth Mediates the Prey Specificity of Venom Potency in Snakes. Toxins. **12**, 74 (2020)
7. Peng, C., et al.: Large-scale snake genome analyses provide insights into vertebrate development. Cell. **186**, 2959–2976e22 (2023)
8. Roberts, J.R., et al.: Whole snake genomes from eighteen families of snakes (Serpentes: Caenophidia) and their applications to systematics. J. Hered. **115**, 487–497 (2024)
9. Holding, M.L., Biardi, J.E., Gibbs, H.L.: Coevolution of venom function and venom resistance in a rattlesnake predator and its squirrel prey. *Proceedings of the Royal Society B: Biological Sciences* 283, 20152841 (2016)
10. Rao, W., et al.: The rise of genomics in snake venom research: recent advances and future perspectives. *GigaScience* 11, giac024 (2022)
11. Schield, D.R., et al.: The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. Genome Res. **29**, 590–601 (2019)
12. Suryamohan, K., et al.: The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. Nat. Genet. **52**, 106–117 (2020)
13. Margres, M.J., et al.: The Tiger Rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. PNAS **118**, (2021)
14. Leonardi, A., Sajevic, T., Pungerčar, J., Križaj, I.: Comprehensive Study of the Proteome and Transcriptome of the Venom of the Most Venomous European Viper: Discovery of a New Subclass of Ancestral Snake Venom Metalloproteinase Precursor-Derived Proteins. J. Proteome Res. **18**, 2287–2309 (2019)
15. Georgieva, D., et al.: Comparative Analysis of the Venom Proteomes of Vipera ammodytes ammodytes and Vipera ammodytes meridionalis. J. Proteome Res. **7**, 866–886 (2008)
16. Ursenbacher, S., et al.: Molecular phylogeography of the nose-horned viper (Vipera ammodytes, Linnaeus (1758)): Evidence for high genetic diversity and multiple refugia in the Balkan peninsula. Mol. Phylogenet. Evol. **46**, 1116–1128 (2008)
17. Ritonja, A., Gubenek, F., Ammodytoxin, A.: a highly lethal phospholipase A2 from Vipera ammodytes ammodytes venom. Biochim. et Biophys. Acta (BBA) - Protein Struct. Mol. Enzymol. **828**, 306–312 (1985)
18. Hempel, B.-F., et al.: Comparative Venomics of the Vipera ammodytes transcaucasiana and Vipera ammodytes montandoni from Turkey Provides Insights into Kinship. Toxins (Basel) **10**, (2018)
19. Leonardi, A., et al.: Structural and biochemical characterisation of VaF1, a P-IIIa fibrinogenolytic metalloproteinase from Vipera ammodytes ammodytes venom. Biochimie **109**, (2014)
20. Casewell, N.R., Wüster, W., Vonk, F.J., Harrison, R.A., Fry, B.G.: Complex cocktails: the evolutionary novelty of venoms. Trends Ecol. Evol. **28**, 219–229 (2013)

21. Daltry, J.C., Wüster, W., Thorpe, R.S.: Diet and snake venom evolution. Nature. **379**, 537 (1996)
22. Chin, C.-S., et al.: Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods. **13**, 1050–1054 (2016)
23. ID 858072 - BioProject - NCBI: https://www.ncbi.nlm.nih.gov/bioproject/858072
24. Vipera berus berus: isolate:VBER.BE-female (ID 170536) - BioProject - NCBI. https://www.ncbi.nlm.nih.gov/bioproject/170536
25. Hu, J., et al.: An efficient error correction and accurate assembly tool for noisy long reads. 03.09.531669 Preprint at (2023). https://doi.org/10.1101/2023.03.09.531669 (2023)
26. Hu, J., Fan, J., Sun, Z., Liu, S.: NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics. **36**, 2253–2255 (2020)
27. Guan, D., et al.: Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. **36**, 2896–2898 (2020)
28. Alonge, M., et al.: RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. **20**, 224 (2019)
29. Saint Girons, H.: Caryotypes et évolution des Vipères Européennes (Reptilia. Viperidae). Bull. Soc. zool. Fr. **102**, 39–49 (1977)
30. Xu, J., et al.: Genomic, transcriptomic, and epigenomic analysis of a medicinal snake, *Bungarus multicinctus*, to provides insights into the origin of Elapidae neurotoxins. Acta Pharm. Sinica B. **13**, 2234–2249 (2023)
31. Westeen, E.P., et al.: A genome assembly for the southern Pacific rattlesnake, Crotalus oreganus helleri, in the western rattlesnake species complex. J. Hered. **114**, 681–689 (2023)
32. Westeen, E.P., et al.: A reference genome assembly for the continentally distributed ring-necked snake, *Diadophis punctatus*. J. Hered. **114**, 690–697 (2023)
33. Schöneberg, Y., et al.: Genomics reveals broad hybridization in deeply divergent Palearctic grass and water snakes (*Natrix* spp). Mol. Phylogenet. Evol. **184**, 107787 (2023)
34. Peng, C., et al.: Large-scale snake genome analyses provide insights into vertebrate development. Cell. **186**, 2959–2976e22 (2023)
35. Flynn, J.M., et al.: RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* 117, 9451–9457 (2020)
36. Kanehisa, M., Goto, S.K.E.G.G.: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. **28**, 27–30 (2000)
37. Boeckmann, B., et al.: The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. **31**, 365–370 (2003)
38. Apweiler, R., et al.: The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res. **29**, 37–40 (2001)
39. De Bie, T., Cristianini, N., Demuth, J.P., Hahn, M.: W. CAFE: a computational tool for the study of gene family evolution. Bioinformatics. **22**, 1269–1271 (2006)
40. Hedges, S.B., Dudley, J., Kumar, S.: TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics. **22**, 2971–2972 (2006)
41. Spehr, M., Munger, S.D.: Olfactory receptors: G protein-coupled receptors and beyond. J. Neurochem. **109**, 1570–1583 (2009)
42. Li, A., et al.: Two Reference-Quality Sea Snake Genomes Reveal Their Divergent Evolution of Adaptive Traits and Venom Systems. Mol. Biol. Evol. **38**, 4867–4883 (2021)
43. Zhang, Z.-Y., et al.: The structural and functional divergence of a neglected three-finger toxin subfamily in lethal elapids. Cell. Rep. **40**, 111079 (2022)

