

Chromosome-level reference genome for the medically important Arabian horned viper (*Cerastes gasperettii*)

Gabriel Mochales-Riaño^{1,*}, Samuel R. Hirst², Adrián Talavera¹, Bernat Burriel-Carranza^{1,3}, Viviana Pagone¹, Maria Estarellas¹, Theo Busschau⁴, Stéphane Boissinot⁴, Michael P. Hogan^{5,6}, Jordi Tena-Garcés⁷, Davinia Pla⁷, Juan J. Calvete⁷, Johannes Els⁸, Mark J. Margres², and Salvador Carranza¹

¹Institute of Evolutionary Biology (IBE), CSIC-Universitat Pompeu Fabra, 08003 Barcelona, Spain

²Department of Integrative Biology, University of South Florida, Tampa, FL 33620, USA

³Museu de Ciències Naturals de Barcelona, P^o Picasso s/n, Parc Ciutadella, 08003 Barcelona, Spain

⁴New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

⁵Department of Biological Sciences, Florida State University, Tallahassee, FL 33306, USA

⁶Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109-1085, USA

⁷Evolutionary and Translational Venomics Laboratory, Instituto de Biomedicina de Valencia, Consejo Superior de Investigaciones Científicas (CSIC) 46010 Valencia, Spain

⁸Breeding Centre for Endangered Arabian Wildlife, Environment and Protected Areas Authority, Sharjah, United Arab Emirates

*Correspondence address. Gabriel Mochales-Riaño, Institute of Evolutionary Biology (IBE), CSIC-Universitat Pompeu Fabra, 08003 Barcelona, Spain. E-mail: gabriel.mochales@csic.es

Abstract

Venoms have traditionally been studied from a proteomic and/or transcriptomic perspective, often overlooking the true genetic complexity underlying venom production. The recent surge in genome-based venom research (sometimes called “venomics”) has proven to be instrumental in deepening our understanding of venom evolution at the molecular level, particularly through the identification and mapping of toxin-coding loci across the broader chromosomal architecture. Although venomous snakes are a model system in venom research, the number of high-quality reference genomes in the group remains limited. In this study, we present a chromosome-resolution reference genome for the Arabian horned viper *Cerastes gasperettii* (NCBI: txid110202), a venomous snake native to the Arabian Peninsula. Our highly contiguous genome (genome size: 1.63 Gbp; contig N50: 45.6 Mbp; BUSCO: 92.8%) allowed us to explore macrochromosomal rearrangements within the Viperidae family, as well as across squamates. We identified the main highly expressed toxin genes within the venom glands comprising the venom’s core, in line with our proteomic results. We also compared microsyntenic changes in the main toxin gene clusters with those of other venomous snake species, highlighting the pivotal role of gene duplication and loss in the emergence and diversification of snake venom metalloproteinases and snake venom serine proteases for *C. gasperettii*. Using Illumina short-read sequencing data, we reconstructed the demographic history and genome-wide heterozygosity of the species, revealing how historical aridity likely drove population expansions. Finally, this study highlights the importance of using long-read sequencing as well as chromosome-level reference genomes to disentangle the origin and diversification of toxin gene families in venomous snake species.

Keywords: toxin evolution, gene synteny, genomics, transcriptomics, venom

Background

The rise of genomics in non-model organisms has led to an increase in the number of high-quality reference genomes available in recent years [1–6]. Advances in sequencing technologies have catalyzed the study of several complex traits from a genomic perspective, such as coloration, domestication, or venom, among others [3, 7–10]. Among these, venom genomic research has been particularly important in enhancing our understanding of the origin, evolution, and dynamics of this medically relevant trait [11–14]. Venom is a potentially lethal cocktail rich in proteins and peptides (from now on referred to as “toxins”) which are actively secreted by specialized venom glands [11, 15]. Toxins can have different effects depending on their type, interactions with other molecules, and the organism in which they are introduced, with convergent outcomes in different taxa [15, 16]. Historically, venom research has primarily been conducted using proteomic and tran-

scriptomic approaches (see [7] and references therein). The identification of venom toxins and the characterization of their evolution using reference genomes is a recent and novel field [17]. Previous works have shown that changes in gene regulation can result in the activation and deactivation of venom-coding genes at all taxonomic levels and within the same individual [2, 16, 18, 19]. This suggests that transcriptomic and proteomic data are critical for studying venoms in conjunction with well annotated reference genomes to disentangle the complete number and biochemical nature of the toxins an individual can potentially transcribe [7]. Ultimately, the study of venom genomics may yield evolutionary insights into antivenom or drug discovery, as it enables the identification of unexpressed toxin-coding genes. These genes, often overlooked by transcriptomic or proteomic approaches unless ontogeny analyses or in-depth venom expression studies are performed, may target unique physiological pathways. Such

Received: July 5, 2024. Revised: December 13, 2024. Accepted: March 4, 2025

© The Author(s) 2025. Published by Oxford University Press GigaScience. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

discoveries could lead to novel therapies for human illnesses including but not limited to cancer [11, 20–22]. Unexpressed toxin-coding genes are particularly noteworthy because they may represent evolutionary “reservoirs” of bioactive molecules. These genes could encode toxins with unique mechanisms of action, offering untapped potential for drug discovery or therapeutic innovation.

Venom has evolved independently in multiple groups including cnidarians, molluscs, arthropods, squamates, and even mammals [11, 15]. Venomous snakes are one of the most life-threatening animal groups to humans [23] and, therefore, a medically relevant model system in venom research. Venomous snakes are a diverse group with more than 600 species [24], in which venom has evolved with the objective of immobilizing and digesting their prey [25]. More than 370 species of venomous snake have been classified as of medically important by the World Health Organization (WHO) due to their potential severe effects on humans [26]. Snakebite is considered a neglected tropical disease, with annual mortality exceeding 100,000 victims worldwide [23, 27]. The most medically important venomous snake families are Elapidae, Viperidae, and Atractaspididae [28], although within Colubridae (*sensu lato*) there are certain medically important venomous species as well [29]. Envenomation by certain members of these families can result in a range of pathologies, spanning neurotoxic, hemotoxic, and/or cytotoxic effects depending on the number and composition of toxins. Neurotoxic venoms primarily target the central nervous system and are mainly composed of small proteins including three-finger toxins (3FTs), snake venom phospholipases A₂ group I (SV-G^I-PLA₂), or dendrotoxins, and are usually associated with elapid snakes [30]. Conversely, hemotoxic and cytotoxic venoms generally are comprised of large enzymatic proteins and protein complexes, including snake venom metalloproteases (SVMs), serine proteases (SPs) or snake venom phospholipases A₂ group II (SV-G^{II}-PLA₂), and are typically associated with viperid snakes [28, 31, 32]. Although these historical classifications have proven to be somewhat useful for treating envenomations medically, recent studies have revealed that the presence of these toxins are not exclusive to specific snake families [33].

Vipers (family Viperidae) are a monophyletic lineage of venomous snakes found across Eurasia, Africa, and America [34], and have received extensive research attention primarily due to their medical relevance [35–39]. The majority of venom studies in this group have primarily been investigated using a proteomic approach, with early venom work being highly motivated by the medical field, with a limited number of studies employing genomic approaches (but see [2, 3, 5, 40–42]). Sequencing efforts to obtain high-quality reference genomes have mainly focused on pitvipers (Crotalinae subfamily, 11 reference genomes, NCBI last accessed 13 March 2024), especially within the *Crotalus* (*n* = 6) genus, and have focused on the study of venom evolution [2, 3, 5, 43, 44]. Other viperids have also been sequenced (although in lower numbers) from both Azemiopinae and Viperinae subfamilies (one and four, respectively) [41, 45, 46]. Currently, reference genomes are only available for 16 viper species out of the total 387 total species via the NCBI genomic database [24]. Vipers display extensive variation in venom composition between and within genera [47, 48] and even intraspecifically [49, 50]. Such differences are most likely due to the high diversity of venom genes and their different effects on prey, but are also, at least in some cases, the result of introgression with related species [19, 49, 50]. This provides an extraordinary opportunity to study trait evolution both at inter- and intraspecific levels.

Native to the Arabian Peninsula, the Arabian horned viper (*Cerastes gasperettii*, family Viperidae) is a venomous snake cur-

rently recognized within the highest medical importance category (WHO; accessed July, 2024). Extending from the Sinai Peninsula to southwestern Iran in the north and reaching as far as Yemen and Oman in the south, its distribution is widespread (Supplementary Fig. S1). Found mainly in sandy habitats, this arid-adapted ground-dwelling snake with generalist requirements [51–53] is one of the most common venomous snakes found in Arabia and is responsible for occasional snakebite envenomations [54–56].

In this study, we present a high-quality chromosome-level reference genome assembly for *C. gasperettii* (NCBI: txid110202), being one of the first within the Viperinae subfamily. Our highly contiguous genome showcases a high level of similarity at the chromosome level within the Viperidae family with some minor rearrangements with elapids. Moreover, combining genomics, transcriptomics, and proteomics, we characterized the main toxins found in its venom and the location of those toxins in the genome, comparing their evolutionary history and gene copy number variation with those of other venomous species. We deciphered its adequate levels of genetic diversity. Finally, we reconstructed the demographic history for the species, revealing how historical increases in aridity likely drove population expansions. Overall, the genomic resources generated in this study provide an essential reference resource for forthcoming studies on venom evolution.

Methods

Sampling

Three adult specimens (two females and one male) of *C. gasperettii* were used for this study (Supplementary Table S1). Blood was extracted only from a single female individual (the heterogametic sex, sample CG1) to obtain high-molecular-weight (HMW) genomic DNA (gDNA). We anesthetized the individual, extracted blood from the heart, and stored this in ethanol and EDTA. For each of the 3 individuals, we extracted 12 different tissues, including the venom gland, which was stored in RNAlater until RNA extraction (Supplementary Table S1 and Supplementary Fig. S2). Before dissections, venom was extracted and snakes were allowed to recover for 4 days to maximize the venom gland transcription. We only extracted the left venom gland per individual because previous research within the same family has shown that both venom glands provide indistinguishable results [57].

DNA extraction, library preparation and sequencing

We extracted gDNA from the blood of a female individual (CG1 in Supplementary Table S1) using the MagAttract HMW Kit (Qiagen, Germany) following the manufacturer's protocols without modifications. Then, we sequenced a total of two 8 M SMRT HiFi cells in a Sequel II PacBio machine, aiming for a ~30× of coverage, at the University of Leiden. Hi-C libraries were prepared using the Omni-C kit (Dovetail Genomics), following the manufacturer's protocol and using blood stored in EDTA, at the National Center for Genomic Analyses (CNAG), in Barcelona, Spain. The library was paired-end sequenced on a NovaSeq 6000 (2 × 150 bp) following the manufacturer's protocol for dual indexing and aiming for a coverage of ~60×. Finally, we sequenced short-read whole-genome data of the same individual using a NEB Ultra II FS DNA kit; the library was paired-end sequenced on a NovaSeq 6000 (2 × 150 bp) at the Core sequencing platform from the New York University of Abu Dhabi, aiming for ~70× depth of coverage.

RNA extraction, library preparation and sequencing

We extracted RNA from the same 3 individuals described above (Supplementary Table S1 and Supplementary Fig. S2). RNA was isolated using a HighPurity Total RNA Extraction Kit (Canvax, Valladolid, Spain). We selected a total of 35 samples (Supplementary Table S2). RNA libraries were prepared with the VAHTS Universal V8 RNA-seq Library Prep Kit, and being strand-specific, were sequenced on a NovaSeq 6000 (2×150 bp) aiming for an average of 40 M read pairs per sample (Supplementary Table S2), but we first sequenced the reference individual and subsequently the other 2 samples. Moreover, we sequenced one 8 M SMRT HiFi cell on a Sequel II PacBio machine containing 2 Iso-seq HiFi libraries at the University of Leiden: one containing only the venom gland, and the second library being a pool of 8 high-quality tissues (brain, kidney, liver, gallbladder, spleen, tongue, pancreas, and ovary).

Genome assembly and scaffolding

Quality control of HiFi and Illumina reads was performed using FastQC (RRID:SCR_014583) v.0.12.1 [58] and adapters were removed with cutadapt (RRID:SCR_011814) v.4.9 [59]. In order to initially explore the genome size, heterozygosity levels, and coverage data, we generated a k-mer profile with Meryl (RRID:SCR_026366) v.1.4.1 [60], using the raw HiFi reads and default parameters, and visualized it with GenomeScope2 (RRID:SCR_017014) v.2.0.1 [61]. Then, we assembled the genome following the VGP assembly pipeline v.2.0 [62]. PacBio HiFi reads were assembled into contigs using the software Hifiasm (RRID:SCR_021069) v.0.21.0 [63], producing primary and alternate assemblies. We used purge_dups (RRID:SCR_021173) [64] to remove haplotypic duplicates from the primary assembly and added them to the alternate assembly. Then, we scaffolded the resulting haplotypic assembly using the Hi-C data with SALSA2 (Salsa, RRID:SCR_022013) v.1 [65], with default parameters. Following the VGP assembly pipeline [62], manual curation was performed with Pretext (RRID:SCR_022024) v.0.2.5. Breaks were not manually created and we joint contigs on gaps previously identified by SALSA2 (Salsa, RRID:SCR_022013). We used the $\sim 78\times$ Illumina data to polish the assembly with one round of Pilon (RRID:SCR_014731) v.1.24 [66]. The mitochondrial genome was obtained with GetOrganelle (RRID:SCR_022963) v.1.7.7.1 [67], using the available mitochondrial genome of several *Echis* species (*E. coloratus*, *E. carinatus*, and *E. omanensis*) to seed the assembly (NCBI: SRX18902082, SRX18902083, SRX18902084, respectively).

Genome assembly quality evaluation

Quality assessment and general metrics for the final assembly were estimated with both QUAST (Quast, RRID:SCR_011228) v.5.1.0 [68] and gfastats (RRID:SCR_026368) v.1.3.8 [69]. Possible contaminations were evaluated with BlobToolKit (Blobtools, RRID:SCR_017618) v.4.4.0 [70] using the NCBI taxdump database. We also used MitoFinder v.1.4.2 [71, 72] to confirm that the mitochondrial genome was absent in the assembled nuclear reference genome. Completeness of the genome assembly was assessed with BUSCO (Busco, RRID:SCR_015008) v.5.3.0. against the sauripsida_odb10 database ($n = 7,480$).

Genome annotation

First, we identified repetitive elements using RepeatModeler (RRID:SCR_015027) v.2.0.3 [73] for *de novo* predictions of repeat families. To annotate genome-wide complex repeats, we used RepeatMasker (RRID:SCR_012954) v.4.1.3 [74] with default settings

to identify known Tetrapoda repeats present in the curated Repeat database [75]. Then, we ran 3 iterative rounds of RepeatMasker to annotate the known and unknown elements identified by RepeatModeler in order to maximize the known elements at the expense of diminishing the unknown elements. Later, we soft-masked the genome for simple repeats. We used GeMoMa (GeMoMa, RRID:SCR_017646) v.1.9 [76] to annotate protein-coding genes, combining both the RNA-seq data generated in this study as described above (already mapped in to our new assembly) as well as annotations from 6 other squamate genomes already published: *Crotalus adamanteus* [2], *Crotalus tigris* [3], *Ophiophagus hannah* [17], *Naja naja* [6], *Crotalus ruber* [42] and *Crotalus viridis* [5]. We quality checked and removed the adapters of the RNA-seq data using fastp v.0.23.3 [77], as well as mapped the transcriptomic data to our new reference genome with Hisat2 (RRID:SCR_015530) v.2.2.1 [78]. Additionally, we also removed the adapters for the Iso-seq data with fastp (RRID:016962) v.0.23.3 [77] and mapped the long-read transcriptomic data to our new reference genome with pbmm2 (RRID:SCR_025549), collapsing mapped reads into unique isoforms with isoseq3 and annotating with GeneMarkS-T (GeneMark, RRID:SCR_011930) v.5.1 [79]. We combined both annotations (GeMoMa and GeneMarkS-T) with TSEBRA [80]. We BLASTp (blastp, RRID:SCR_001010) our predicted proteins to a Uniprot protein database for a total of 10 species (*C. gasperettii*, *C. vipera*, *C. cerastes*, *Anolis carolinensis*, *C. viridis*, *C. tigris*, *C. ruber*, *C. adamanteus*, *O. hannah* and *N. naja*). Simultaneously, we ran Interproscan v.5.72 [81] on our predicted proteins. Then, we combined both functional annotations with AGAT v.1.4.1 [82]. Finally, as toxin-coding gene families are known to occur in large tandem arrays and the number of paralogs can be underestimated in particular gene families [5], we performed additional annotation steps for toxin genes: Following [3], we used a combination of empirical annotation in FGENSESH+ (FGENSESH, RRID:SCR_011928) [83], as well as manual annotation using RNA-seq and Iso-seq alignments; the former identified all genes regardless of expression, whereas the latter was used to explicitly identify expressed toxins.

Chromosome-level analyses

Chromosomal synteny was explored between our new chromosome-level reference genome for the Arabian horned viper together with the Eastern diamondback rattlesnake (*C. adamanteus*) [2], the Indian cobra (*N. naja*) [6], and the brown anole (*Anolis sagrei*) [84] using MCscan (RRID:SCR_017650) v.1.4.23 [85]. Protein sequences from each of the 3 venomous snakes were extracted using AGAT v.1.2.1 [82] and were pairwise aligned with LAST [86], implemented in the JCVI python module [87]. A first alignment was used between the 3 species to identify chromosomes assembled in the reverse complement, which were corrected using SAMtools faidx (samtools, RRID:SCR_002105) v.1.18.1 [88] using both options reverse-complement and mark-strand. Gene annotations for the new reference (with the corresponding reversed chromosomes) were annotated using GeMoMa v.1.9 [76], and MCscan was rerun. The last 4 scaffolds (14, 15, 16, and 17) from *A. sagrei* were removed because no orthologous groups were found.