44. Peier, A.M., et al.: A TRP channel that senses cold stimuli and menthol. Cell. **108**, 705–715 (2002)
45. Cordero-Morales, J.F., Gracheva, E.O., Julius, D.: Cytoplasmic ankyrin repeats of transient receptor potential A1 (TRPA1) dictate sensitivity to thermal and chemical stimuli. Proc. Natl. Acad. Sci. U S A. **108**, E1184–E1191 (2011)
46. Gracheva, E.O., et al.: Molecular Basis of Infrared Detection by Snakes. Nature. **464**, 1006–1011 (2010)
47. Leonardi, A., Sajevic, T., Pungerčar, J., Križaj, I.: Comprehensive Study of the Proteome and Transcriptome of the Venom of the Most Venomous European Viper: Discovery of a New Subclass of Ancestral Snake Venom Metalloproteinase Precursor-Derived Proteins. J. Proteome Res. **18**, 2287–2309 (2019)
48. Manjunatha Kini, R.: Excitement ahead: structure, function and mechanism of snake venom phospholipase A2 enzymes. Toxicon. **42**, 827–840 (2003)
49. Lomonte, B., Gutiérrez, J.M.: Phospholipases A2 from viperidae snake venoms: how do they induce skeletal muscle damage? Acta Chim. Slov. **58**, 647–659 (2011)
50. Brglez, V., Pucer, A., Pungerčar, J., Lambeau, G., Petan, T.: Secreted phospholipases A2 are differentially expressed and epigenetically silenced in human breast cancer cells. Biochem. Biophys. Res. Commun. **445**, 230–235 (2014)
51. Montecucco, C., Gutiérrez, J.M., Lomonte, B.: Cellular pathology induced by snake venom phospholipase A2 myotoxins and neurotoxins: common aspects of their mechanisms of action. Cell. Mol. Life Sci. **65**, 2897–2912 (2008)
52. Fox, J.W., Serrano, S.M.T.: Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases. Toxicon. **45**, 969–985 (2005)
53. Fox, J.W., Serrano, S.M.T.: Insights into and speculations about snake venom metalloproteinase (SVMP) synthesis, folding and disulfide bond formation and their contribution to venom complexity. FEBS J. **275**, 3016–3030 (2008)
54. Di Nicola, M.R., et al.: Vipers of Major clinical relevance in Europe: Taxonomy, venom composition, toxicology and clinical management of human bites. Toxicology. **453**, 152724 (2021)
55. Niimura, Y., Nei, M.: Extensive Gains and Losses of Olfactory Receptor Genes in Mammalian Evolution. PLOS ONE. **2**, e708 (2007)
56. Griffin, C.A., Kafadar, K.A., Pavlath, G.K.: MOR23 promotes muscle regeneration and regulates cell adhesion and migration. Dev. Cell. **17**, 649–661 (2009)
57. Teshera, M.S., Clark, R.W.: Strike-Induced Chemosensory Searching in Reptiles: A Review. Herpetological Monogr. **35**, 28–52 (2021)
58. Kardong, K.V., Mackessy, S.P.: The Strike Behavior of a Congenitally Blind Rattlesnake. J. Herpetology. **25**, 208–211 (1991)
59. Castoe, T.A., et al.: The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proceedings of the National Academy of Sciences* 110, 20645–20650 (2013)
60. Perry, B.W., et al.: Molecular Adaptations for Sensing and Securing Prey and Insight into Amniote Genome Diversity from the Garter Snake Genome. Genome Biol. Evol. **10**, 2110–2129 (2018)
61. Vandewege, M.W., et al.: Contrasting Patterns of Evolutionary Diversification in the Olfactory Repertoires of Reptile and Bird Genomes. Genome Biol. Evol. **8**, 470–480 (2016)
62. Kishida, T., et al.: Loss of olfaction in sea snakes provides new perspectives on the aquatic adaptation of amniotes. Proc. Biol. Sci. **286**, 20191828 (2019)
63. Silva-Júnior, L.N., da, et al.: Geographic variation of individual venom profile of *Crotalus durissus* snakes. J. Venom. Anim. Toxins incl. Trop. Dis. **26**, (2020)
64. Kalita, B., Mackessy, S.P., Mukherjee, A.K.: Proteomic analysis reveals geographic variation in venom composition of Russell's Viper in the Indian subcontinent: implications for clinical manifestations post-envenomation and antivenom treatment. Expert Rev. Proteomics. **15**, 837–849 (2018)

65. Kalita, B., Mukherjee, A.K.: Recent advances in snake venom proteomics research in India: a new horizon to decipher the geographical variation in venom proteome composition and exploration of candidate drug prototypes. J. Proteins Proteom. **10**, 149–164 (2019)
66. Núñez, V., et al.: Snake venomics and antivenomics of Bothrops atrox venoms from Colombia and the Amazon regions of Brazil, Perú and Ecuador suggest the occurrence of geographic variation of venom phenotype by a trend towards paedomorphism. J. Proteom. **73**, 57–78 (2009)
67. Požek, K., et al.: Genomic Confirmation of the P-IIIe Subclass of Snake Venom Metalloproteinases and Characterisation of Its First Member, a Disintegrin-Like/Cysteine-Rich Protein. Toxins. **14**, 232 (2022)
68. Schield, D.R., et al.: The roles of balancing selection and recombination in the evolution of rattlesnake venom. Nat. Ecol. Evol. **6**, 1367–1380 (2022)
69. Jackson, T.N.W., Koludarov, I.: How the Toxin got its Toxicity. Front. Pharmacol. **11**, (2020)
70. Fry, B.G.: 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 Res. **15**, 403–420 (2005)
71. Casewell, N.R., Jackson, T.N.W., Laustsen, A.H., Sunagar, K.: Causes and Consequences of Snake Venom Variation. Trends Pharmacol. Sci. **41**, 570–581 (2020)
72. Lynch, V.J.: Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. BMC Evol. Biol. **14** (2007)
73. Fry, B.G., et al.: The Toxicogenomic Multiverse: Convergent Recruitment of Proteins Into Animal Venoms. Annu. Rev. Genom. Hum. Genet. **10**, 483–511 (2009)
74. Giorgianni, M.W., et al.: The origin and diversification of a novel protein family in venomous snakes. Proc. Natl. Acad. Sci. USA. **117**, 10911–10920 (2020)
75. Jackson, T.N.W., et al.: Rapid Radiations and the Race to Redundancy: An Investigation of the Evolution of Australian Elapid Snake Venoms. Toxins (Basel). **8**, 309 (2016)
76. Guo, X., et al.: CNSA: a data repository for archiving omics data. *Database (Oxford)* baaa055 (2020). (2020)
77. Chen Fengzhen, Y.L., CNGBdb: China National GeneBank DataBase. Hereditas(Beijing). **42**, 799–809 (2020)
78. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M.: BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol. Biol. Evol. **38**, 4647–4654 (2021)
79. Chen, N.: Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Current Protocols in Bioinformatics* 5, 4.10.1–4.10.14 (2004)
80. Ou, S., Jiang, N., LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons1[OPEN]. Plant. Physiol. **176**, 1410–1422 (2018)
81. Birney, E., Clamp, M., Durbin, R.: GeneWise and Genomewise. Genome Res. **14**, 988–995 (2004)
82. Stanke, M., Steinkamp, R., Waack, S., Morgenstern, B.: AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. **32**, W309 (2004)
83. Pertea, M., et al.: StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. **33**, 290–295 (2015)
84. Campbell, M.S., Holt, C., Moore, B., Yandell, M.: Genome Annotation and Curation Using MAKER and MAKER-P. *Curr Protoc Bioinformatics* 48, 4.11.1–4.11.39 (2014)
85. Ye, J., et al.: WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. **34**, W293–W297 (2006)
86. Jones, P., et al.: InterProScan 5: genome-scale protein function classification. Bioinformatics. **30**, 1236–1240 (2014)
87. Lowe, T.M., Eddy, S.R.: tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**, 955–964 (1997)

88. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q.: IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol. Biol. Evol. **32**, 268–274 (2015)
89. Li, H., et al.: TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Res. **34**, D572–D580 (2006)
90. Guindon, S., Delsuc, F., Dufayard, J.-F., Gascuel, O.: Estimating Maximum Likelihood Phylogenies with PhyML. In: Posada, D. (ed.) Bioinformatics for DNA Sequence Analysis, pp. 113–137. Humana, Totowa, NJ (2009). 10.1007/978-1-59745-251-9_6
91. Yang, Z., Rannala, B.: Bayesian Estimation of Species Divergence Times Under a Molecular Clock Using Multiple Fossil Calibrations with Soft Bounds. Mol. Biol. Evol. **23**, 212–226 (2006)
92. Yang, Z.: PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. **24**, 1586–1591 (2007)
93. Emms, D.M., Kelly, S.: OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. **16**, 157 (2015)
94. Das, I., et al.: Taxonomic revision of the king cobra Ophiophagus hannah (Cantor, 1836) species complex (Reptilia: Serpentes: Elapidae), with the description of two new species. Eur. J. Taxonomy. **961**, 1–51 (2024)