Transcriptomics

After adapter trimming and quality control using fastp v.0.23.3 [77], we mapped our RNA-seq reads to the reference genome of *C. gasperettii* using Hisat2 (RRID:SCR_015530) v.2.2.1 [78]. Gene expression raw counts per gene across all samples were calculated with StringTie (RRID:SCR_016323) [89]. Initial exploration of our

transcriptomic data revealed a clear batch effect for one of the 3 samples (Supplementary Fig. S4), due to the low mapping of that sample to our reference genome. Therefore, we decided to remove individual CG1 from future RNA-seq analyses. Moreover, to avoid pseudoreplication, we also removed the accessory gland from individual CG009 due to its high similarity with the venom gland, suggesting that the venom gland rather than the accessory gland was sampled (Supplementary Fig. S4). Differential expression analyses were carried out with the DESeq2 package (RRID:SCR_015687) v.1.42.0 [90] from R v.4.4.2 [91]. Prior to analysis, genes with <10 counts across all samples were filtered out. For comparisons, we defined 2 groups: venom glands versus all other tissues. DESeq2 uses a negative binomial generalized linear model to estimate differences in gene expression, and the P-values were adjusted for multiple testing using the Benjamini-Hochberg method to control the false discovery rate. Genes with an adjusted P-value <0.01 and a fold change (FC) >2 were considered significantly differentially expressed. Finally, we identified the highly expressed genes found in the venom gland as well as the toxins uniquely expressed in the venom gland (following [6]) which were defined as: (1) genes expressed in the venom gland (transcripts per million (TPM) > 500), (2) differential upregulated genes with a FC > 2 comparing venom glands with all other tissues, and (3) unique to venom glands (TPM < 500 in all other tissues).

Proteomics

A bottom-up mass spectrometry (MS) strategy [92] was used to characterize the venom of *C. gasperettii*. Briefly, the venom proteome (pooled from individuals CN6134 and CN6135, both from the United Arab Emirates (UAE); Supplementary Table S1) was submitted to reverse-phase high-performance liquid chromatography decomplexation followed by SDS-PAGE analysis in 12% polyacrylamide gels run under non-reducing and reducing conditions. Protein bands were excised from Coomassie Brilliant Blue-stained gels and subjected to automated in-gel reduction and alkylation on a Genomics Solution ProGest Protein Digestion Workstation. Tryptic digests were submitted to MS/MS analysis on a nano-Acquity UltraPerformance LC (UPLC) equipped with a BEH130 C₁₈ (100 µm × 100 mm, 1.7 µm particle size) column in-line with a Waters SYNAPT G2 high-definition mass spectrometer. Doubly and triply charged ions were selected for CID-MS/MS. Fragmentation spectra were matched against a customized database including the bony vertebrates taxonomy dataset of the NCBI nonredundant database (release 258, 15 October 2023) plus the species-specific venom gland transcriptomic and genomic protein sequences gathered in this work. Search parameters were as follows: enzyme: trypsin (two-missed cleavage allowed); MS/MS mass tolerance for monoisotopic ions: ±0.6 Da; carbamidomethyl cysteine and oxidation of methionine were selected as fixed and variable modifications, respectively. Assignments with significance protein score threshold of $P < 0.05$ (Mascot score > 43) were taken into consideration, and all associated peptide ion hits were manually validated. Unmatched MS/MS spectra were *de novo* sequenced and manually matched to homologous snake toxins available in the NCBI nonredundant protein sequences database using the default parameters of the BLASTP program.

Local synteny analyses

To explore toxin genomic organization across (sub)families, we used BLASTn, incorporating both toxin and nontoxin paralogs to identify the genomic location of SVMs, SVSPs, and PLA₂ toxin families, across the genome of *C. gasperettii*, *C. adamanteus*, *N. naja*

and *A. feae*. We excluded *A. feae* for SVSP and SVM local synteny analyses because those families were not assembled onto a single contig in the *A. feae* genome. Then, we aligned those regions using Mafft [93]: for SVMs in CHR8:16.506.135 to CHR8:17.374.029, for SVSPs in CHR9: 17.531.416 to CHR9:17.788.049 and for PLA₂ in CHR17:7.882.542 to CHR17:7.916.827. Each species was annotated within the MSA using its own annotation as a reference in Geneious Prime 2023.0.4. Results were plotted using the gggenomes package [94] from R v.4.4.2 [91].

Toxin phylogenies

We used phylogenetic inference to study the evolutionary history for the main groups of toxins (i.e., SVMs and SVSPs), which were the most abundant in the proteome of *C. gasperettii*, as well as PLA₂ because this family has been widely studied within the Viperidae family [12, 41]. For the 3 main toxin families, we selected available toxin genes as well as nontoxin paralogous genes from venomous species; we also included other nontoxin paralogous genes from nontoxic species (for details, see Supplementary Datasets for the 3 main toxins). When nuclear sequences were obtained, we translated CDS to protein sequence, and then protein sequences were aligned with Mafft (RRID:SCR_011811) v.7 [93]. Following [13], we built a phylogeny for each of the toxin groups with the translated CDS sequences, as explained above, using Phylml (RRID:SCR_014629) v.3.3 [95], implementing the Dayhoff substitution model and validating our inferred tree with aBayes support.

Demographic history

We inferred the demographic history of *C. gasperettii* by implementing the Pairwise Sequential Markovian Coalescent (PSMC) software (RRID:SCR_017229) v.0.6.5 [96] on the short-read whole-genome data. Heterozygous positions were obtained from bam files with the Samtools v.1.9 mpileup function [97], and data were filtered for low mapping (<30) and base quality (<30). Minimum and maximum depths were set at a third (27×) and twice (156×) the average coverage. Only autosomal chromosomes were considered. We used a squamate mutation rate of 2.4×10^{-9} substitutions/site/generation and a generation time of 3 years, following [98, 99], respectively. A total of 10 bootstraps were calculated, plotting the final results with the psmc_plot.pl function from PSMC.

Genomic diversity

We downloaded Illumina data for *Bothrops jararaca* (SRR13839751 from [40], *Crotalus viridis* (SRR19221440; [5]), *Naja kaouthia* (SRR8224383; [100]), *N. naja* (SRR10428156; [6]) and *Sistrurus tergeminus* (SRR12802282; [101]). Then, we filtered for quality (Phred score: 30) and removed adapters with fastp v.0.23.3 [77]. Trimming of poly-G/X tails and correction in overlapped regions were specified. All other parameters were set as default. Filtered sequences were visually explored with FastQC (fastQC, RRID:SCR_014583) v.0.12.1 [58] to ensure data quality and absence of adapters. *C. gasperettii* filtered reads were mapped against the new reference genome of *C. gasperettii* using the bwa mem algorithm (bwa, RRID:SCR_010910) v.0.7.17 [102]. *B. jararaca*, *C. viridis*, and *S. tergeminus* were mapped against the *C. viridis* [5] reference genome and *N. naja* and *N. kaouthia* were mapped against the *N. naja* reference genome [6]. Mapped reads were sorted with Samtools (RRID:SCR_002105) v.1.9 [97] and duplicated reads were marked and removed with PicardTools (Picard, RRID:SCR_006525) v.2.28.0 [103]. Reads with mapping quality <30 were discarded. SNP calling was carried out with HaplotypeCaller (GATK, RRID:SCR_001876)

v.4.1.3.0 [104], with BP_resolution and split by chromosome. For each chromosome, individual genotypes were joined using CombineGVCFs with convert-to-base-pair-resolution, and the GenotypeGVCFs tool was then applied to include nonvariant sites. Finally, for each individual, the whole dataset split by chromosome was concatenated with bcftools concat (bcftools, [RRID:SCR_005227](#)) [88], keeping only the autosomes. Then, for each sample, we used the raw dataset to calculate average genome heterozygosity. We generated nonoverlapping sliding windows for each of the reference genomes and included only sites (both variant and invariant) with site quality >30 (QUAL field in a VCF file from GATK). Only windows containing more than 60,000 unfiltered sites were considered. Visualization was carried out with ggplot2 ([RRID:SCR_014601](#)) [105] in R v.4.4.2 [91].

Results and Discussion

Genome assembly and annotation

We generated a high-quality chromosome-level assembly for *C. gasperettii* by combining PacBio HiFi (65 Gbp of data), Hi-C (96 Gbp of data), and Illumina data (135 Gbp of data) (Fig. 1 and [Supplementary Fig. S3](#)). First, we *de novo* assembled the HiFi reads into 1,018 contigs (N50 = 45.7 Mbp; longest contig: 149.99 Mbp). Then, using the proximity ligation data (i.e., Hi-C), we scaffolded the genome into 319 scaffolds (N50 = 111.38 Mbp; largest scaffold: 345.38 Mbp). After manual curation, the scaffolding parameters of our genome were improved (N50 = 214.14 Mbp; largest scaffold: 361.99 Mbp), containing 99.44% of the genome present in 19 scaffolds or pseudochromosomes (7 macro-, 10 micro-, Z and W sex chromosomes; Table 1 and Fig. 1B). The total genome length was 1.63 Gb, similar to other venomous snakes [3, 5, 6, 17] (Table 1), with a contig N50 of 45.6 Mbp, ~3.3 times more contiguous than the *N. naja* genome [6], ~228 times more contiguous than the *A. sagrei* genome [84], but 0.67 times less contiguous than the recently published *C. adamanteus* genome [2], making it one of the most contiguous chromosomal squamate genomes assembled to date (Table 1). We assessed the completeness of the assembly using BUSCO [106] with the sauropsida gene set ($n = 7,480$). Upon evaluation, we successfully identified 92.8% of the genes (91.4% single-copy, 1.4% duplicated), while the remaining genes were fragmented (1%) or missing (6.2%; Fig. 1). For the *de novo* assembly, GC content and repeat content were 37.87% and 43.63%, respectively. The repetitive landscape was dominated by retroelements (30.25%), with a majority of LINEs (21.25%) ([Supplementary Table S3](#)). Finally, we annotated 27,158 different protein-coding genes within our assembly, with a total of 194 putative toxins or toxin-paralog genes. Toxin genes were found in both macro- and microchromosomes (Fig. 1), and were found on individual contigs. Finally, we also found a battery of 3FTxs and myotoxin-like genes, but they were not represented in our proteome and RNA-seq dataset (see below).

Genomic architecture highly conserved among vipers

Whole-genome synteny comparisons showed similarity between *C. gasperettii* and *C. adamanteus*, with large syntenic blocks both within macro- and microchromosomes (Fig. 2). Some chromosomal rearrangements were observed between vipers and elapids, as previously discussed by [6], with a fission of chromosome 4 in *N. naja* to form chromosomes 5 and 7 in vipers, and a fusion of chromosomes 5 and 6 in *N. naja* to form chromosome 4 in vipers. Interestingly, several chromosomal rearrangements between lizards

and snakes have occurred: we found several fission events in the *A. sagrei* genome, including one fission from chromosome 2 that originates the current Z chromosome in snakes (Fig. 2).

Toxins uniquely expressed in the venom glands

Our analyses of multitissue transcriptomic data (23 samples from 2 individuals covering 13 different tissues) reported a total of 23,178 expressed genes (TPM > 1). A heatmap of the 2,000 most variable genes reported unique upregulated genes for each of the analyzed tissues ([Supplementary Fig. S5](#)). The venom gland transcriptome contained a total of 7,237 genes expressed (TPM > 500), including a total of 65 putative toxin genes. From those, we did not detect any 3FTxs and/or myotoxin-like gene transcripts. Differential gene expression analyses revealed a total of 161 genes (33 putative toxin genes) that were differentially upregulated (FC > 2 and 1% false discovery rate) in venom glands compared to other tissues (Fig. 3A and [Supplementary Figs S6 and S7](#)). Finally, a total of 10 toxin genes (CRISP2, SVMP9, SVMP10, SVSP8, SVSP7, SVSP5, CTL14, CTL15, SVSP4, and SVMP13) were uniquely expressed in the venom gland, encoding for the minimal core venom effector (Fig. 3A) [6], and in line with the main toxins found within the proteome (Fig. 3B), although some differences were observed (such as the absence of PLA₂ within the highly expressed genes), possibly due to individual venom differences. These 10 genes, together with other SVMPs, SVSPs, and C-type lectins (CTLs), were highly expressed in the venom gland and form the core toxic effector components of the venom. Targeting the core toxins together with other well-known modulators of venom may help manufacture of synthetic antivenom treatments and improve neutralization tests of current antivenoms [6]. However, more transcriptomic data should be incorporated to correct for potential ontogenetic and geographical variation in venom composition in *C. gasperettii* [18, 107].

SVSPs and SVMPs as main toxins

Venom proteomics identified SVSPs and SVMPs as the most abundant toxin families within the venom of *C. gasperettii*, with 37.38% and 22.19% of the venom being composed by peptides from those 2 families, respectively (Fig. 3B); the dominance of these 2 toxin families is consistent with previous research on the same genus [108, 109]. Other toxin families identified were DISI (12.74%), CTL (7.25%), PLA₂ (5.47%), cysteine-rich secretory proteins (CRISP; 4.34%) or L-amino acid oxidase (LAO; 1.71%) (Fig. 3B). We did not detect any 3FTx or myotoxin-like peptides within the proteome.

SVMPs

We analyzed the evolution of venom of the most abundant venom toxin groups (i.e., SVMPs and SVSPs, as well as PLA₂). After a thorough manual curation, we used comparative genomics to evaluate the number and position of those genes in comparison with the Indian cobra (*N. naja*), the Eastern diamondback rattlesnake (*C. adamanteus*), and Fea's viper (*Azemiops feae*). We reported a total of 13 fully contiguous tandem array SVMPs for *C. gasperettii* (Fig. 4A), next to the nontoxic paralogous gene ADAM28 and flanked by the NEFL and NEFM nontoxic genes. Microsyntenic analyses showed gene copy number variation between the studied species (Fig. 4A). Overall, we can see an expansion in the number of SVMPs within the Viperidae family, particularly in *C. adamanteus* (22 copies unique to vipers and 10 lineage-specific copies) but also in *C. gasperettii* (12 copies unique to vipers and 1 lineage-specific copy) (Fig. 4A). Then, we reconstructed the evolutionary history of this toxin family (Figs 4B and [Supplementary Fig. S8](#)). Phylogenetic

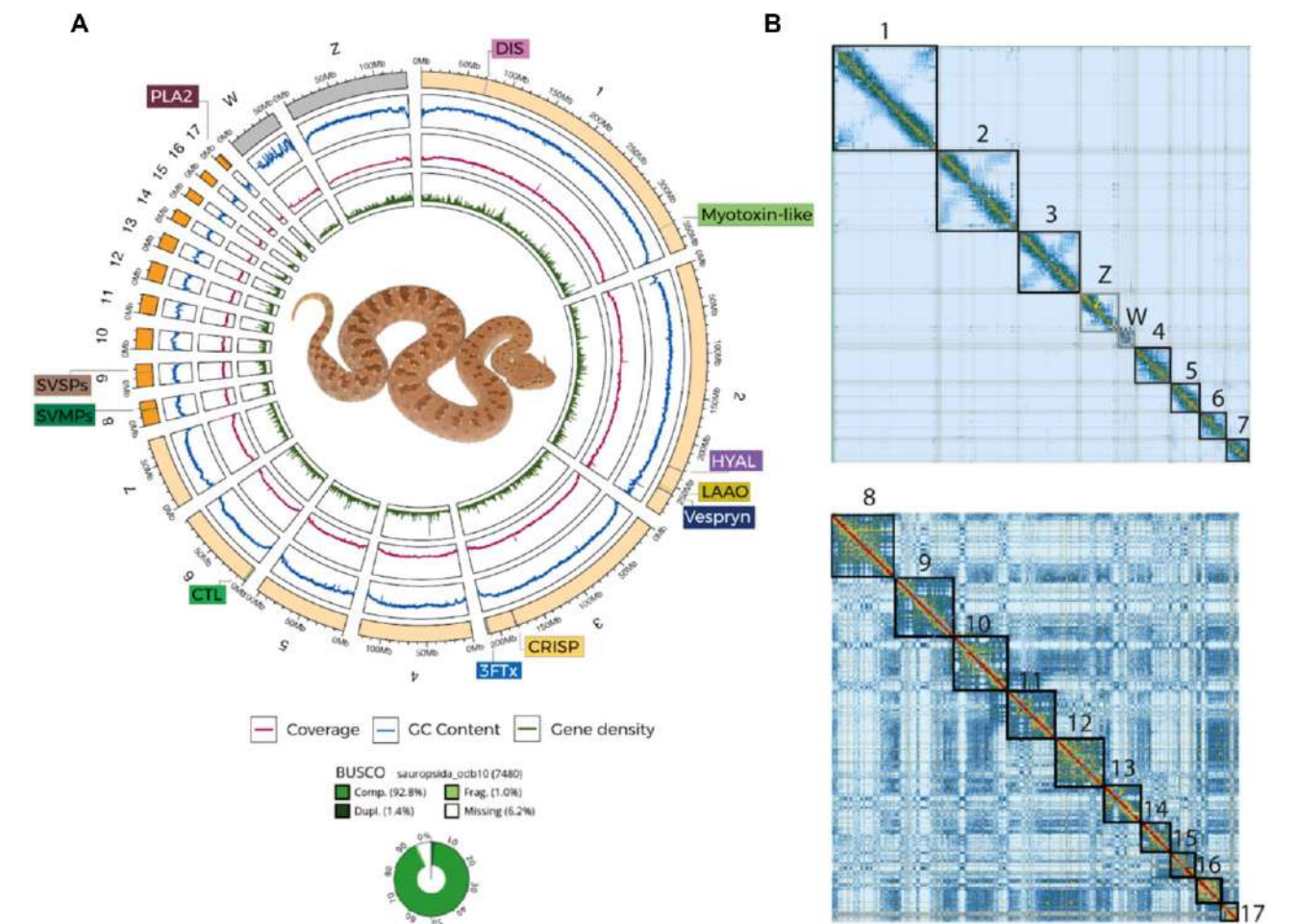


Figure 1: (A) Reference genome for *C. gasperettii*, including BUSCO score, GC content, coverage level, and the main toxins found within the genome. Macrochromosomes are shown in light orange; microchromosomes are shown in bright orange. Sex chromosomes are shown in gray. DIS: disintegrins; HYAL: hyaluronidases; LAAO: L-amino acid oxidase; CRISP: cysteine-rich secreted proteins; CTL: C-type lectins. (B) HiC contact map for the genome.