Figures

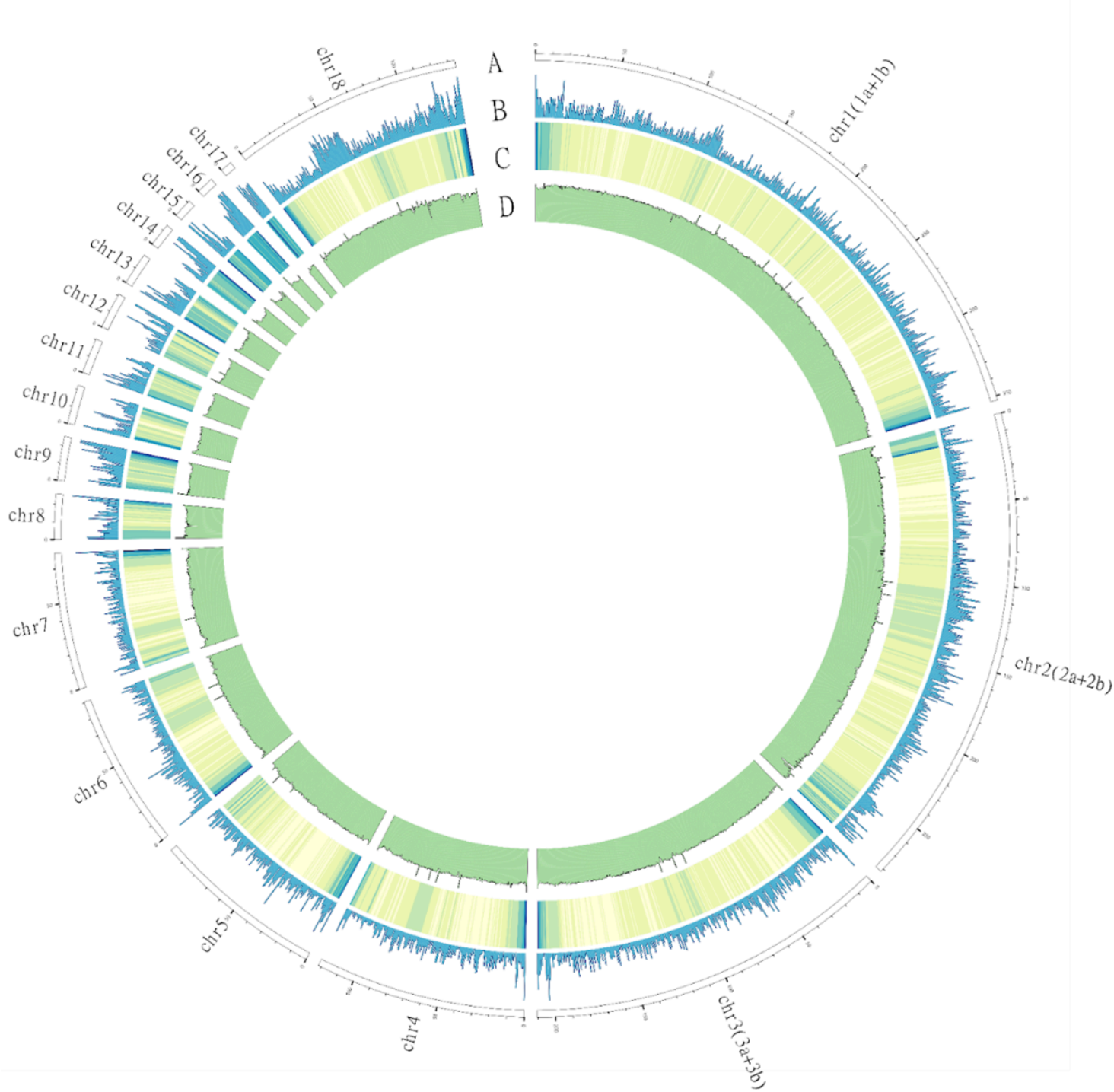

Figure 1

Chromosomal assembly of *V ammodytes ammodytes* genome using *C. viridis* chromosomal assembly as reference. *Circos plot of V. ammodytes genome assembly showing sizes (each single scale represents 15 Mb) of 18 pseudo-chromosomes, corresponding to a complete set of the presumed 21 chromosomes. Tracks (outer to inner circles) indicate the following: A, Location of toxin genes; B, Gene density at 500 Kb bins (gene numbers per Mb; minimum 0, maximum 87; darker color indicates more genes); C, DNA transposon ratios; D, Retrotransposon ratios; E, GC content (window size of 100 Kb bins). The outer circle represents the chromosome length of V. ammodytes, with units in MB. The*

assembly based on C. viridis distribution is composed by 17 autosomes (7 macro-, 11 micro-) and a sexual chromosome (Chr Z).

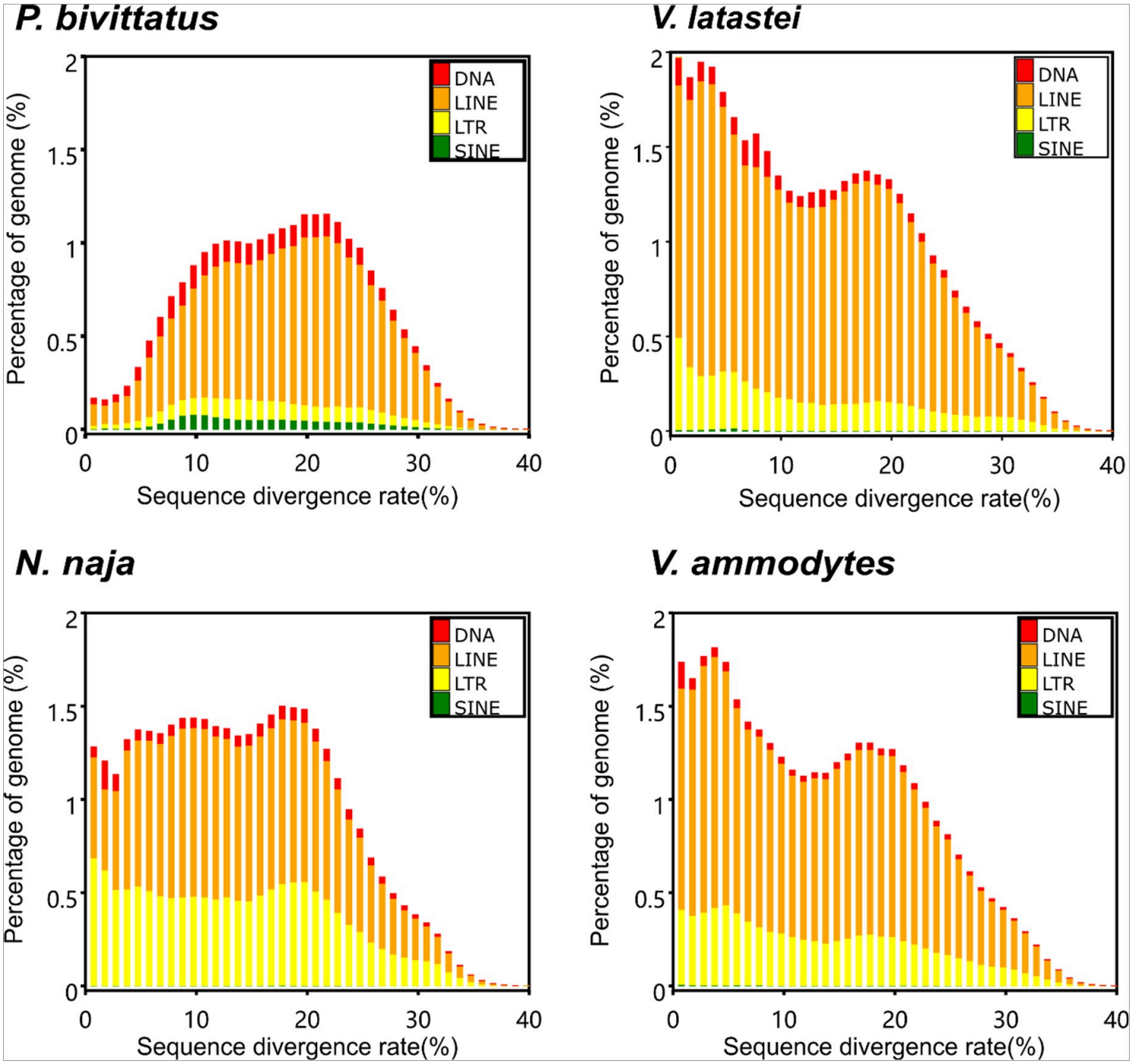

Figure 2

Distribution of divergence rate of each type of transposable elements (TEs) in the genome of four snakes (*P. bivittatus*, *N. naja*, *V. latastei* and *V. ammodytes ammodytes*) based on same analysis. *Y-axis represents genome coverage for each type of TEs (i.e., DNA transposons, SINE, LINE and LTR retrotransposons), and X-axis represent sequence divergence rate. The divergence rate was calculated between the identified TEs in the genome using a homology-based method and the consensus sequence in the Repbase database.*