Table 1: Comparison of our new reference genome for *C. gasperettii* with other high-quality squamate genomes. Best value per category is shown in bold

	<i>C. gasperettii</i>	<i>C. adamanteus</i>	<i>N. naja</i>	<i>A. sagrei</i>
Genome size	1.63 Gbp	1.69 Gbp	1.79 Gbp	1.92 Gbp
Number of scaffolds	221	27	1,897	3,738
Scaffold N50	214.14 Mbp	208.9 Mbp	223.35 Mbp	253.58 Mbp
Scaffold L50	3	3	3	4
Contig N50	45.6 Mbp	67.5 Mbp	13.06 Mbp	0.2 Mbp

analyses for this toxin group reported a highly supported clade comprising ADAM28 peptides, the nontoxic paralogous gene. The second clade of orthologous toxin-peptides were found within both elapid and viperid families (including species from Crotalinae and Viperinae subfamilies in viperids; [Supplementary Fig. S8](#)) as well as 2 SVMPs from *A. ferox*. Interestingly, we report a new toxin-coding gene within *C. gasperettii* with a different evolutionary history, as it did not share orthology with any other gene (Fig. 4B). This new gene likely arose from a duplication event of SVMP13, within the group of SVMP MDC1 toxins ([Supplementary Fig. S8](#)). Our discovery of a novel SVMP gene in *C. gasperettii* adds to the growing body of work on the dynamic evolution of venom

systems. Similar gene expansions and duplications have been observed in other species, such as PLA₂ toxin-coding genes found in the venom of *A. ferox* [41], highlighting the lineage-specific nature of venom evolution. The gene we identified, possibly arising from an SVMP13 duplication, does not share orthology with genes in other species, suggesting the presence of hidden toxin diversity in venom systems. This discovery highlights the importance of using genomics in studying venom evolution, as this putatively toxic gene was not found to be differentially upregulated in the venom gland or recovered in the proteome (Fig. 3). More genomic data will indicate if SVMP12 is unique for the Viperinae subfamily, the *Cerastes* genus, or if it is only found in *C. gasperettii*. All other

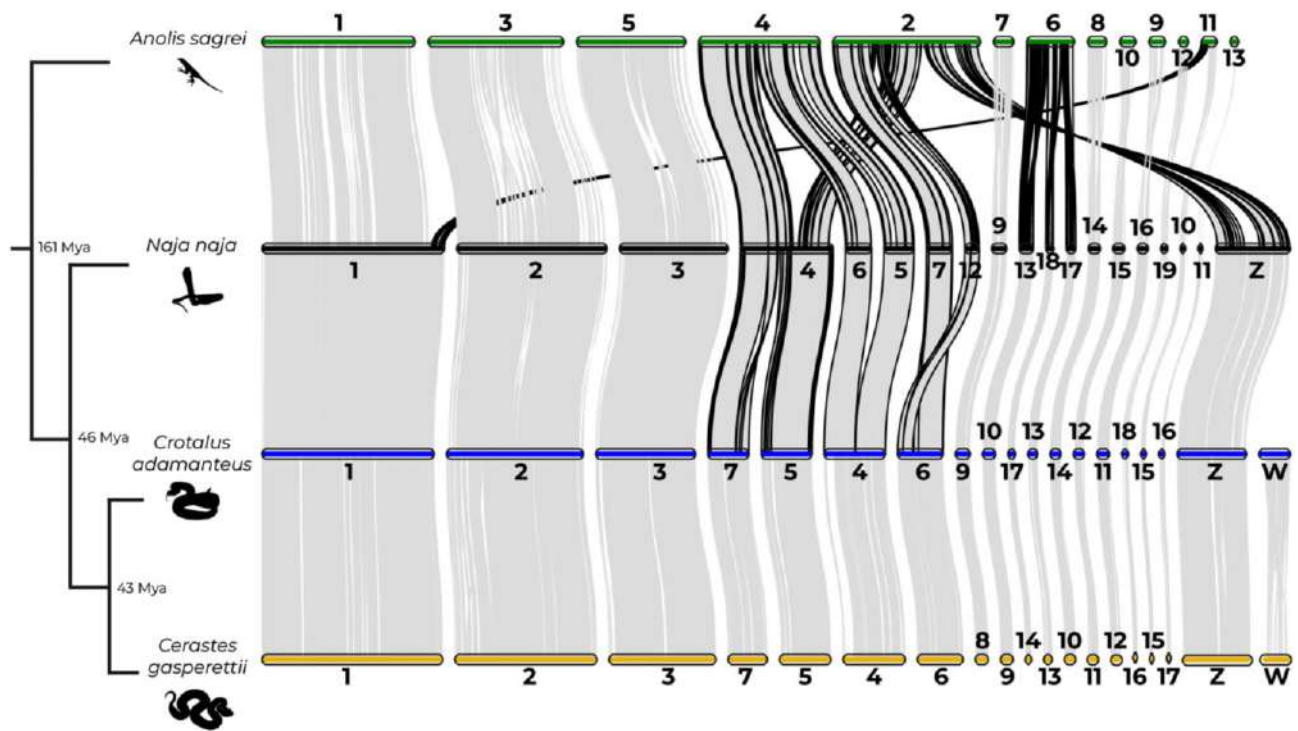


Figure 2: Chromosome-level analyses for one Elapidae (*N. naja*), one Crotalinae (*C. adamanteus*), and one Viperinae (*C. gasperettii*) species, with *A. sagrei* as the outgroup. The 4 smallest scaffolds (14, 15, 16, and 17) of *A. sagrei* were removed because no orthologous groups were found with other species. Borders of regions showing evidence for chromosomal rearrangements are shown in black. Estimates for branch times obtained from TimeTree.org based on divergence times between Iguania and Serpentes, Elapidae and Viperidae, and Crotalinae and Viperinae, respectively.

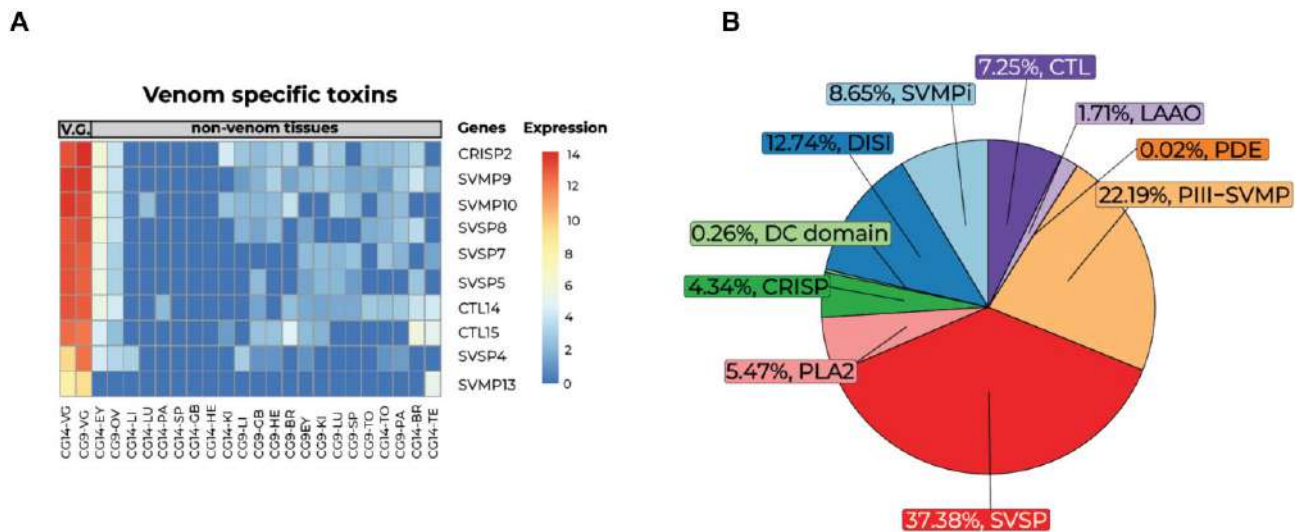


Figure 3: Main toxins found in both the transcriptome and proteome of *C. gasperettii*. (A) Transcriptomic results with genes upregulated and exclusively found in the venom gland for both individuals. Each column represents a different tissue type per sample. Rows show the different genes, and colors correspond to different expression levels. VG: venom gland; EY: eye; OV: ovary; LI: liver; LU: lung; PA: pancreas; SP: spleen; GB: gallbladder; HE: heart; KI: kidney; LI: liver; BR: brain; TO: tongue; TE: testis. (B) Proteomic results of venom composition for a pool of 2 individuals of *C. gasperettii*. The pie chart displays the relative abundances of the toxin families found in the proteome of the *C. gasperettii* venom. PDE, phosphodiesterases.

clades were unique to viperids (and some exclusive only to crotalids), except for a clade composed by SVMPs unique to elapids, as previously discussed in [6]. Interestingly, one of the toxins (SVMP8) was not a class P-III SVMP, as it clusters within the MAD-4/5 clade (class P-II SVMP), contrary to the proteomic results where all SVMPs were categorized within the class P-III (Fig. 3B). Although there has been a clear expansion of the SVMP family within the

Crotalus genus, our results suggest that the origin of that expansion was at the beginning of the Viperidae family, as most of the groups are also present within the Viperinae subfamily.

PLA₂

Regarding PLA₂, we report 2 tandem repeat venom genes for *C. gasperettii* within the non-toxic PLA₂-g2E and PLA₂-g2F array

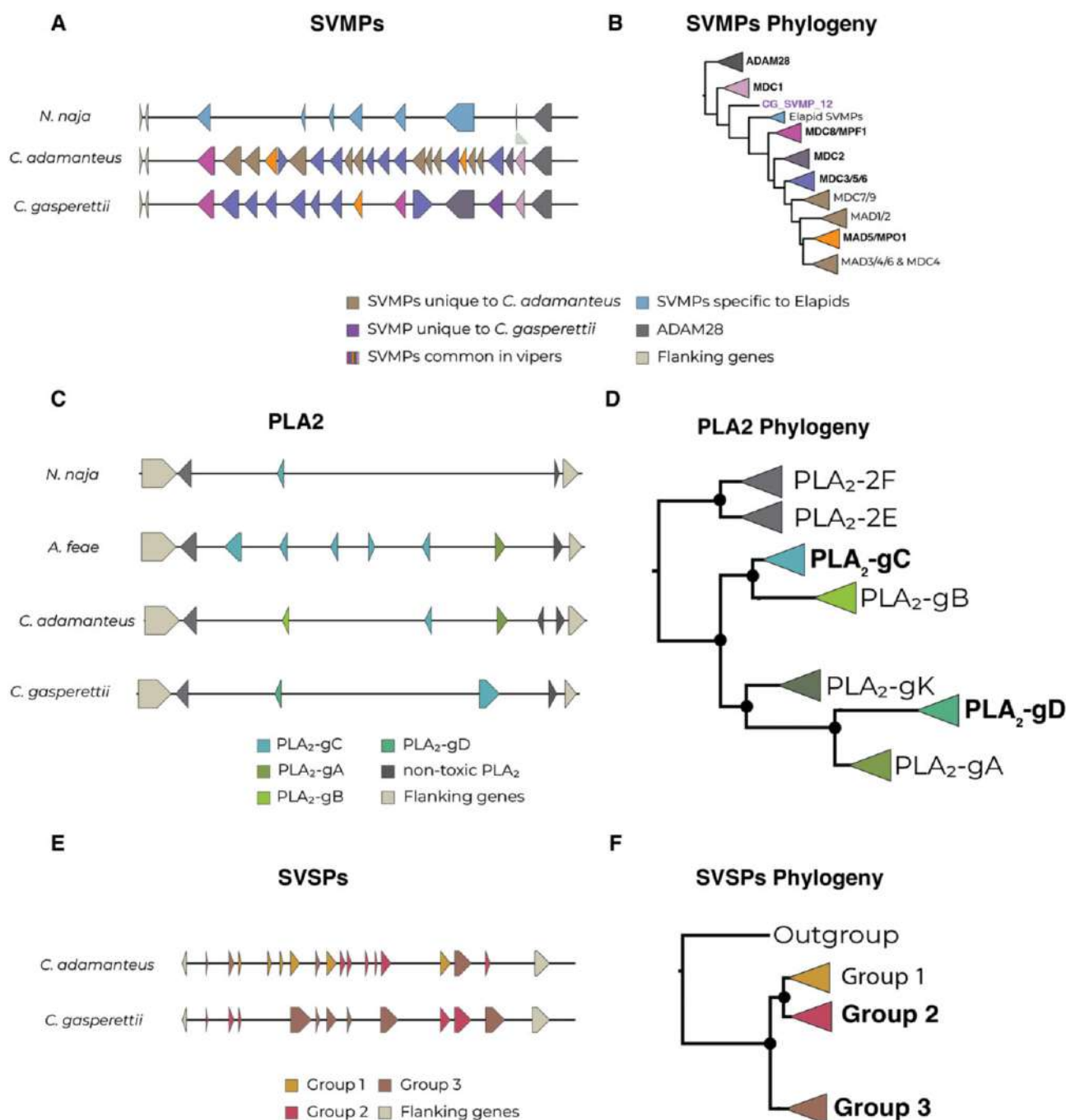


Figure 4: (A) Local synteny analyses for the SVMP toxin family in *N. naja*, *C. adamanteus*, and *C. gasperettii*. Different colors indicate orthologous genes unique to *C. gasperettii*, crotalids, true vipers, or elapids. ADAM28 (right) as well as flanking genes (left) are also indicated. (B) Phylogeny of SVMPs. Bold type indicates groups that contained SVMPs from *C. gasperettii*; purple indicates the gene is exclusively found in *C. gasperettii*. (C) Local synteny analyses for PLA₂ in *N. naja*, *A. feae*, *C. adamanteus*, and *C. gasperettii*. Nontoxic PLA₂ and flanking genes are also shown. (D) Phylogeny of the PLA₂ gene family, with 2 non-toxic PLA₂s as outgroups. Some samples that did not fit in any category have been removed. For a complete phylogeny see [Supplementary Fig. S9](#). Note that PLA₂-gK is present in the phylogeny but not in the local synteny analyses, as any of the studied species contains it. (E) Local synteny analyses for SVSPs for *C. adamanteus* and *C. gasperettii*. Flanking genes are also shown. (F) Phylogeny for SVSPs with a nontoxic outgroup. For the 3 different phylogenies the groups that contained toxins from *C. gasperettii* are highlighted in bold.

(Fig. 4C), flanked by *OTUD3* and *MUL1* nontoxic genes, as previously reported in other species [3, 12, 41]. The number of venomous PLA₂s in *C. gasperettii* was lower than in *A. feae* and *C. adamanteus*. Phylogenetic results for PLA₂ genes showed a fully supported clade containing both nontoxic PLA₂-g2E and PLA₂-g2F as outgroups (Fig. 4D and [Supplementary Fig. S9](#)). We also found all other PLA₂ groups reported in previous studies: PLA₂-

gC, PLA₂-gK, PLA₂-gB, PLA₂-gD, and PLA₂-gA [12, 41]. The 2 genes for our target species clustered in different groups (Fig. 4D and [Supplementary Fig. S9](#)). The first PLA₂ was a PLA₂-gD, which is a group of PLA₂s exclusively found in true vipers (subfamily Viperinae). The second one was a PLA₂-gC which is more ancestral as it is also found in other pitvipers and nonvenomous snakes such as pythons [12]. The genomic results are consistent with the

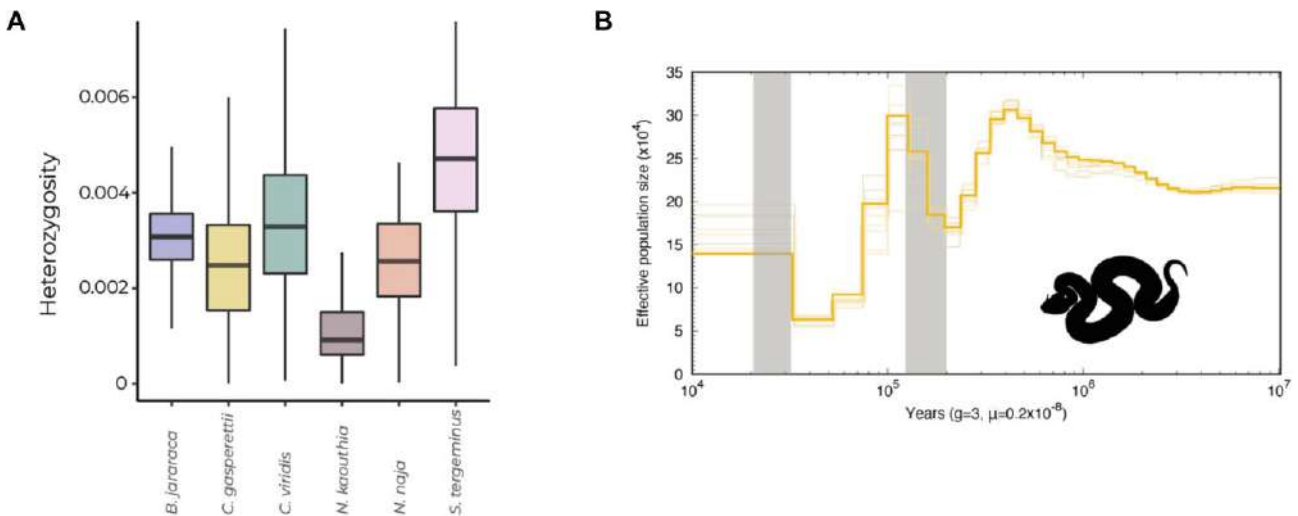


Figure 5: (A) Genome-wide diversity for 6 different venomous snakes: *B. jararaca*, *C. gasperettii*, *C. viridis*, *N. kaouthia*, *N. naja*, and *S. tergeminus*. (B) PSMC analysis recovering the ancient demographic history of *C. gasperettii*. Generation time was set to 3 years and the substitution rate to 2.4×10^{-9} per site per year. Shaded lines represent 10 bootstrap estimates. Two last glacial periods are shown with gray lines.

proteomics, indicating that specific duplications of PLA₂ toxin-coding genes have not occurred in *C. gasperettii*.