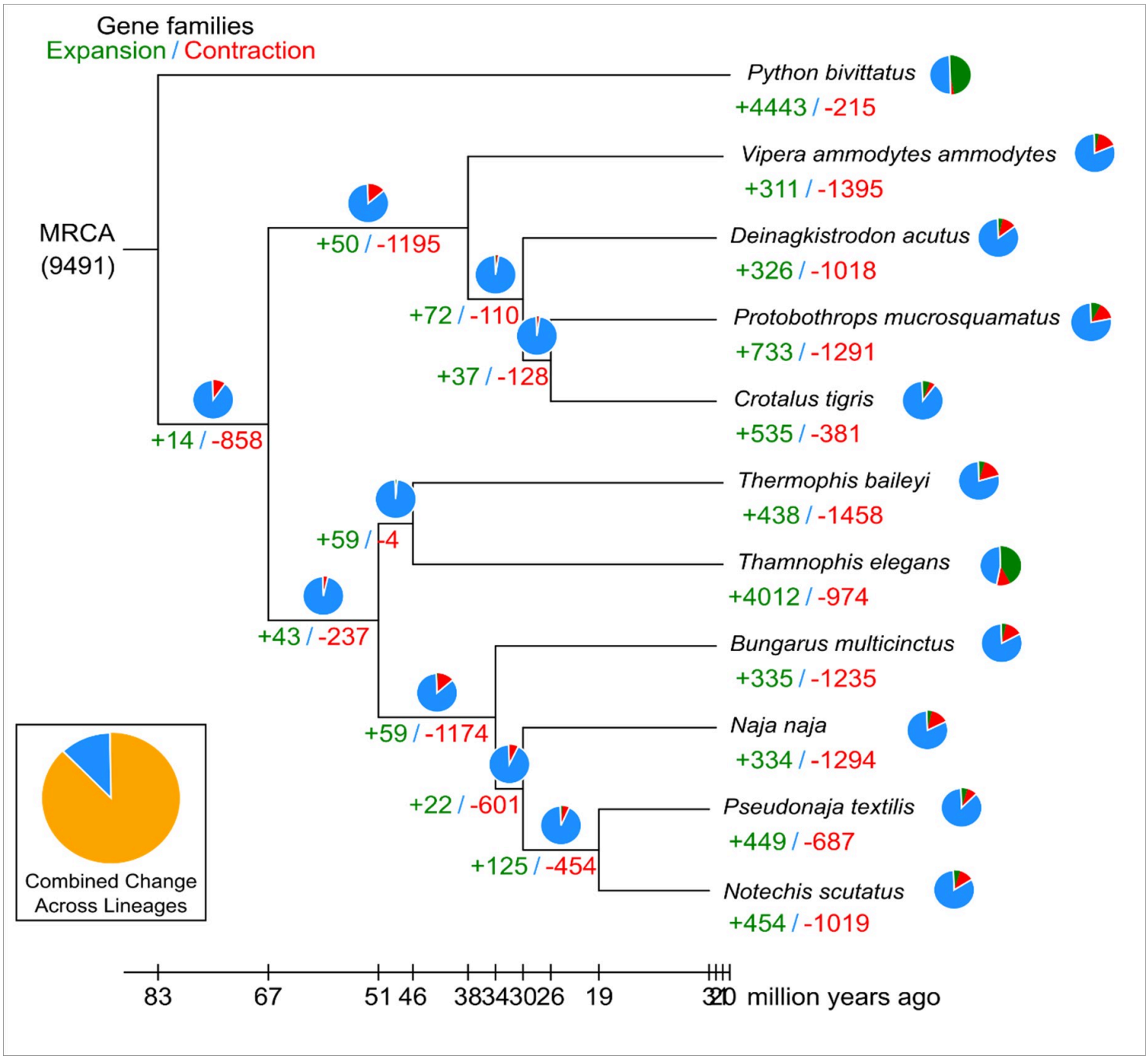

Figure 3

Phylogenetic tree with timescale indicating divergence times. *Green indicates the numbers of significantly expanded gene families (+) and brown represents the contracted gene families' numbers (-). Pie charts indicate proportions of expanded (green), contracted (reddish brown), and other (light red) gene families of each node. Timescale bar at the bottom shows the divergence time.*

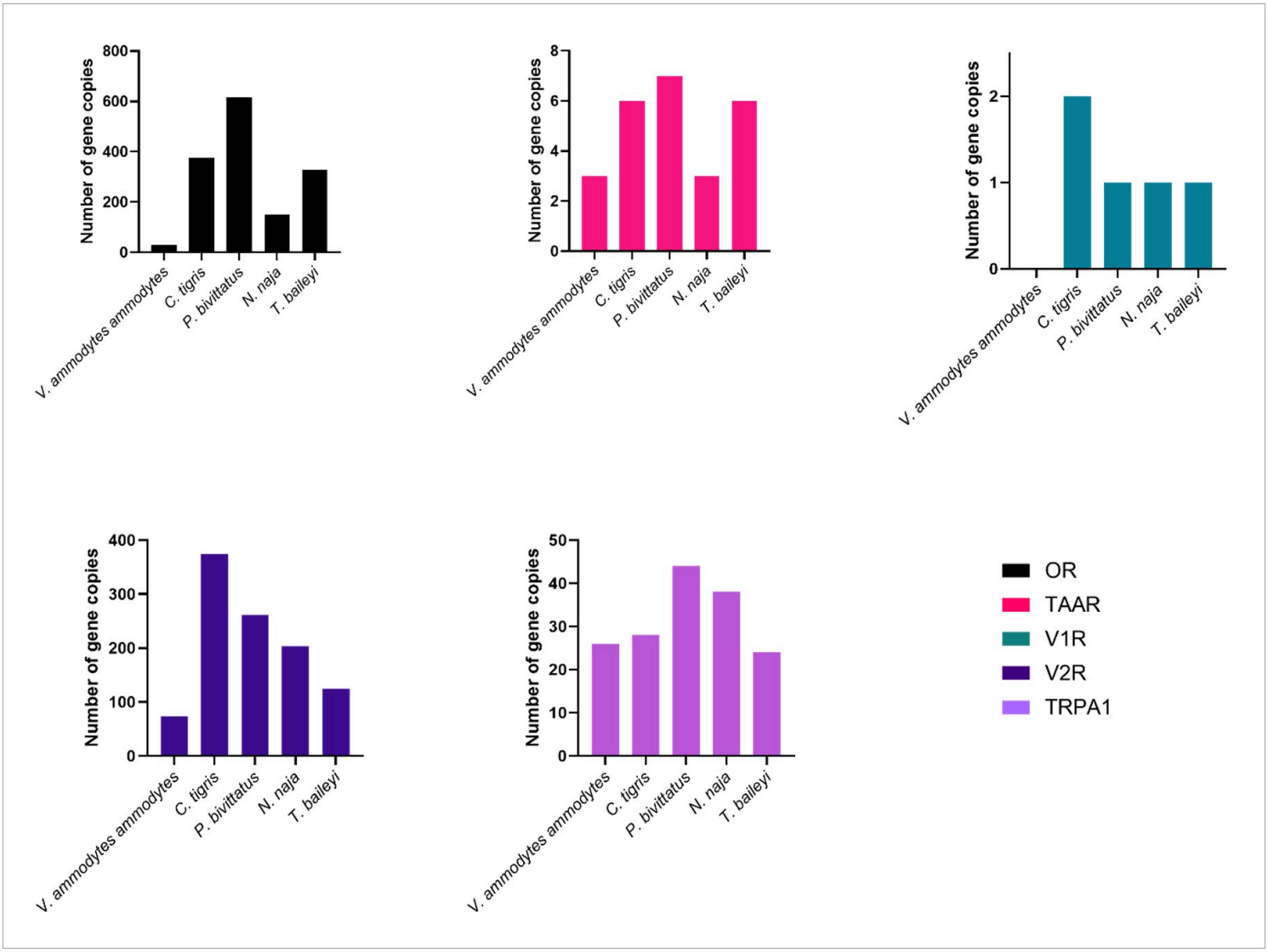

Figure 4

Number of gene copies for olfactory receptors and thermal sensing channels for V. ammodytes amodytes, C. tigris, P. bivittatus, and T. baileyi. Olfatory receptors include include olfactory receptors (ORs) trace amine-associated receptors (TAARs), vomeronasal type-1 receptors (V1Rs), and vomeronasal type-2 receptors (V2Rs) TRP channel genes involved in thermal sensing include transient receptor potential ankyrin 1 (TRPA1).

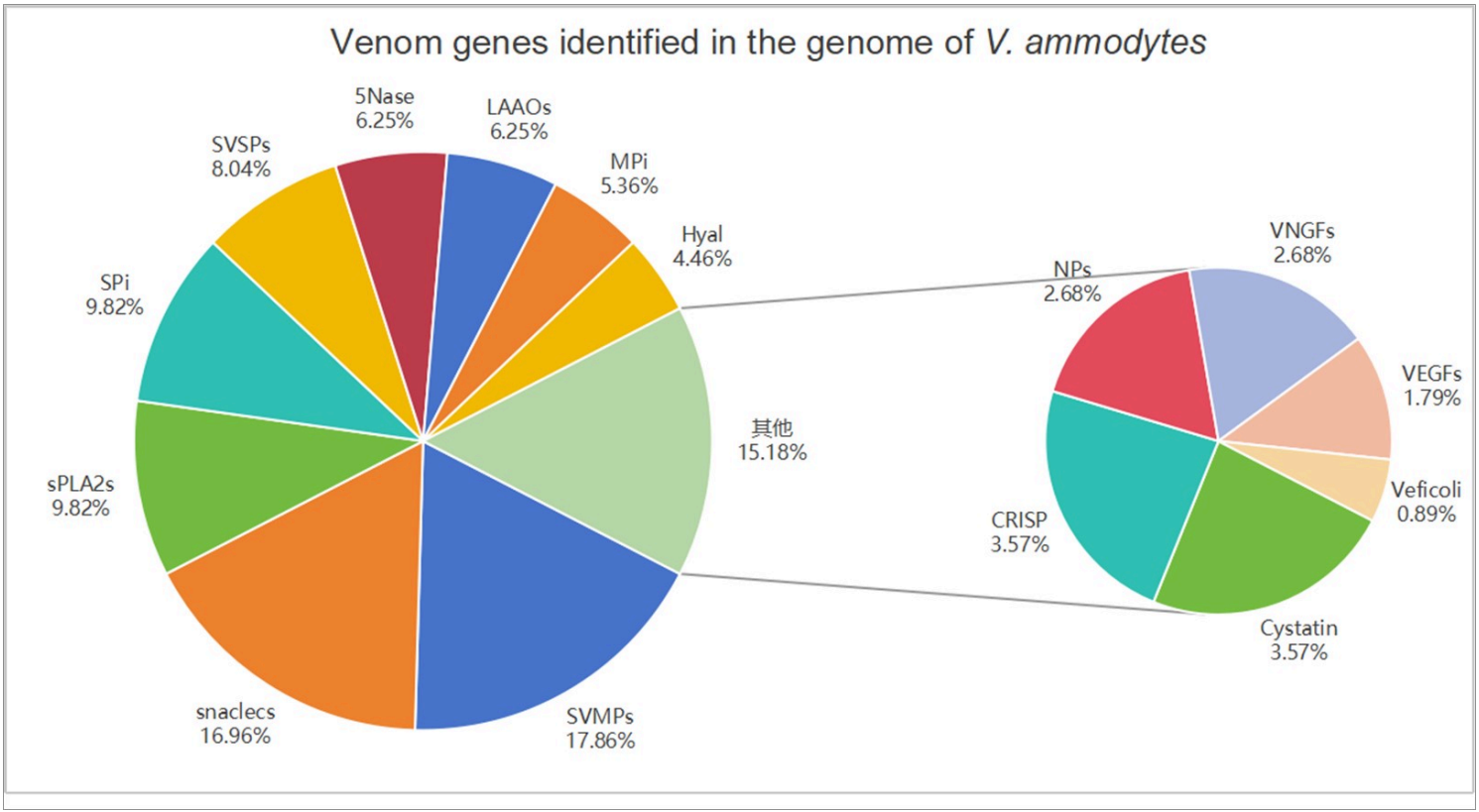

Figure 5

Summary of the venom genes of *V. ammodytes*. *Pie chart displays the percentage of venom genes identified in the genome of V. ammodytes. There are 112 venom coding genes identified for 15 protein families in its genome. SVMPs, Snaclec, sPLA2s, SPi, and SVSPs are the five major toxin families, accounting for more than 60% of the total toxin genes in the V. ammodytes ammodytes venom (62.5 %).*