SVSPs

We found 8 different SVSPs within the genome of *C. gasperettii*, flanked by RBM42 and GRAMD1A nontoxic genes (Fig. 4E). For this toxin family, we were only able to compare the results with *C. adamanteus*. We were unable to confidently determine the location of SVSPs in the *N. naja* genome (several regions matched our venomous SVSP genes as well as the flanking genes). Moreover, *A. feae* was also not compared because SVSPs were not assembled in a single contig. Phylogenetic results showed 3 clades, with 2 containing *C. gasperettii* genes (Fig. 4F and Supplementary Fig. S10). Group 1 was mainly present within *Crotalus*, although there was the presence of some true vipers species, but not in *C. gasperettii* (Fig. S10). Group 2 contained 6 genes within *C. adamanteus* and only 2 for *C. gasperettii*. Interestingly, Group 3 was expanded in *C. gasperettii* (Fig. 4E) with a total of 6 copies, while 4 were found within *C. adamanteus*. Most of the toxins included in the analyses for true vipers were also found in Group 3 (Supplementary Fig. S10), indicating a possible expansion of this group of toxins in true vipers (or gene losses in pit vipers). Overall, our high-quality chromosome-level reference genome has shed light on the evolution of the main toxin-coding gene families, indicating a compelling correlation between the abundance of toxin-coding genes and the prevalence of these toxins in the venom of *C. gasperettii*.

Glacial periods drove population expansions of *C. gasperettii*

The Arabian horned viper (*C. gasperettii*) is a widespread species, categorized as “Least Concern” by the IUCN [109]. Genome-wide diversity was in line with its conservation status, as it showed similar heterozygosity levels compared to other venomous snakes (Fig. 5A). However, more individuals should be sampled along its distribution to verify that similar heterozygosity levels are found across its range. PSMC analyses showed several population expansions and contractions in the last 400 kya, whilst the effective population size of *C. gasperettii* remained relatively constant from 1 until 10 Mya (Fig. 5B). Interestingly, population expansions were

coincident with the last glacial and penultimate glacial periods (gray lines on Fig. 5B), with a large population increase during the penultimate glacial period (1.94–1.35 mya) (Fig. 5B). In fact, during glacial periods, the global sea level dropped around 150 m, exposing the floor and the sand to the wind, which promoted aridification in the Arabian Peninsula and potentially increased habitat suitability for the species [110, 111]. PSMC results may vary depending on the generation time as well as the mutational rate specified. The absence of species-specific data for this analyses may bias our results, although it is a general consensus in the literature when inferring demographic analyses such as ours in snakes (e.g., [5]).

Conclusions

Our high-quality chromosome-level reference genome for *C. gasperettii* showed that chromosomal architecture is highly conserved between Crotalinae and Viperinae subfamilies, and differs from elapid genomes by a small number of chromosomal rearrangements. We also found the genomic coordinates of the main toxin-encoding genes, highlighting gene duplication as the main driver in the evolution of SVMP and SVSP toxins. We identified a new SVMP toxin-coding gene, showcasing the importance of using high-quality reference genomes (combined with other -omic techniques) for thoroughly characterizing toxin-encoding genes. Finally, this is a new and important resource for a large clade with few reference genomes available. Future genomic studies focusing on Old World viper evolution will benefit greatly from this resource, which will help unveil the origin and diversification of venom and serve as an essential genomic tool for further venom studies on the subfamily Viperinae.

Additional Files

Supplementary Fig. S1. Distribution map for the studied species *C. gasperettii* with the location of our samples. Countries where the species is present are indicated. JO: Jordan; SA: Saudi Arabia; YE: Yemen; OM: Oman; UAE: United Arab Emirates; IQ: Iraq; IR: Iran; KU: Kuwait; QA: Qatar.

Supplementary Fig. S2. Drawing of an Arabian horned viper depicting all the tissues sampled for RNA-seq analyses.

Supplementary Fig. S3. Histogram from GenomeScope showing the frequency of reads in relation with their coverage.

Supplementary Fig. S4. Heatmap for the 2,000 most variable genes within our 3 samples, showing a clear batch effect of sample CG1 (possibly due to differences in sequencing time) as well as a high similarity between the putative accessory gland and the venom gland. Each column represents a different sampled tissue. The 3 different samples are depicted with different colors at the top of the heatmap. Abbreviations are as follows: G. bladder: gall bladder; V. gland: venom gland.

Supplementary Fig. S5. Heatmap for the 2,000 most variable genes for both samples, reporting highly expressed genes unique for each tissue type. Each column represents one tissue sampled per individual. Expression levels were normalized. Te: testis; Ov: ovary; G.b.: gallbladder; Sp: spleen; V.G.: venom gland.

Supplementary Fig. S6. Heatmap for the 161 upregulated genes found in the venom gland of *C. gasperettii* transcriptome including the 65 putative expressed toxins for both venom gland samples. Each column represents one tissue sampled per individual. VG: venom Gland; Ki: kidney; GB: gall bladder; Lu: lung; Sp: spleen; He: heart; Li: liver; Pa: pancreas; To: tongue; Te: Testis; Ov: ovary.

Supplementary Fig. S7. Heatmap for the venom gland transcriptome for the 65 putative expressed toxins for both venom gland samples. Each column represents one tissue sampled per individual. VG: venom Gland; Ki: kidney; GB: gall bladder; Lu: lung; Sp: spleen; He: heart; Li: liver; Pa: pancreas; To: tongue; Te: Testis; Ov: ovary.

Supplementary Fig. S8. Maximum-likelihood phylogeny for SVMF genes and its non-toxic paralog (ADAM28). Genes for *C. gasperettii* are highlighted in bold. Toxin groups are identified following previous categorizations. Asterisks indicate if *C. gasperettii* genes are present in that specific group. Branch support with aBayes values >90 are depicted as circles.

Supplementary Fig. S9. Maximum-likelihood phylogeny for PLA₂, with the 2 nontoxic genes as outgroups (PLA₂-2F and PLA₂-2E). Asterisks in group labels indicate if *C. gasperettii* genes are present in that specific group. Branch support with aBayes values >90 are depicted as circles.

Supplementary Fig. S10. Maximum-likelihood phylogeny for SVSPs, with 1 sample from *Thamnophis elegans* as outgroup. Asterisks in group labels indicate if *C. gasperettii* genes are present in that specific group. Branch support with aBayes values >90 are depicted as circles.

Supplementary Table S1. Individuals sampled in this study with their sex, sampling coordinates, and data sequenced.

Supplementary Table S2. Id, tissue type and number of reads sequenced per sample.

Supplementary Table S3. Different types of repetitive elements masked within the genome.

Supplementary Table S4. Abundances for the different toxin families identified in the proteome of *C. gasperettii*.

Acknowledgments

G.M.-R. is supported by an FPI grant from the Ministerio de Ciencia, Innovación y Universidades, Spain (PRE2019-088729); S.R.H. is supported by a National Science Foundation Graduate Research Fellowship Program (grant no. 2136515); A.T. is supported by “la Caixa” doctoral fellowship program (LCF/BQ/DR20/11790007); B.B.-C. is supported by an FPU grant from Ministerio de Cien-

cia, Innovación y Universidades, Spain (FPU18/04742); and M.E. is supported by an FPI grant from Ministerio de Ciencia e Innovación (PRE2022-101473). In the UAE, we wish to thank His Highness Sheikh Dr Sultan bin Mohammed Al Qasimi, Supreme Council Member and Ruler of Sharjah, H. E. Ms Hana Saif al Suwaidi (Chairperson of the Environment and Protected Areas Authority, Sharjah) for their continuous support. Some of this research was carried out on the High Performance Computing resources at New York University Abu Dhabi. We thank Jonathan Wood and Klara Eleftheriadi for their input during the genome assembly and manual curation processes. We also thank Valéria Marques for her help in building the figures and Prem Aguilar for reviewing a previous version of the manuscript.

Author's Contribution

Conceptualization: G.M.R., A.T., B.B.C., J.C., J.E., M.M., S.C. Investigation: S.H., V.P., M.E., T.B., S.B., M.H., J.T.G., D.P., J.C., M.M. Funding acquisition: S.C. Writing—original draft: G.M.R. Writing—review and editing: all authors read, revised, and approved the manuscript final version.

Funding

This work was funded by grant PID2021-128901NB-I00 (MCIN/AEI/10.13039/501100011033 and by ERDF, A way of making Europe; Spain) and grant 2021-SGR-00751 from the Departament de Recerca i Universitats from the Generalitat de Catalunya, Spain to SC.

Data Availability

Final assembly and raw reads files have been deposited in NCBI under bioproject no. PRJNA1068073. Proteomic data have been published at PRIDE [112, 113] under project accession numbers PXD060777 and PXD060783. All additional supporting data are available in the GigaScience repository, GigaDB [114].

Competing Interests

The authors declare that they have no competing interests.

Ethics statement

No in vivo experiments were performed. Specimens were collected and manipulated with the authorization and under strict control and permission of the government of the United Arab Emirates (Environment and Protected Areas Authority, Government of Sharjah), who approved the study. Specimens were captured and processed following the guidelines and protocols stated in the agreements obtained from the competent authority of the United Arab Emirates. Members of the government supervised collecting activities. All efforts were made to minimize animal suffering. All the research in the United Arab Emirates was done under the supervision and permission of the Environment and Protected Areas Authority, Government of Sharjah.

References

1. Dussex N, van der Valk T, Morales HE, et al. Population genomics of the critically endangered kākāpō. *Cell Genomics*. 2021;1(1):100002. <https://doi.org/10.1016/j.xgen.2021.100002>.
2. Hogan MP, Holding ML, Nystrom GS, et al. The genetic regulatory architecture and epigenomic basis for age-related

- changes in rattlesnake venom. *Proc Natl Acad Sci USA*. 2024;121(16):e2313440121. <https://doi.org/10.1073/pnas.2313440121>.
3. Margres MJ, Rautsaw RM, Strickland JL, et al. The tiger rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. *Proc Natl Acad Sci USA*. 2021;118(4):e2014634118. <https://doi.org/10.1073/pnas.2014634118>.
 4. Pardos-Blas JR, Irisarri I, Abalde S, et al. The genome of the venomous snail *lautoconus ventricosus* sheds light on the origin of conotoxin diversity. *GigaScience*. 2021;10(5):giab037. <https://doi.org/10.1093/gigascience/giab037>.
 5. Schield DR, Card DC, Hales NR, et al. The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome Res*. 2019;29(4):590–601. <https://doi.org/10.1101/gr.240952.118>.
 6. Suryamohan K, Krishnankutty SP, Guillory J, et al. The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. *Nat Genet*. 2020;52(1):106–17. <https://doi.org/10.1038/s41588-019-0559-8>.
 7. Drukewitz SH, Von Reumont BM. The significance of comparative genomics in modern evolutionary venomomics. *Front Ecol Evolution*. 2019;7:163. <https://doi.org/10.3389/fevo.2019.00163>.
 8. Frantz LAF, Bradley DG, Larson G, et al. Animal domestication in the era of ancient genomics. *Nat Rev Genet*. 2020;21(8):449. <https://doi.org/10.1038/s41576-020-0225-0>.
 9. Orteu A, Jiggins CD. The genomics of coloration provides insights into adaptive evolution. *Nat Rev Genet*. 2020;21(8):461. <https://doi.org/10.1038/s41576-020-0234-z>.
 10. San-Jose LM, Roulin A. Genomics of coloration in natural animal populations. *Phil Trans R Soc B*. 2017;372(1724):20160337. <https://doi.org/10.1098/rstb.2016.0337>.
 11. Casewell NR, Wüster W, Vonk FJ, et al. Complex cock-tails: the evolutionary novelty of venoms. *Trends Ecol Evol*. 2013;28(4):219–29. <https://doi.org/10.1016/j.tree.2012.10.020>.
 12. Dowell NL, Giorgianni MW, Kassner VA, et al. The deep origin and recent loss of venom toxin genes in rattlesnakes. *Curr Biol*. 2016;26(18):2434–45. <https://doi.org/10.1016/j.cub.2016.07.038>.
 13. Giorgianni MW, Dowell NL, Griffin S, et al. The origin and diversification of a novel protein family in venomous snakes. *Proc Natl Acad Sci USA*. 2020;117(20):10911–10920. <https://doi.org/10.1073/pnas.1920011117>.
 14. Werren JH, Richards S, Desjardins CA, et al. The Nasonia Genome Working Group. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science*. 2010;327(5963):343–48. <https://doi.org/10.1126/science.1178028>.
 15. Fry BG, Roelants K, Champagne DE, et al. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu Rev Genom Hum Genet*. 2009;10:483–511. <https://doi.org/10.1146/annurev.genom.9.081307.164356>.
 16. Zancolli G, Reijnders M, Waterhouse RM, et al. Convergent evolution of venom gland transcriptomes across Metazoa. *Proc Natl Acad Sci USA*. 2022;119(1):e2111392119. <https://doi.org/10.1073/pnas.2111392119>.
 17. Vonk FJ, Casewell NR, Henkel CV, et al. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc Natl Acad Sci USA*. 2013;110(51):20651–56. <https://doi.org/10.1073/pnas.1314702110>.
 18. Avella I, Calvete JJ, Sanz L, et al. Interpopulational variation and ontogenetic shift in the venom composition of Lataste's viper (*Vipera latastei*, Boscá 1878) from northern Portugal. *J Proteomics*. 2022;263:104613. <https://doi.org/10.1016/j.jprot.2022.104613>.
 19. Margres MJ, Wray KP, Sanader D, et al. Varying intensities of introgression obscure incipient venom-associated speciation in the timber rattlesnake (*Crotalus horridus*). *Toxins*. 2021;13(11):782. <https://doi.org/10.3390/toxins13110782>.
 20. King GF. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert Opin Biol Ther*. 2011;11(11):1469–1484. <https://doi.org/10.1517/14712598.2011.621940>.
 21. Li L, Huang J, Lin Y. Snake venoms in cancer therapy: past, present and future. *Toxins*. 2018;10(9):346. <https://doi.org/10.3390/toxins10090346>.
 22. Vyas VK, Brahmabhatt K, Bhatt H, et al. Therapeutic potential of snake venom in cancer therapy: current perspectives. *Asia Pac J Trop Biomedicine*. 2013;3(2):156–62. [https://doi.org/10.1016/S2221-1691\(13\)60042-8](https://doi.org/10.1016/S2221-1691(13)60042-8).
 23. Williams DJ, Faiz MA, Abela-Ridder B, et al. Strategy for a globally coordinated response to a priority neglected tropical disease: snakebite envenoming. *PLoS Negl Trop Dis*. 2019;13(2):e0007059. <https://doi.org/10.1371/journal.pntd.0007059>.
 24. Uetz P. The Reptile Database: curating the biodiversity literature without funding. *BISS*. 2021;5:e75448. <https://doi.org/10.3897/biss.5.75448>.
 25. Fry BG, Wüster W. Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. *Mol Biol Evol*. 2004;21(5):870–83. <https://doi.org/10.1093/molbev/msh091>.
 26. Gutiérrez JM, Warrell DA, Williams DJ, et al. Global Snakebite Initiative. The need for full integration of snakebite envenoming within a global strategy to combat the neglected tropical diseases: the way forward. *PLoS Negl Trop Dis*. 2013;7(6):e2162. <https://doi.org/10.1371/journal.pntd.0002162>.
 27. Gutiérrez JM, Calvete JJ, Habib AG, et al. Snakebite envenoming. *Nat Rev Dis Primers*. 2017;3(1):Article 1. <https://doi.org/10.1038/nrdp.2017.63>.
 28. Tasoulis T, Isbister G. A review and database of snake venom proteomes. *Toxins*. 2017;9(9):290. <https://doi.org/10.3390/toxins9090290>.
 29. Weinstein SA, White J, Keyler DE, et al. Non-front-fanged colubrid snakes: a current evidence-based analysis of medical significance. *Toxicon*. 2013;69:103–13. <https://doi.org/10.1016/j.toxicon.2013.02.003>.
 30. Ferraz CR, Arrahman A, Xie C, et al. Multifunctional toxins in snake venoms and therapeutic implications: from pain to hemorrhage and necrosis. *Front Ecol Evol*. 2019;7. <https://doi.org/10.3389/fevo.2019.00218>.
 31. Fry B. (Ed.). *Venomous Reptiles and Their Toxins: Evolution, Pathophysiology, and Biodiscovery*. Oxford University Press, 2015. ISBN: 9780199309399.
 32. Fry BG, Scheib H, van der Weerd L, et al. Evolution of an arsenal: structural and functional diversification of the venom system in the advanced snakes (Caenophidia). *Mol Cell Proteomics*. 2008;7(2):215–46. <https://doi.org/10.1074/mcp.M700094-MCP200>.
 33. Osipov A, Utkin Y. What are the neurotoxins in hemotoxic snake venoms? *Int J Mol Sci*. 2023;24(3):2919. <https://doi.org/10.3390/ijms24032919>.
 34. Vitt LJ, Caldwell JP. *Herpetology: An Introductory Biology of Amphibians and Reptiles* (Fourth edition). Elsevier, AP,