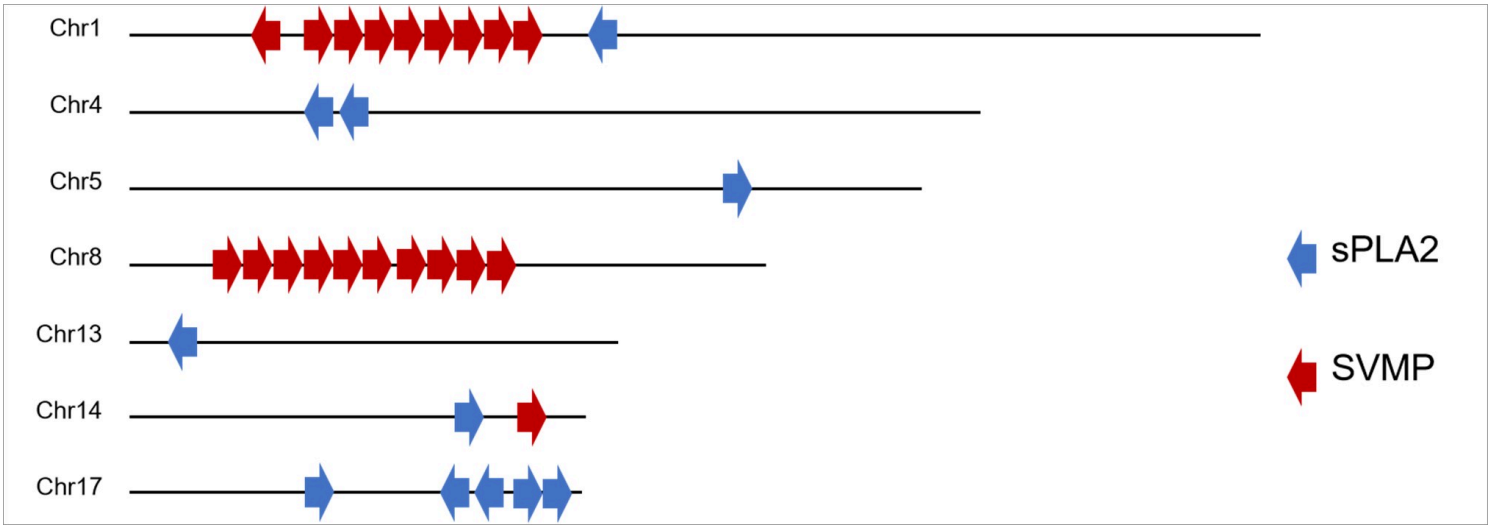

Figure 6

Syntenic regions of the *V. ammodytes ammodytes* containing $sPLA_2$ and SVMP genes.

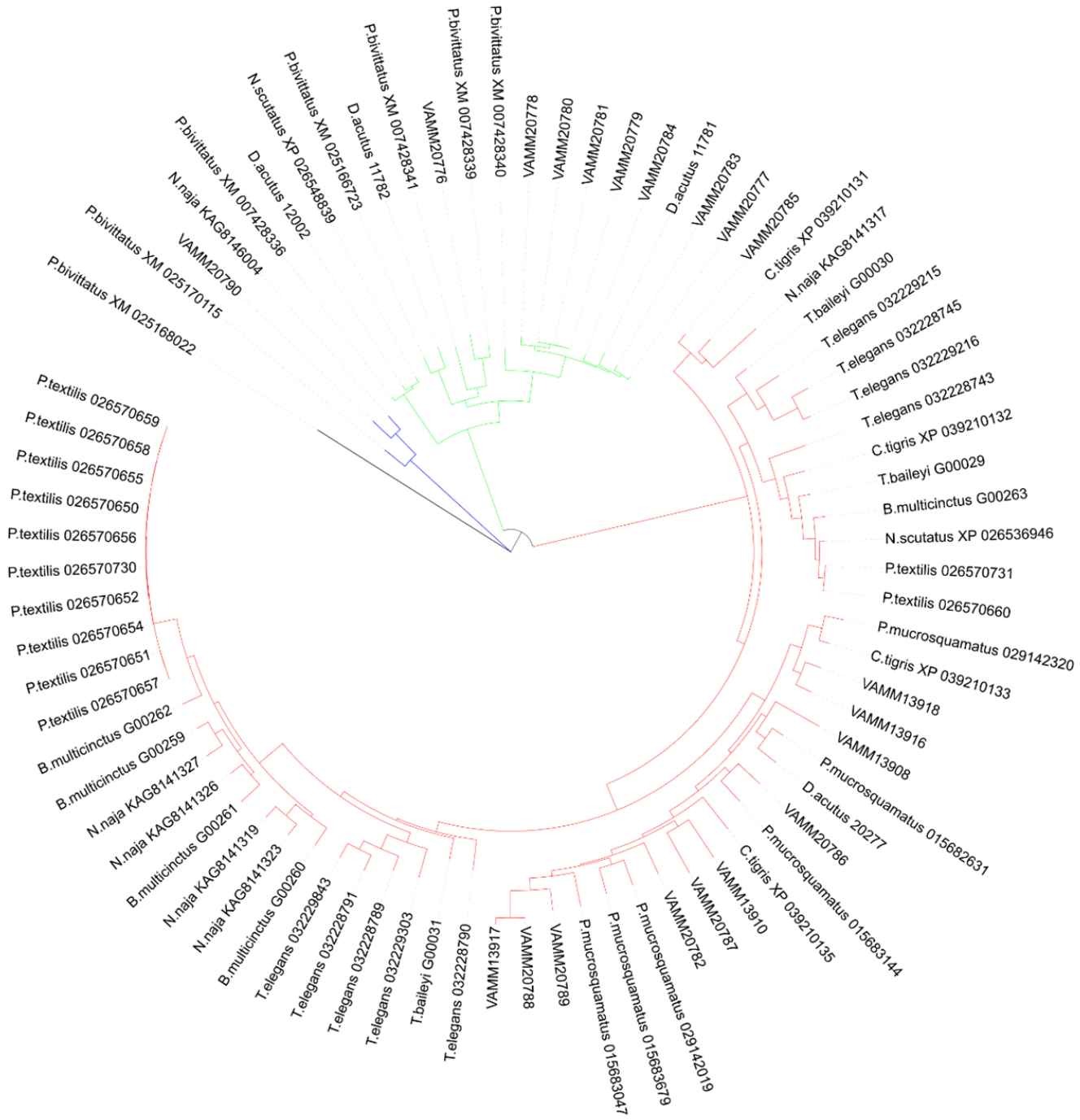

Figure 7

Evolutionary tree of the SVMPs gene family in ten snakes. Maximum likelihood tree for 78 SVMP genes. Different colors indicate different tree clades.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFiguresandTables.docx