- Academic Press is an imprint of Elsevier, 2014. <https://doi.org/10.1016/C2010-0-67152-5>
35. Arnold NE, Robinson MD, Carranza S. A preliminary analysis of phylogenetic relationships and biogeography of the dangerously venomous carpet vipers, *Echis* (Squamata, Serpentes, Viperidae) based on mitochondrial DNA sequences. *Amphib Reptilia*. 2009;30(2):273–82. <https://doi.org/10.1163/156853809788201090>.
 36. Casewell NR, Harrison RA, Wüster W, et al. Comparative venom gland transcriptome surveys of the saw-scaled vipers (Viperidae: *echis*) reveal substantial intra-family gene diversity and novel venom transcripts. *BMC Genomics* [Electronic Resource]. 2009;10(1):564. <https://doi.org/10.1186/1471-2164-10-564>.
 37. Pook CE, Joger U, Stümpel N, et al. When continents collide: phylogeny, historical biogeography and systematics of the medically important viper genus *Echis* (Squamata: Serpentes: Viperidae). *Mol Phylogenet Evol*. 2009;53(3):792–807. <https://doi.org/10.1016/j.ympev.2009.08.002>.
 38. Šmíd J, Tolley KA. Calibrating the tree of vipers under the fossilized birth-death model. *Sci Rep*. 2019;9(1):5510. <https://doi.org/10.1038/s41598-019-41290-2>.
 39. Wüster W, Peppin L, Pook CE, et al. A nesting of vipers: phylogeny and historical biogeography of the Viperidae (Squamata: Serpentes). *Mol Phylogenet Evol*. 2008;49(2):445–59. <http://doi.org/10.1016/j.ympev.2008.08.019>.
 40. Almeida DD, Viala VL, Nachtigall PG, et al. Tracking the recruitment and evolution of snake toxins using the evolutionary context provided by the *Bothrops jararaca* genome. *Proc Natl Acad Sci USA*. 2021;118(20):e2015159118. <https://doi.org/10.1073/pnas.2015159118>.
 41. Myers EA, Strickland JL, Rautsaw RM, et al. De novo genome assembly highlights the role of lineage-specific gene duplications in the evolution of venom in Fea's viper (*Azemiops feae*). *Genome Biol Evolut*. 2022;14(7):evac082. <https://doi.org/10.1093/gbe/evac082>.
 42. Hirst SR, Rautsaw RM, VanHorn CM, et al. Where the “ruber” meets the road: using the genome of the red diamond rattlesnake to unravel the evolutionary processes driving venom evolution. *Genome Biol Evolut*. 2024;16(9):evae198. <https://doi.org/10.1093/gbe/evae198>.
 43. Gilbert C, Meik JM, Dashevsky D, et al. Endogenous hepadnaviruses, bornaviruses and circoviruses in snakes. *Proc R Soc B*. 2014;281(1791):20141122. <https://doi.org/10.1098/rspb.2014.1122>.
 44. Westeen EP, Escalona M, Holding ML, et al. A genome assembly for the southern Pacific rattlesnake, *Crotalus oreganus helleri*, in the western rattlesnake species complex. *J Hered*. 2023;114(6):681–89. <https://doi.org/10.1093/jhered/esad045>.
 45. Saethang T, Somporn P, Payungporn S, et al. Identification of *Daboia siamensis* venom using integrated multi-omics data. *Sci Rep*. 2022;12(1):Article 1. <https://doi.org/10.1038/s41598-022-17300-1>.
 46. Talavera A, Palmada-Flores M, Martínez-Freiría F, et al. Unveiling the evolutionary history of European vipers and their venoms from a multi-omic approach. *bioRxiv*. 2024; 2024–2012. <https://doi.org/10.1101/2024.12.10.627732>.
 47. Ali SA, Jackson TNW, Casewell NR, et al. Extreme venom variation in Middle Eastern vipers: a proteomics comparison of *Eristicophis macmahonii*, *Pseudocerastes fieldi* and *Pseudocerastes persicus*. *J Proteomics*. 2015;116(26):106–13. <https://doi.org/10.1016/j.jprot.2014.09.003>.
 48. Mackessy SP. Evolutionary trends in venom composition in the western rattlesnakes (*Crotalus viridis sensu lato*): toxicity vs. tenderizers. *Toxicon*. 2010;55(8):1463–1474. <https://doi.org/10.1016/j.toxicon.2010.02.028>.
 49. Jan V, Maroun RC, Robbe-Vincent A, et al. Toxicity evolution of *Vipera aspis aspis* venom: identification and molecular modeling of a novel phospholipase A2 heterodimer neurotoxin. *FEBS Lett*. 2002;527(1):263–68. [https://doi.org/10.1016/S0014-5793\(02\)03205-2](https://doi.org/10.1016/S0014-5793(02)03205-2).
 50. Smith CF, Nikolakis ZL, Perry BW, et al. The best of both worlds? Rattlesnake hybrid zones generate complex combinations of divergent venom phenotypes that retain high toxicity. *Biochimie*. 2023;213:176–89. <https://doi.org/10.1016/j.biochi.2023.07.008>.
 51. Carranza S, Els J, Burriel-Carranza B. A Field Guide to the Reptiles of Oman. Madrid: Consejo Superior de Investigaciones Científicas. 2021. ISBN:978-84-00-10877-9
 52. Mochales-Riaño G, Burriel-Carranza B, Barros MI, et al. Hidden in the sand: phylogenomics unravel an unexpected evolutionary history for the desert-adapted vipers of the genus *Cerastes*. *Mol Phylogenet Evol*. 2024;191:107979. <https://doi.org/10.1016/j.ympev.2023.107979>.
 53. Russell FE, Campbell JR. Venomous Terrestrial Snakes of the Middle East. 2015. Edition Chimaira, ISBN:978-3-89973-446-1.
 54. Al-Sadoon MK, Paray BA. Ecological aspects of the horned viper, *Cerastes cerastes gasperettii* in the central region of Saudi Arabia. *Saudi J Biol Sci*. 2016;23(1):135–38. <https://doi.org/10.1016/j.sjbs.2015.10.010>.
 55. Amr ZS, Abu Baker MA, Warrell DA. Terrestrial venomous snakes and snakebites in the Arab countries of the Middle East. *Toxicon*. 2020;177:1–15. <https://doi.org/10.1016/j.toxicon.2020.01.012>.
 56. Schneemann M, Cathomas R, Laidlaw ST, et al. Life-threatening envenoming by the Saharan horned viper (*Cerastes cerastes*) causing micro-angiopathic haemolysis, coagulopathy and acute renal failure: clinical cases and review. *Quart J Med*. 2004;97(11):717–27. <https://doi.org/10.1093/qjmed/hch118>.
 57. Rokytá DR, Margres MJ, Ward MJ, et al. The genetics of venom ontogeny in the eastern diamondback rattlesnake (*Crotalus adamanteus*). *PeerJ*. 2017;5:e3249. <https://doi.org/10.7717/peerj.3249>.
 58. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequencing Data. 2010. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
 59. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17(1):10. <https://doi.org/10.14806/ej.17.1.200>.
 60. Rhie A, Walenz BP, Koren S, et al. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020;21(1):245. <https://doi.org/10.1186/s13059-020-02134-9>.
 61. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11(1):Article 1. <https://doi.org/10.1038/s41467-020-14998-3>.
 62. Rhie A, McCarthy SA, Fedrigo O, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 2021;592(2021):737–46. <https://doi.org/10.1038/s41586-021-03451-0>.
 63. Cheng H, Concepcion GT, Feng X, et al. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021;18(2):170. <https://doi.org/10.1038/s41592-020-01056-5>.
 64. Guan D, McCarthy SA, Wood J, et al. Identifying and removing haplotypic duplication in primary genome assemblies.

- Bioinformatics. 2020;36(9):2896–98. <https://doi.org/10.1093/bioinformatics/btaa025>.
65. Ghurye J, Rhie A, Walenz BP, et al. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput Biol*. 2019;15(8):e1007273. <https://doi.org/10.1371/journal.pcbi.1007273>.
 66. Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;9(11):e112963. <https://doi.org/10.1371/journal.pone.0112963>.
 67. Jin J-J, Yu W-B, Yang J-B, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020;21(1):241. <https://doi.org/10.1186/s13059-020-02154-5>.
 68. Gurevich A, Saveliev V, Vyahhi N, et al. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072–75. <https://doi.org/10.1093/bioinformatics/btt086>.
 69. Formenti G, Abueg L, Brajuka A, et al. Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics*. 2022;38(17):4214–16. <https://doi.org/10.1093/bioinformatics/btac460>.
 70. Challis R, Richards E, Rajan J, et al. BlobToolKit—interactive quality assessment of genome assemblies. *G3 (Bethesda)*. 2020;10(4):1361–1374. <https://doi.org/10.1534/g3.119.400908>.
 71. Allio R, Schomaker-Bastos A, Romiguier J, et al. MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol Ecol Resour*. 2020;20(4):892–905. <https://doi.org/10.1111/1755-0998.13160>.
 72. Li D, Luo R, Liu C-M, et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*. 2016;102:3–11. <https://doi.org/10.1016/j.ymeth.2016.02.020>.
 73. Flynn JM, Hubley R, Goubert C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA*. 2020;117(17):9451–57. <https://doi.org/10.1073/pnas.1921046117>.
 74. Tempel S. Using and understanding RepeatMasker. In: Bigot Y. (Ed.), *Mobile Genetic Elements* (Vol. 859, pp. 29–51). Humana Press, 2012. https://doi.org/10.1007/978-1-61779-603-6_2.
 75. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*. 2015;6(1):11. <https://doi.org/10.1186/s13100-015-0041-9>.
 76. Keilwagen J, Hartung F, Grau J. GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. In: Kollmar M. (Ed.), *Gene Prediction* (Vol. 1962, pp. 161–77). Springer, New York, 2019. https://doi.org/10.1007/978-1-4939-9173-0_9.
 77. Chen S, Zhou Y, Chen Y, et al. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
 78. Kim D, Paggi JM, Park C, et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907–15. <https://doi.org/10.1038/s41587-019-0201-4>.
 79. Tang S, Lomsadze A, Borodovsky M. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res*. 2015;43(12):e78–e78. <https://doi.org/10.1093/nar/gkv227>.
 80. Gabriel L, Hoff KJ, Bruna T, et al. TSEBRA: transcript selector for BRAKER. *BMC Bioinf*. 2021;22(1):566. <https://doi.org/10.1186/s12859-021-04482-0>.
 81. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236–40. <https://doi.org/10.1093/bioinformatics/btu031>.
 82. Dainat J, Hereñú D, Murray KDDr, et al. NBISweden/AGAT: AGAT-v1.2.0 (v1.2.0) [Computer software]. Zenodo. 2023. <https://doi.org/10.5281/ZENODO.3552717>.
 83. Solovyev V, Kosarev P, Seledsov I, et al. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol*. 2006;7(Suppl 1):S10. <https://doi.org/10.1186/gb-2006-7-s1-s10>.
 84. Geneva AJ, Park S, Bock DG, et al. Chromosome-scale genome assembly of the brown anole (*Anolis sagrei*), an emerging model species. *Commun Biol*. 2022;5(1):1126. <https://doi.org/10.1038/s42003-022-04074-5>.
 85. Tang H, Bowers JE, Wang X, et al. Synteny and collinearity in plant genomes. *Science*. 2008;320(5875):486–88. <https://doi.org/10.1126/science.1153917>.
 86. Kiełbasa SM, Wan R, Sato K, et al. Adaptive seeds tame genomic sequence comparison. *Genome Res*. 2011;21(3):487–93. <https://doi.org/10.1101/gr.113985.110>.
 87. Tang H, Krishnakumar V, Jingping L, et al. tanghaibao / jvci: JCVI v0.7.5 (v0.7.5) [Computer software]. Zenodo. 2017. <https://doi.org/10.5281/ZENODO.846919>.
 88. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10(2):giab008. <https://doi.org/10.1093/gigascience/giab008>.
 89. Pertea M, Pertea GM, Antonescu CM, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290–95. <https://doi.org/10.1038/nbt.3122>.
 90. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>.
 91. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, 2021. <http://www.R-project.org/>.
 92. Calvete JJ, Pla D, Els J, et al. Combined molecular and elemental mass spectrometry approaches for absolute quantification of proteomes: application to the venomomics characterization of the two species of desert black cobras, *Walterinnesia aegyptia* and *Walterinnesia morgani*. *J Proteome Res*. 2021;20(11):5064–78. <https://doi.org/10.1021/acs.jproteome.1c00608>.
 93. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80. <https://doi.org/10.1093/molbev/mst010>.
 94. Hackl T, Ankenbrand M, van Adrichem B, et al. Gggenomes: effective and versatile visualizations for comparative genomics. *arXiv*. 2024; <https://doi.org/10.48550/arXiv.2411.13556>.
 95. Guindon S, Dufayard J-F, Lefort V, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59(3):307–21. <https://doi.org/10.1093/sysbio/syq010>.
 96. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature*. 2011;475(7357):493–96. <https://doi.org/10.1038/nature10231>.
 97. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–79. <https://doi.org/10.1093/bioinformatics/btp352>.
 98. Green RE, Braun EL, Armstrong J, et al. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science*. 2014;346(6215):1254449. <https://doi.org/10.1126/science.1254449>.

99. Schield DR, Perry BW, Adams RH, et al. The roles of balancing selection and recombination in the evolution of rattlesnake venom. *Nat Ecol Evol.* 2022;6(9):1367–80. <https://doi.org/10.1038/s41559-022-01829-5>.
100. Thongchum R, Singchat W, Laopichienpong N, et al. Diversity of PBI-DdeI satellite DNA in snakes correlates with rapid independent evolution and different functional roles. *Sci Rep.* 2019;9(1):15459. <https://doi.org/10.1038/s41598-019-51863-w>.
101. Bylsma R, Walkup DK, Hibbitts TJ, et al. Population genetic and genomic analyses of Western massasauga (*Sistrurus tergeminus* ssp.): implications for subspecies delimitation and conservation. *Conserv Genet.* 2022;23(2):271–83. <https://doi.org/10.1007/s10592-021-01420-8>.
102. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.* 2013; <https://doi.org/10.48550/arXiv.1303.3997>.
103. Broad Institute. Picard Tools. Broad Institute, 2021, GitHub Repository.
104. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–1303. <http://doi.org/10.1101/gr.107524.110>.
105. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag, 2016. <https://doi.org/10.1007/978-0-387-98141-3>.
106. Simão FA, Waterhouse RM, Ioannidis P, et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–12. <https://doi.org/10.1093/bioinformatics/btv351>.
107. Kalita B, Mackessy SP, Mukherjee AK. Proteomic analysis reveals geographic variation in venom composition of Russell's viper in the Indian subcontinent: implications for clinical manifestations post-enuvenomation and antivenom treatment. *Exp Rev Proteomics.* 2018;15(10):837–49. <https://doi.org/10.1080/14789450.2018.1528150>.
108. Casewell NR, Wagstaff SC, Wüster W, et al. Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. *Proc Natl Acad Sci USA.* 2014;111(25):9205–10. <https://doi.org/10.1073/pnas.1405484111>.
109. Egan D, Amr Z, Al Johany A, et al. *The IUCN Red List of Threatened Species: Cerastes gasperettii*, 2012. <https://doi.org/10.2305/IUCN.UK.2012.RLTS.T164599A1060588.en>.
110. Burriel-Carranza B, Tejero-Cicuéndez H, Carné A, et al. The origin of a mountain biota: hyper-aridity shaped reptile diversity in an Arabian biodiversity hotspot. *bioRxiv.* 2023; <https://doi.org/10.1101/2023.04.07.536010>.
111. Glennie KW, Singhvi AK. Event stratigraphy, paleoenvironment and chronology of SE Arabian deserts. *Quat Sci Rev.* 2002;21(7):853–69. [https://doi.org/10.1016/S0277-3791\(01\)00133-0](https://doi.org/10.1016/S0277-3791(01)00133-0).
112. Perez-Riverol Y, Bandla C, Kundu DJ, et al. The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res.* 2025;53(D1):D543–53. <https://doi.org/10.1093/nar/gkae1011>.
113. Deutsch EW, Bandeira N, Perez-Riverol Y, et al. The ProteomeXchange Consortium at 10 years: 2023 update. *Nucleic Acids Res.* 2023;51:D1539–48. <https://doi.org/10.1093/nar/gkac1040>.
114. Mochales-Riaño G, Hirst SR, Talavera A, et al. Supporting data for “chromosome-level reference genome for the medically important Arabian horned viper (*Cerastes gasperettii*)”. *GigaScience Database.* 2025. <https://doi.org/10.5524/102647>.

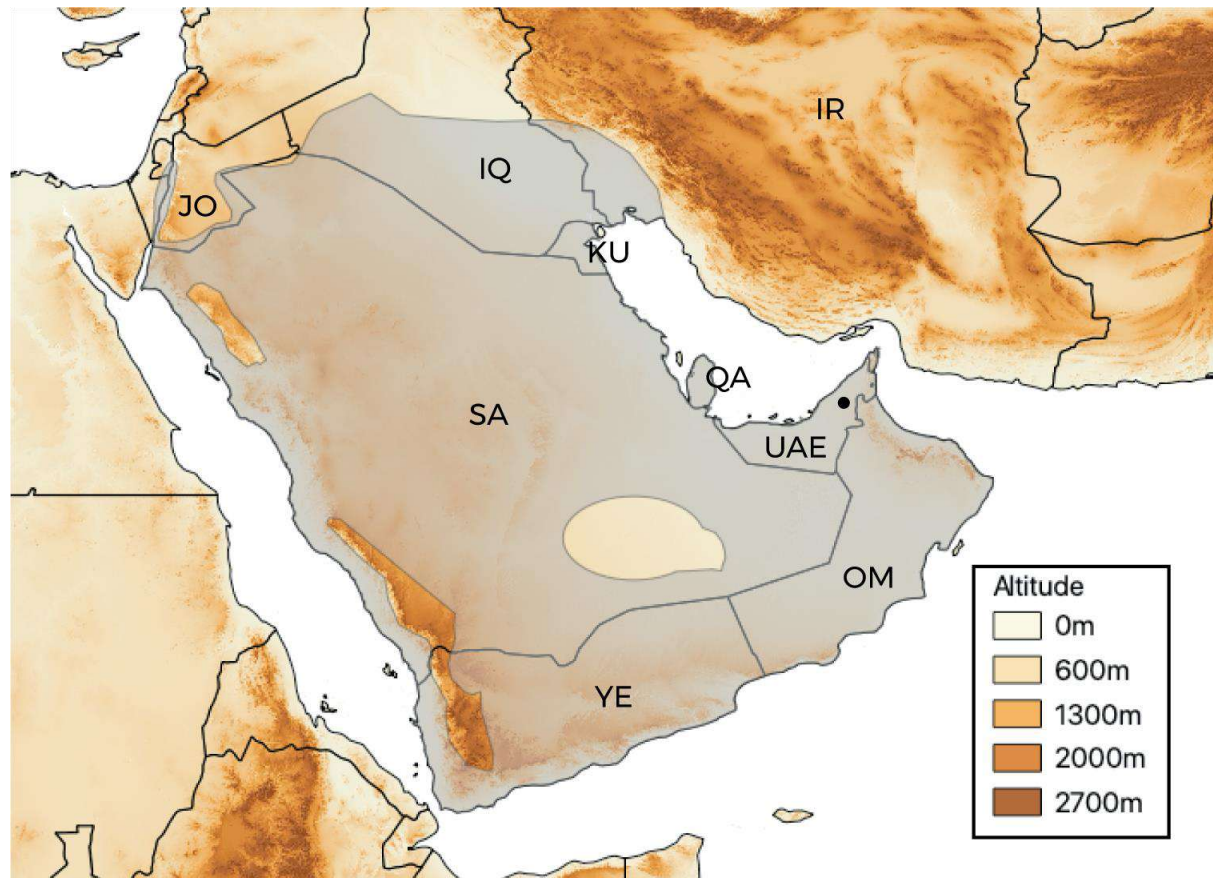
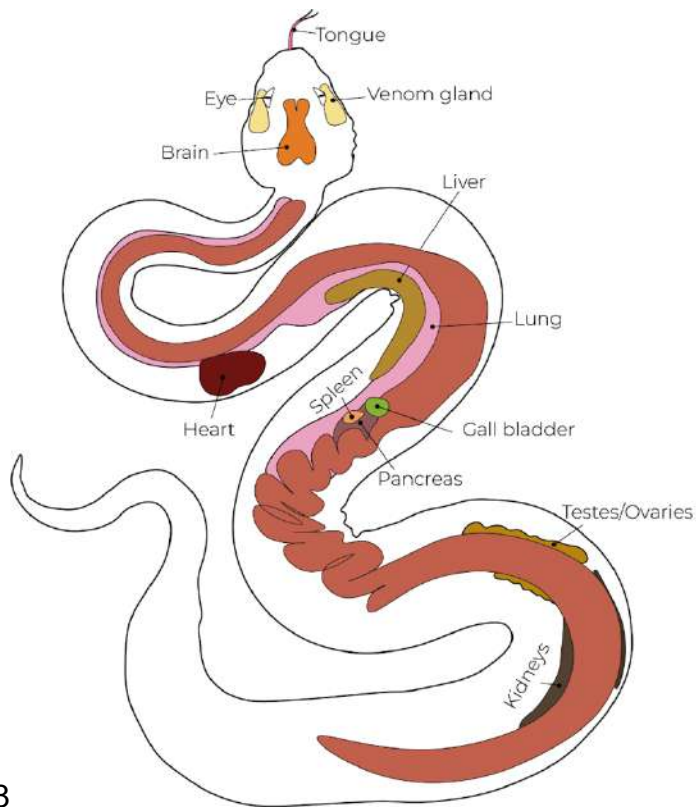


Fig. S1: Distribution map for the studied species *Cerastes gasperettii* with the location of our samples. Countries where the species is present are indicated. Abbreviations are as follows: JO, Jordania; SA, Saudi Arabia; YE, Yemen; OM, Oman; UAE, United Arab Emirates; IQ, Iraq; IR, Iran; KU, Kuwait, QA, Qatar.



668

669

670 Fig. S2: Drawing of an Arabian horned viper depicting all the tissues sampled for RNA-seq analyses.

671

672

673

674

675

676

GenomeScope Profile

len:1,392,502,372bp uniq:68.8%
aa:99% ab:0.984%
kcov:20.6 err:0.146% dup:0.799 k:21 p:2

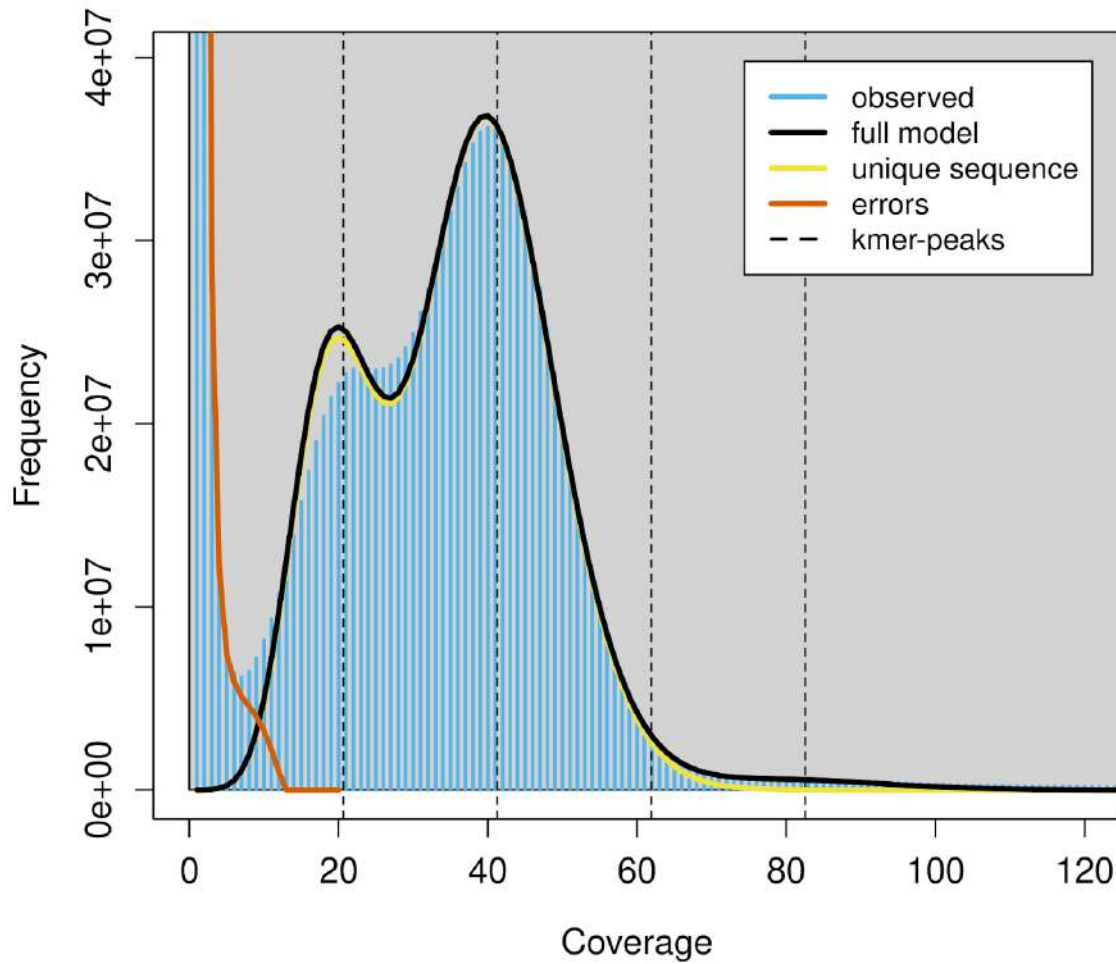


Fig. S3: Histogram from GenomeScope showing the frequency of reads in relation with their coverage.

Top 2,000 most variable genes

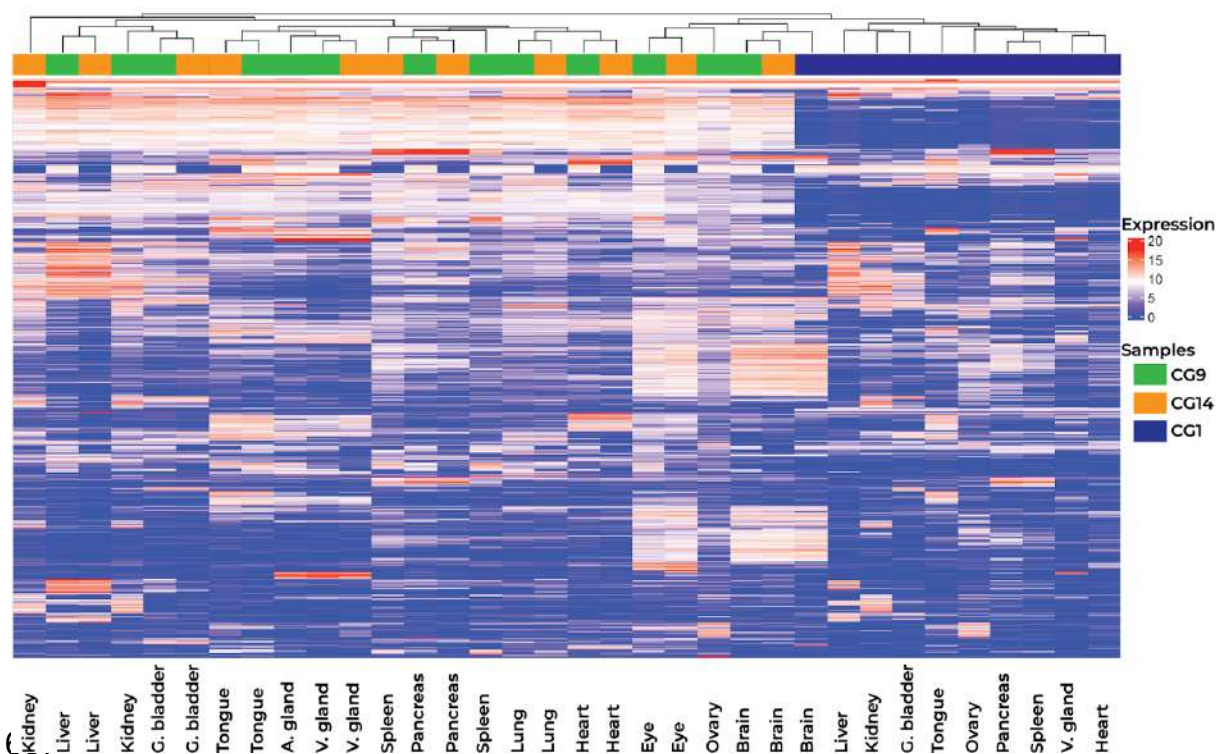


Fig. S4: Heatmap for the 2,000 most variable genes within our three samples, showing a clear batch effect of sample CG1 (possibly due to differences in sequencing time) as well as a high similarity between the putative accessory gland and the venom gland. Each column represents a different sampled tissue. The three different samples are depicted with different colors at the top of the heatmap. Abbreviations are as follows: G. bladder, gallbladder and V. gland, venom gland.

Top 2,000 most variable genes

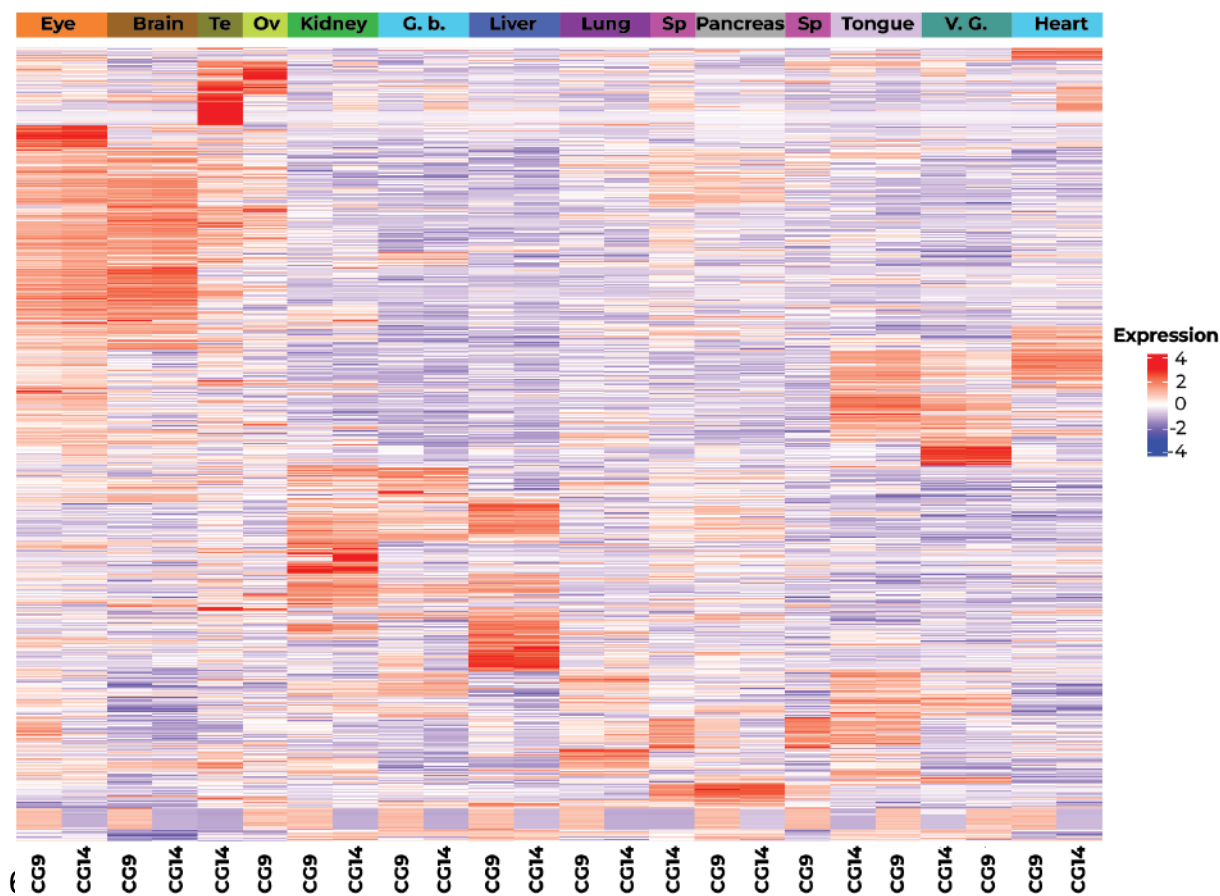
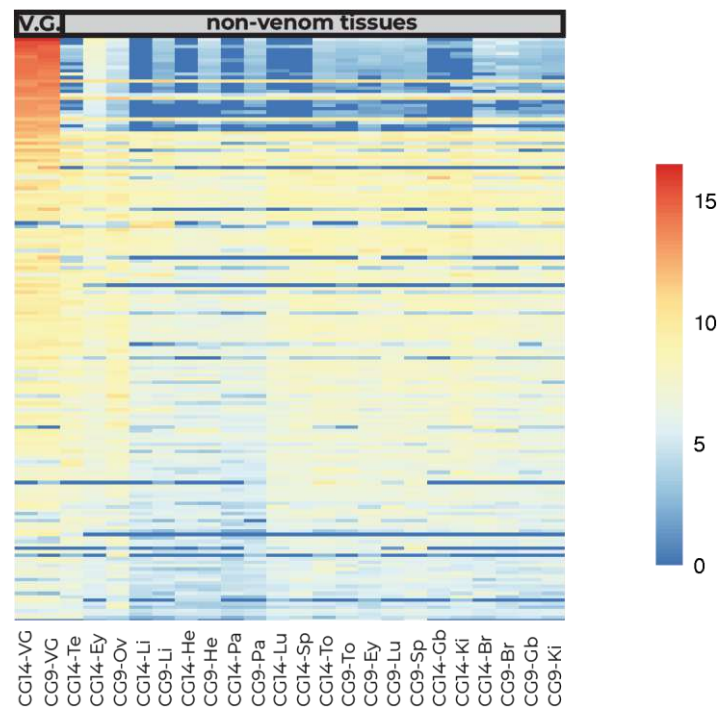


Fig. S5: Heatmap for the 2,000 most variable genes for both samples, reporting highly expressed genes unique for each tissue type. Each column represents one tissue sampled per individual. Expression levels were normalized. Abbreviations are as follows: Te, Testis; Ov, Ovary; G.b., gallbladder; Sp, Spleen and V.G., Venom gland.



702

703 Fig. S6: Heatmap for the 161 upregulated genes found in the venom gland of *C. gasperettii*
 704 transcriptome including the 65 putative expressed toxins for both venom gland samples. Each column
 705 represents one tissue sampled per individual. Abbreviations are as follows: VG, Venom Gland; Ki,
 706 Kidney; GB, Gall Bladder; Lu; Lung; Sp, Spleen; He, Heart; Li, Liver; Pa, Pancreas; To, Tongue; Te,
 707 Testis; Ov, Ovary.

708

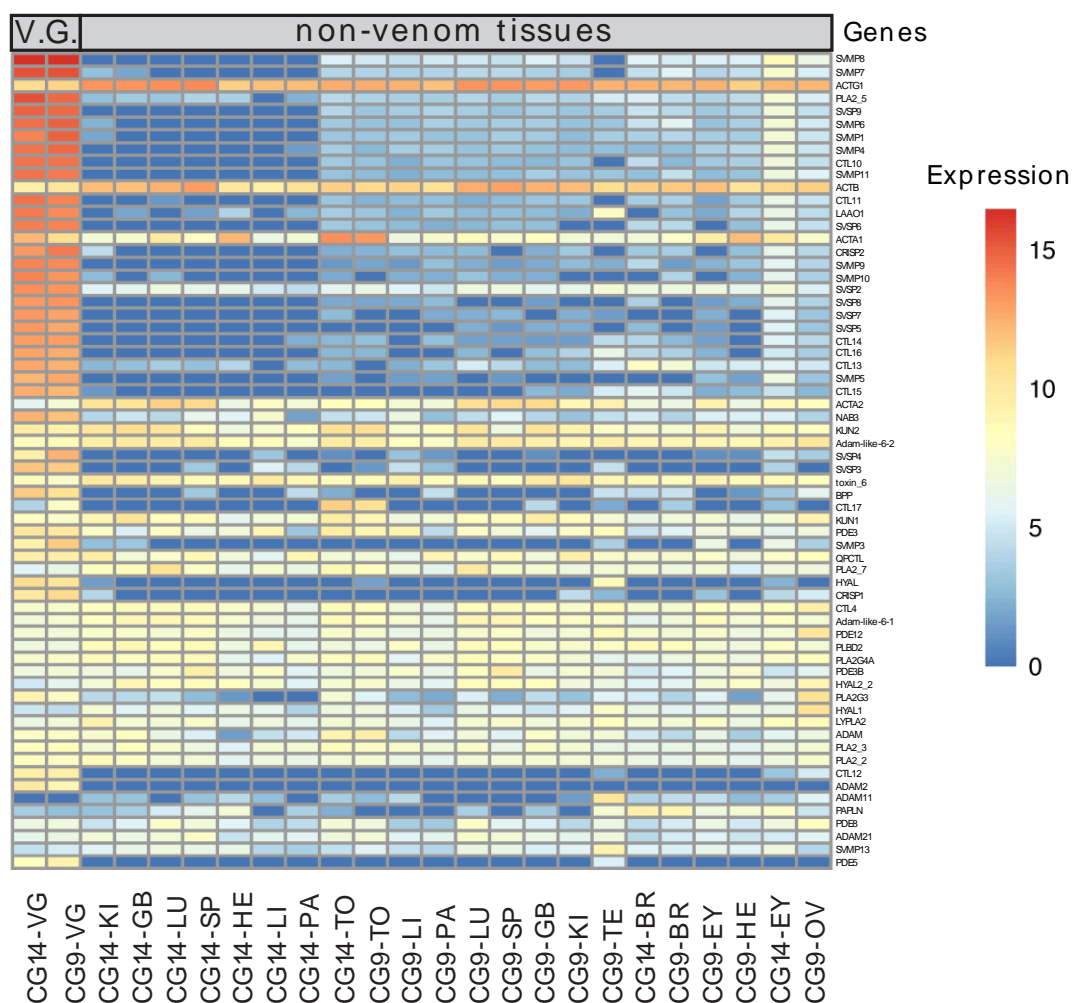
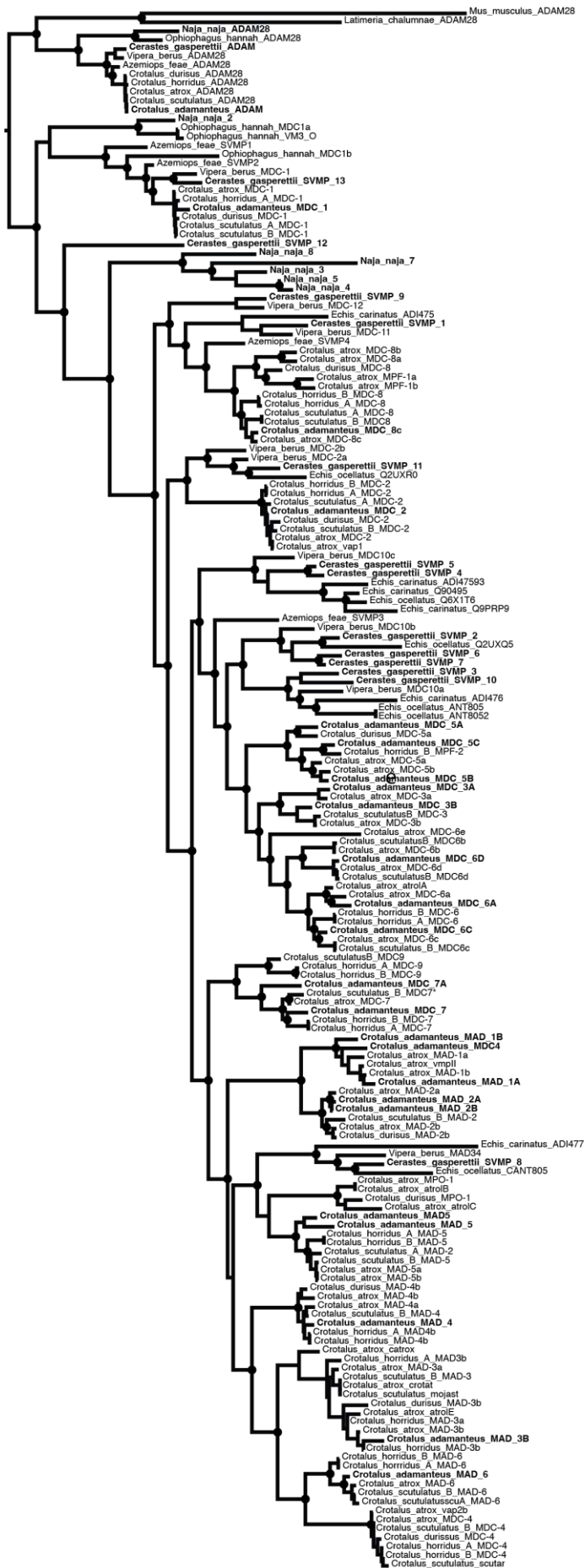


Fig. S7: Heatmap for the venom gland transcriptome for the 65 putative expressed toxins for both venom gland samples. Each column represents one tissue sampled per individual. Abbreviations are as follows: VG, Venom Gland; Ki, Kidney; GB, Gall Bladder; Lu; Lung; Sp, Spleen; He, Heart; Li, Liver; Pa, Pancreas; To, Tongue; Te, Testis; Ov, Ovary.



ADAM28*

MDC-1*

SVMP-CG-12*

Elapid SVMPs

MDC-8b & MPF-1b*

MDC-2*

MDC-3/5/6*

MDC-7/9

MAD-1/2

MAD-4/5*

MAD-3/4/6 & MDC-4

718 likelihood phylogeny for SVMP genes and its non-toxic paralog (ADAM28). Genes for *Cerastes*
719 *gasperettii* are highlighted in bold. Toxin groups are identified following previous categorizations.
720 Asterisks indicate if *Cerastes gasperettii* genes are present in that specific group. Branch support with
721 aBayes values higher than 90 are depicted as circles.

722

723

724

725

726

727

728

729

730

731

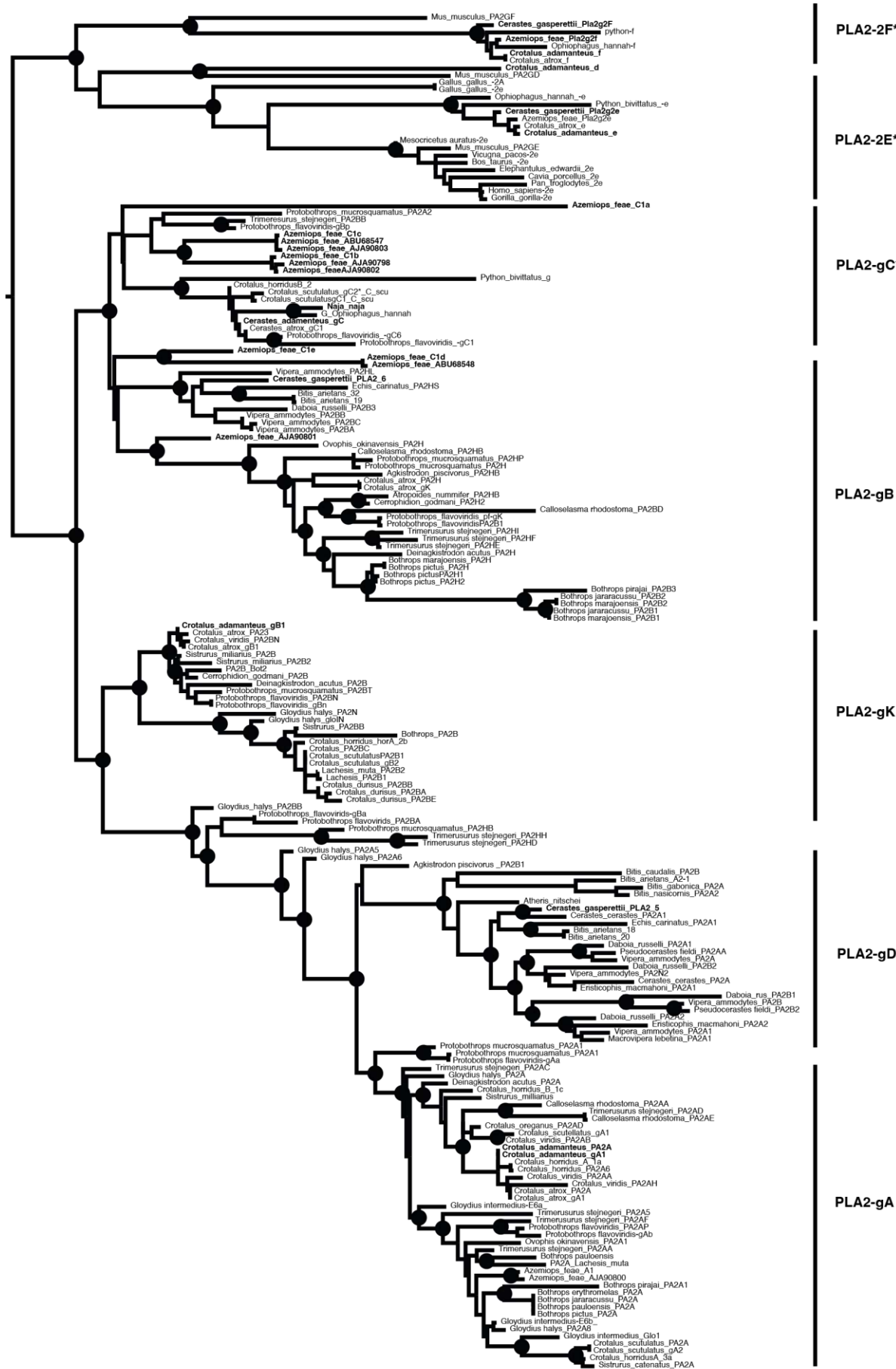


Fig. S9: Maximum likelihood phylogeny for PLA., with the two non-toxic genes as outgroups (PLA-:

CN6134	-	UAE	UAE	Proteome
CN6135	-	UAE	UAE	Proteome

751

752 Table S2: Id, tissue type and number of reads sequenced per sample.

ID	Tissue	Reads
CG9	Tongue	44,672,733
CG9	Venom gland	41,124,132
CG9	Eye	41,951,109
CG9	Brain	42,800,966
CG9	Heart	40,715,947
CG9	Lung	42,518,938
CG9	Liver	42,251,137
CG9	Gallbladder	42,738,665
CG9	Spleen	40,909,550
CG9	Pancreas	40,527,010
CG9	Ovary	41,118,336
CG9	Kidney	40,620,023
CG9	Accessory gland	44,114,293
CG14	Tongue	41,455,346
CG14	Venom gland	41,035,764
CG14	Eye	40,753,220
CG14	Brain	43,413,973
CG14	Heart	42,338,980
CG14	Lung	42,068,410
CG14	Liver	21,549,210
CG14	Gallglabbder	50,571,941

CG14	Spleen	45,447,235
CG14	Pancreas	50,223,941
CG14	Testis	47,495,900
CG14	Kidney	45,945,776
CG1	Heart	47,362,067
CG1	Brain	45,740,571
CG1	Kidney	50,758,869
CG1	Gallbladder	40,546,711
CG1	Liver	48,058,958
CG1	Spleen	44,752,981
CG1	Tongue	46,837,490
CG1	Pancreas	45,023,783
CG1	Venom gland	49,775,424
CG1	Ovary	48,703,420

753
754
755
756
757
758
759

760 Table S3: Different types of repetitive elements masked within the genome:

Element	Number of elements	Length (bp)	Percentage
Retroelements	1524124	493932584	30.25 %
SINEs:	339152	55265721	3.38
Penelope	124778	19471740	1.19
LINEs:	988815	347028895	21.25
CRE/SLACS	0	0	0.00%
L2/CR1/Rex	480371	137654000	8.43
R1/LOA/Jockey	579	99034	0.01
R2/R4/NeSL	41793	10873028	0.67

RTE/Bov-B	128092	79663597	4.88
L1/CIN4	207974	95913575	5.87
LTR elements:	196157	91637968	5.61
BEL/Pao	16545	5263265	0.32
Ty1/Copia	25582	15088781	0.92
Gypsy/DIRS1	102598	63604234	3.90
Retroviral	50617	7642063	0.47
DNA transposons	707499	111444059	6.83
hobo-Activator	265944	30679712	1.88
Tc1-IS630-Pogo	227637	58877559	3.61
En-Spm	0	0	0.00%
MULE-MuDR	44	3962	0.00%
PiggyBac	138	6619	0.00%
Tourist/Harbinger	182161	18395721	1.13
Other	0	0	0.00%
Rolling-circles	2242	136656	0.01
Unclassified	205700	42385187	2.60
Total interspersed repeats	-	647761830	39.67
Small RNA	6134	652217	0.04
Satellites	35838	4217238	0.26
Simple repeats	765726	53044358	3.25
Low complexity	97863	6694649	0.41

Table S4: Abundances for the different toxin families identified in the proteome of *C. gasperettii*

Toxin family	Percentage
SVMPi	8.65%








DISI	12.74%
DC domain	0.26%
CRISP	4.34%
PLA2	5.47%
SVSP	37.38%
SVMP-III	22.19%
PDE	0.02%
LAAO	1.71%
CTL	7.25%

References

1. Dussex, N., van der Valk, T., Morales, H. E., Wheat, C. W., Díez-del-Molino, D., von Seth, J., Foster, Y., Kutschera, V. E., Guschanski, K., Rhie, A., Phillippy, A. M., Korlach, J., Howe, K., Chow, W., Pelan, S., Mendes Damas, J. D., Lewin, H. A., Hastie, A. R., Formenti, G., ... Dalén, L. (2021). Population genomics of the critically endangered kākāpō. *Cell Genomics*, 1(1), 100002. <https://doi.org/10.1016/j.xgen.2021.100002>
2. Hogan, M. P., Holding, M. L., Nystrom, G. S., Colston, T. J., Bartlett, D. A., Mason, A. J., Ellsworth, S. A., Rautsaw, R. M., Lawrence, K. C., Strickland, J. L., He, B., Fraser, P., Margres, M. J., Gilbert, D. M., Gibbs, H. L., Parkinson, C. L., & Rokyta, D. R. (2024). The genetic regulatory architecture and epigenomic basis for age-related changes in rattlesnake venom. *Proceedings of the National Academy of Sciences*, 121(16), e2313440121. <https://doi.org/10.1073/pnas.2313440121>
3. Margres, M. J., Rautsaw, R. M., Strickland, J. L., Mason, A. J., Schramer, T. D., Hofmann, E. P., Stiers, E., Ellsworth, S. A., Nystrom, G. S., Hogan, M. P., Bartlett, D. A., Colston, T. J., Gilbert, D. M., Rokyta, D. R., & Parkinson, C. L. (2021). The Tiger Rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. *Proceedings of the National Academy of Sciences*, 118(4), e2014634118. <https://doi.org/10.1073/pnas.2014634118>
4. Pardos-Blas, J. R., Irisarri, I., Abalde, S., Afonso, C. M. L., Tenorio, M. J., & Zardoya, R. (2021). The genome of the venomous snail *Lautoconus ventricosus* sheds light on the origin of conotoxin diversity. *GigaScience*, 10(5), giab037. <https://doi.org/10.1093/gigascience/giab037>
5. Schield, D. R., Card, D. C., Hales, N. R., Perry, B. W., Pasquesi, G. M., Blackmon, H., Adams, R. H., Corbin, A. B., Smith, C. F., Ramesh, B., Demuth, J. P., Betrán, E., Tollis, M., Meik, J. M., Mackessy, S. P., & Castoe, T. A. (2019). The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome Research*, 29(4), 590–601. <https://doi.org/10.1101/gr.240952.118>

ORIGINAL ARTICLE OPEN ACCESS

Unveiling the Evolutionary History of European Vipers and Their Venoms From a Multi-Omic Approach

Adrián Talavera^{1,2}  | Marc Palmada-Flores¹  | Fernando Martínez-Freiria^{3,4}  | Gabriel Mochales-Riaño¹  | Bernat Burriel-Carranza^{1,5}  | Maria Estarellas¹  | Daniel Fernández-Guiberteau⁶  | Álvaro Camina⁷ | Sylvain Ursenbacher^{8,9}  | Judit Vörös^{10,11}  | Bálint Halpern^{11,12,13}  | Davinia Pla¹⁴  | Juan José Calvete¹⁴  | Alexander S. Mikheyev¹⁵  | Tomàs Marquès-Bonet^{1,16,17,18}  | Salvador Carranza¹ 

¹Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain | ²CREAF, Cerdanyola del Vallès, Catalonia, Spain | ³CIBIO, Centro de Investigação Em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, Vairão, Portugal | ⁴BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Vairão, Portugal | ⁵Museu de Ciències Naturals de Barcelona, Barcelona, Spain | ⁶Centre de Recerca i Educació Ambiental de Calafell (CREAC-GREN-P-Ajuntament de Calafell), Tarragona, Spain | ⁷Grupo Atrox, Faunia, Madrid, Spain | ⁸Section of Conservation Biology, Department of Environmental Sciences, University of Basel, Basel, Switzerland | ⁹Info Fauna – CSCF and Karch, University of Neuchâtel, Neuchâtel, Switzerland | ¹⁰HUN-REN Balaton Limnological Research Institute, Tihany, Hungary | ¹¹MME Birdlife Hungary, Budapest, Hungary | ¹²Department of Systematic Zoology and Ecology, Institute of Biology, ELTE-Eötvös Loránd University, Budapest, Hungary | ¹³HUN-REN – ELTE – MTM Integrative Ecology Research Group, Budapest, Hungary | ¹⁴Evolutionary and Translational Venomics Laboratory, Consejo Superior de Investigaciones Científicas, Valencia, Spain | ¹⁵Research School of Biology, Australian National University, Canberra, Australian Capital Territory, Australia | ¹⁶CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain | ¹⁷Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain | ¹⁸Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain

Correspondence: Adrián Talavera (adrian.talavera@csic.es) | Salvador Carranza (salvicarranza@gmail.com)

Received: 29 January 2025 | **Accepted:** 26 June 2025

Handling Editor: Mitchell Cruzan

Funding: This work was funded by grant PID2021-128901NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF, a way of making Europe, and grant 2021 SGR 00420 from the Departament de Recerca i Universitats de la Generalitat de Catalunya to SC. AT is supported by “la Caixa” doctoral fellowship programme (LCF/BQ/DR20/11790007). MP-F is supported by “la Caixa” doctoral fellowship programme (LCF/BQ/DR20/11790032). FM-F is supported by FCT—Fundação para a Ciência e Tecnologia de Portugal (ref. DL57/2016/CP1440/CT0010; DOI: [10.54499/DL57/2016/CP1440/CT0010](https://doi.org/10.54499/DL57/2016/CP1440/CT0010)). BB-C was funded by FPU grant from Ministerio de Ciencia, Innovación y Universidades, Spain (FPU18/04742). GM-R is funded by an FPI grant from the Ministerio de Ciencia, Innovación y Universidades, Spain (PRE2019-088729). ME is funded by an FPI grant from the Ministerio de Ciencia, Innovación y Universidades, Spain (PRE2022-101473). Some fieldwork campaigns have been supported by the Natuurmuseum Fryslân of Leeuwarden, Netherlands.

Keywords: adaptive introgression | chromosomal rearrangements | genomics | mito-nuclear discordance | snakes | venom evolution | *Vipera*

ABSTRACT

Snake genomes attract significant attention from multiple disciplines, including medicine, drug bioprospection, and evolutionary biology. However, genomic research within the Viperidae family has mostly focused on the subfamily Crotalinae, while the true vipers (Viperinae) have largely been overlooked. European vipers (*Vipera*) have been the subject of extensive research due to their phylogeographic and ecological diversification, as well as their venoms. Nevertheless, phylogeography and systematics in this genus have primarily relied on biased information from mitochondrial genes, which fail to capture the likely effects of introgression and are prone to biases. On the other hand, venom research in this group has been conducted predominantly through proteomics alone. In this study, we generated chromosome-level genome assemblies for three *Vipera* species and whole-genome sequencing data for 94 samples representing 15 *Vipera* lineages. This comprehensive dataset allowed us to disentangle the phylogenomic relationships of this genus, affected by mito-nuclear discordance and pervaded by ancestral introgression.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Molecular Ecology* published by John Wiley & Sons Ltd.

Population-level analyses in the Iberian Peninsula, where the three oldest lineages within *Vipera* meet, revealed signals of recent adaptive introgression between old-diverged and ecologically dissimilar species, whereas chromosomal rearrangements isolate species occupying similar niches. Finally, using transcriptomic and proteomic data, we characterised the *Vipera* toxin-encoding genes, in which opposing selective forces were unveiled as common drivers of the evolution of venom as an integrated phenotype.

1 | Introduction

The advent of genomics has transformed our understanding of introgression and its role in speciation within animals (Taylor and Larson 2019). Far from being strictly deleterious, as originally conceived, introgression serves as a source of genetic variation used both during local adaptation and adaptive radiation (Hedrick 2013). Reticulate speciation due to introgression is more likely to occur in groups with overlapping ranges and weak reproductive isolation barriers. This is the case for the European vipers of the genus *Vipera*, which encompasses ca. 15 species with multiple levels of intraspecific diversification (Dufresnes et al. 2024; Freitas et al. 2020).

The genus *Vipera* includes species with two distinct ecological affinities: a Eurosiberian set of cold-adapted species with relatively septentrional or mountainous ranges, and a Mediterranean set of warm-adapted taxa with parapatric ranges across the Mediterranean Basin (Figure 1). The Mediterranean group comprises one species complex and three species (i.e., *V. ammodytes* complex, *V. aspis*, *V. latastei*, and *V. monticola*) with high subspecific diversity, but uncertain phylogenetic relationships among them (Dufresnes et al. 2024; Freitas et al. 2020). In contrast, the Eurosiberian species form a well-supported monophyletic group nested within the Mediterranean vipers and divided into two clades: (I) *Vipera berus*, widely distributed across the Palearctic and the vicariant *V. seoanei* mostly endemic to Iberia; and (II) a speciose clade severely affected by taxonomic inflation, represented in Europe by the closely related *V. ursinii*, *V. graeca* and *V. renardi* (Freitas et al. 2020) (Figure 1A,B). The

sympatric distribution of certain species, combined with weak reproductive barriers, facilitates natural hybridization, which has been reported even between distantly related taxa (Guiller et al. 2017; Tarroso et al. 2014; Zwahlen et al. 2022), and makes this genus well suited to study reticulate speciation.

However, research on the phylogeography and systematics of this genus has heavily relied on mitochondrial DNA (mtDNA), unable to unveil introgression due to the absence of recombination, or few nuclear loci that lack enough resolution (Martínez-Freiría et al. 2020; Mizsei et al. 2017; Vörös et al. 2022). More recently, the first genome-wide studies, with reduced-representation data, have begun revealing introgression events within some of the *Vipera* lineages (Thanou et al. 2023; Zinenko et al. 2016), while also revealing significant mito-nuclear discordance in the genus (Dufresnes et al. 2024), albeit these findings remain assumption-biased.

Beyond their phylogenetic and ecological diversification, the European vipers have garnered attention also for their venoms (Avella et al. 2023; Giribaldi et al. 2020; Göçmen et al. 2015; Sajevec et al. 2014). Snake venom is a complex cocktail of multiple toxin families but acts as an integrated phenotype while subduing prey (Aird et al. 2015). Regarding its evolution, however, most studies have focused on how specific venom families diversify independently (Calvete et al. 2005; Casewell et al. 2011; Juárez et al. 2008; Lynch 2007; Sunagar et al. 2012), highlighting gene duplication and positive selection as key evolutionary mechanisms (Rao et al. 2022). Altogether, the few studies envisaging the evolution of venom-encoding genes as a whole have revealed accelerated evolution in these genes compared

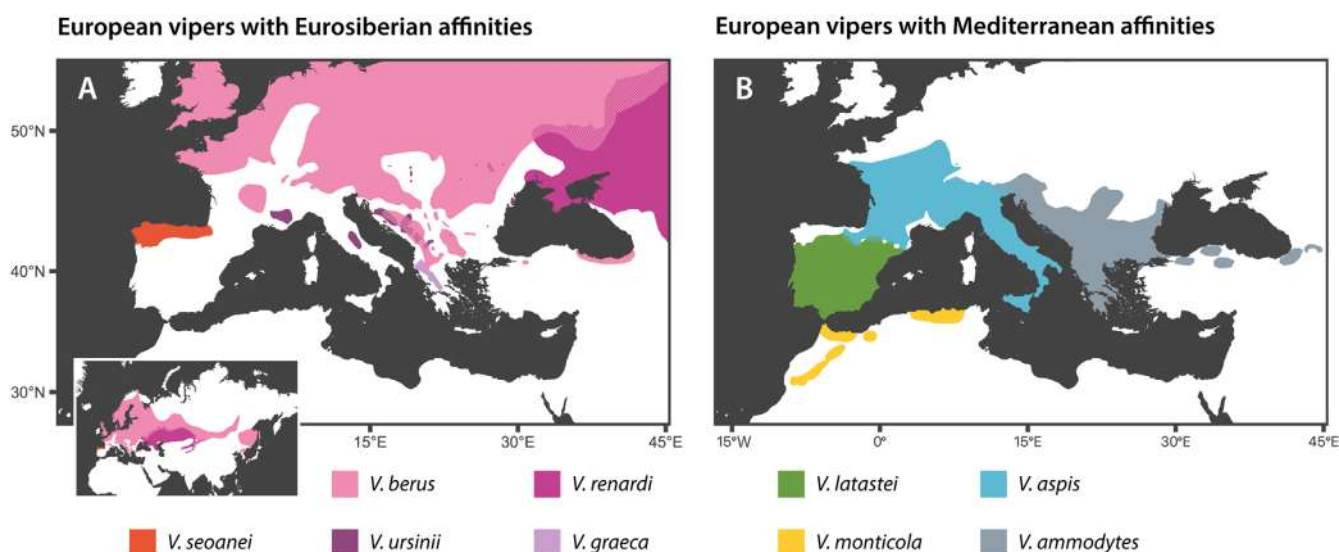


FIGURE 1 | Ranges of the *Vipera* species occurring in Europe, sorted by ecological affinities. (A) Eurosiberian species inhabit septentrional regions, or disjoint mountainous areas towards the South. *Vipera berus* and *V. renardi* extend their distributions out of Europe, shown on bottom-left map. (B) *Vipera* species with Mediterranean affinities have parapatric ranges surrounding the Mediterranean Basin, including one species, *V. monticola*, endemic to North Africa.

to non-venom paralogs (Shibata et al. 2018), especially in those most expressed (Aird et al. 2015). Although snake genomes are receiving increasing attention (Gutiérrez et al. 2017), with around 20 snake genomes assembled to date (Kerkkamp et al. 2016; Rao et al. 2022), venom research within the family Viperidae has focused on the subfamily Crotalinae (Margres et al. 2021; Schield et al. 2019) and, more recently, Azemiopinae (Myers et al. 2022). Nevertheless, few genomic studies in this regard have been conducted on the subfamily Viperinae, the so-called true vipers (but see Mochales-Riño, Hirst, et al. (2024)), contrasting with the extensive research on their venom proteomes (Damm et al. 2021).

In this study, we present chromosome-level genome assemblies for the three true vipers occurring in the Iberian Peninsula: *V. aspis*, *V. latastei*, *V. seoanei*, and whole genome sequencing (WGS) data from 94 specimens representing 15 *Vipera* taxa (Table S1), including all species occurring in Europe but *V. renardi* and *V. graeca*. The aim of this work is to unravel the uncertain phylogenomic relationships of this genus and its evolutionary history, including chromosomal rearrangements, demographic oscillations, ancient introgression events, and current secondary contacts among old-diverged species, for which we discuss their adaptive potential focusing on the Iberian Peninsula. Finally, leveraging the assembly of reference genomes from three old-diverged species within the genus *Vipera*, and integrating available and *de novo* proteomic and transcriptomic data, we characterise the venom encoding-genes and analyse the drivers behind the evolution of this complex and fascinating phenotype in the genus *Vipera*.

2 | Material and Methods

More detailed information on the Materials and Methods used in this work is provided in the Appendix S1.

2.1 | Reference Genome Sequencing, Assembly, and Annotation

In brief, the reference genomes of *V. latastei*, *V. seoanei*, and *V. aspis* were sequenced and assembled with data from different platform combinations, among ONT and PacBio HiFi long reads, Illumina short reads, and Omni-C, Hi-C contact map or Bionano optical mapping (see Appendix S1 for assembly pipelines), yielding 18, 18, and 21 chromosomes, respectively. For *V. seoanei*, we scaffolded contigs using the closely related *V. ursinii* as reference instead. Genome assemblies were depicted as snailplots with BlobToolsKit (Figures S1–S3), and their contiguity and completeness were assessed with gfastats (Formenti et al. 2022) and BUSCOs vertebrate database (Simão et al. 2015). In addition, data from 19 RNA-Seq samples from 12 different tissues of the three species, full-length transcriptome Iso-Seq data from pooled tissues and venom glands, as well as protein databases and other snake genomes served as evidence to annotate the *Vipera* genomes following Gene-MarkS-T, BRAKER 1 and 2, GeMoMa, and Tsebra pipelines (Brûna et al. 2021; Gabriel et al. 2021; Hoff et al. 2016; Keilwagen et al. 2016; Tang et al. 2015).

Finally, we explored macrosynteny between the three Iberian reference genomes and *Crotalus viridis* as the outgroup (Pasquesi et al. 2018) with MCscan (Tang et al. 2008). Protein sequences were aligned between genomes with LAST (Kielbasa et al. 2011) and some chromosome reorientation was carried out to correctly visualise chromosomal rearrangements and identify the sexual chromosomes.

2.2 | Whole-Genome Sequencing Data Production

Furthermore, we build genomic libraries following Carøe et al. (2018) to sequence five additional genomes (apart from the three references) at high coverage (~40×) with 150-paired-end (PE) Illumina short reads to represent other species/lineages within the genus (i.e., *V. ursinii*, *V. berus*, *V. ammodytes*, *V. monticola* and Eastern *V. aspis*). Other 86 genomes were sequenced at low coverage to explore finer-scale population structure and introgression patterns focusing on the Western Mediterranean, including all subspecies within *V. latastei*, *V. monticola*, and *V. aspis*. Although *V. seoanei* is thought not to harbour subspecific diversity, its range was exhaustively sampled as well (Martínez-Freiria et al. 2015).

Afterwards, WGS data was quality filtered and adapter-trimmed with fastp (Chen et al. 2018) and mapped to *V. latastei* reference genome (rVipLat1) with BWA (Li 2013). PCR duplicates were removed with PicardTools and we called SNPs and filtered the resulting dataset with GATK (Poplin et al. 2018). Further filtering with VCFtools (Danecek et al. 2011) for subsequent analyses included filtering by missingness, keeping just autosomal bi-allelic SNPs, purging singletons, and Linkage-Disequilibrium pruning, whose decay function was calculated with PopLDdecay (Zhang, Dong, et al. 2019) (Figure S4 and Table S2). Finally, mitochondrial sequences were extracted and assembled from WGS data from 15 representatives of different taxa within *Vipera* with GetOrganelle (Jin et al. 2020). Resulting mitogenomes were annotated with MitoFinder (Allio et al. 2020).

2.3 | Phylogenomic and Phylogenetic Analyses

First, genomic species trees were inferred with SNAPP implemented on BEAST2 (Bouckaert et al. 2014) for two unlinked SNP (uSNPs) datasets without missingness, at species- and subspecies-level, with 1.23 M and 126 k uSNPs, respectively (Table S2), and three runs of 20 M iterations for each dataset. These phylogenies were time-calibrated with the estimated divergence between Viperinae and Crotalinae (38.61 ± 4.85 mya) for the species dataset, and the crown diversification of the *V. berus* + *V. seoanei* + *V. ursinii* clade, that is, the *berus* group (5.97 ± 0.75 mya) for the subspecies phylogeny without outgroup, following fossil-calibrated phylogenetic studies (Šmíd and Tolley 2019). On the other hand, from the mitogenomes of all subspecies plus *Crotalus* as outgroup, we aligned their protein coding genes, investigated the best partition scheme with the greedy algorithm of PartitionFinder (Lanfear et al. 2017) and ran three runs of BEAST2 to estimate the mitochondrial tree, accounting for 150 M iterations in total. We time-calibrated this last dataset with the crown age of the

berus group as well. The mitochondrial and genomic trees were depicted overlapped for a better visualisation of mitochondrial discordance.

2.4 | Demographic Inference, Heterozygosity and Runs of Homozygosity

To explore the influence of climate change on the demographic history of the European vipers, we ran PSMC on each high-coverage representative sample of the studied *Vipera* spp., following parameter selection by Schield et al. (2022), a rate of $2.4\text{e-}9$ substitutions/site/year (Pasquesi et al. 2018) and a 6-year generation period (Barbanera et al. 2009). Ten bootstrap replicates were calculated per genome, and results were divided into Eurosiberian or Mediterranean species according to ecological affinities.

In addition, we calculated mean autosomal heterozygosity per individual throughout 100-kbp windows, following Mochales-Riaño et al. (2023), using 94 low-coverage genomes. Means per species were plotted as bubble plots with ggplot2 (Wickham 2016). Runs of homozygosity (ROHs) across autosomes for the 8 high-coverage samples were estimated with bcftools (Narasimhan et al. 2016), filtered by quality (PhredScore > 70), sorted out by length, and plotted as a percentage of the total autosomal genome size.

2.5 | Estimating Interspecific Introgression

We explored interspecific introgression among identified *Vipera* lineages/subspecies by means of *f*-branch statistics, derived from ABBA-BABA tests in Dsuite (Malinsky et al. 2021). We excluded F1 hybrids and individuals resulting from secondary contacts between intraspecific lineages (see Admixture results), and included a *Crotalus* species as outgroup, yielding a low-coverage dataset of 86 individuals and > 12M linked SNPs. We tested both sensitive and robust clustering thresholds implemented in Dsuite to avoid false positives due to ABBA homoplasy and plotted the statistically significant results. Secondly, we ran TreeMix (Pickrell and Pritchard 2012) to investigate the direction of the introgression events on a subspecies dataset. Models with different migration edges between 1 and 10 were estimated by blocks of varying numbers of SNPs, and the optimum number of edges was estimated with the OptM package (Fitak 2021) in R. Finally, we investigated the landscape of the most important introgression event in the phylogeny with TWISST (Martin and Van Belleghem 2017), highlighting the windows supporting introgression between *V. seoanei* and *V. latastei-monticola*.

2.6 | Population Genomic Analyses in the Western Mediterranean

First, we explored with genomic PCAs the population structure of Western Mediterranean vipers, using data from 90 low-coverage whole genomes of four species. Following, we delve into the population structure and potential inter-specific secondary

contacts with Admixture (Alexander et al. 2009) ($K=1-12$). We depicted the optimum K as individual pie charts in the map. In addition, for the two best-sampled species: that is, *V. latastei* and *V. seoanei*, we examined whether the discovered intra-specific structure could be explained by Isolation-by-distance patterns with the R package prabclus (Hennig and Hausdorf 2019). In addition, per-individual mean heterozygosity values for the three Iberian species were estimated and used to interpolate a heterozygosity raster of each species. Finally, we inferred the climatic niches and investigated niche overlap between these two species and their respective lineages, both excluding or including localities assigned to admixed populations. We extracted 19 environmental variables from WorldClim2 (Fick and Hijmans 2017) belonging to *V. latastei* ($n=1105$) and *V. seoanei* ($n=508$) and summarised them into climatic PCs. Niche overlap was estimated with both Sørensen and Overlap Indexes (Lucchini et al. 2023; Martínez-Freiría et al. 2020; Mochales-Riaño, Burriel-Carranza, et al. 2024), using the R package hypervolume (Blonder et al. 2014), and visualisation was performed with ggplot2 (Wickham 2016) through 2-D PCAs. We highlighted admixed individuals to illustrate if their niches are congruent with the expected niches for non-admixed conspecifics.

2.7 | Venom Characterisation and Evolution

In brief, we integrated transcriptomic and proteomic data from the three species to characterise the toxin-encoding genes of the Iberian vipers. On the one hand, RNA-Seq data from 19 samples from venom glands and other tissues (i.e., gonads, brain, pancreas, eye, salivary gland, heart, kidneys, stomach, gall bladder and liver; Figure S5) were used to identify positively up-regulated genes exclusive to the venom gland (common to the three species, but with *V. latastei* as reference). We used DESeq2 (Love et al. 2014) to standardise gene expression and test for significantly upregulated genes (\log_2 fold-change > 2) in the venom glands vs. all other tissues, correcting for multiple tests (FDR: 5%).

On the other hand, we used available proteome data from venoms of *V. seoanei* (Avella et al. 2023) and *V. aspis* (Giribaldi et al. 2020), as well as *de novo* produced proteomes from three *V. latastei* (see Appendix S1). We looked for exact matches of the peptide sequences of each specific proteome among the translated proteins annotated in its respective genome to identify the genes involved in the expression of those venoms. Venom-encoding genes derived from proteomics were compared to those derived from transcriptomics in the case of *V. latastei* to verify that proteome data is accurate and sufficient.

Finally, we performed d_N/d_S ratio tests with BUSTED (Murrell et al. 2015) and FEL (Kosakovsky Pond and Frost 2005) analyses implemented on Datamonkey (Weaver et al. 2018) to look for episodic diversifying selection and pervasive selection, respectively, in each toxin-encoding family. The dependence of the corrected relative ubiquity of pervasive diversifying or purifying selection from FEL tests was assessed based on the number of gene copies within a family and the abundance of this family in the venom composition (Table S3). Linear regressions were plotted with ggplot2.

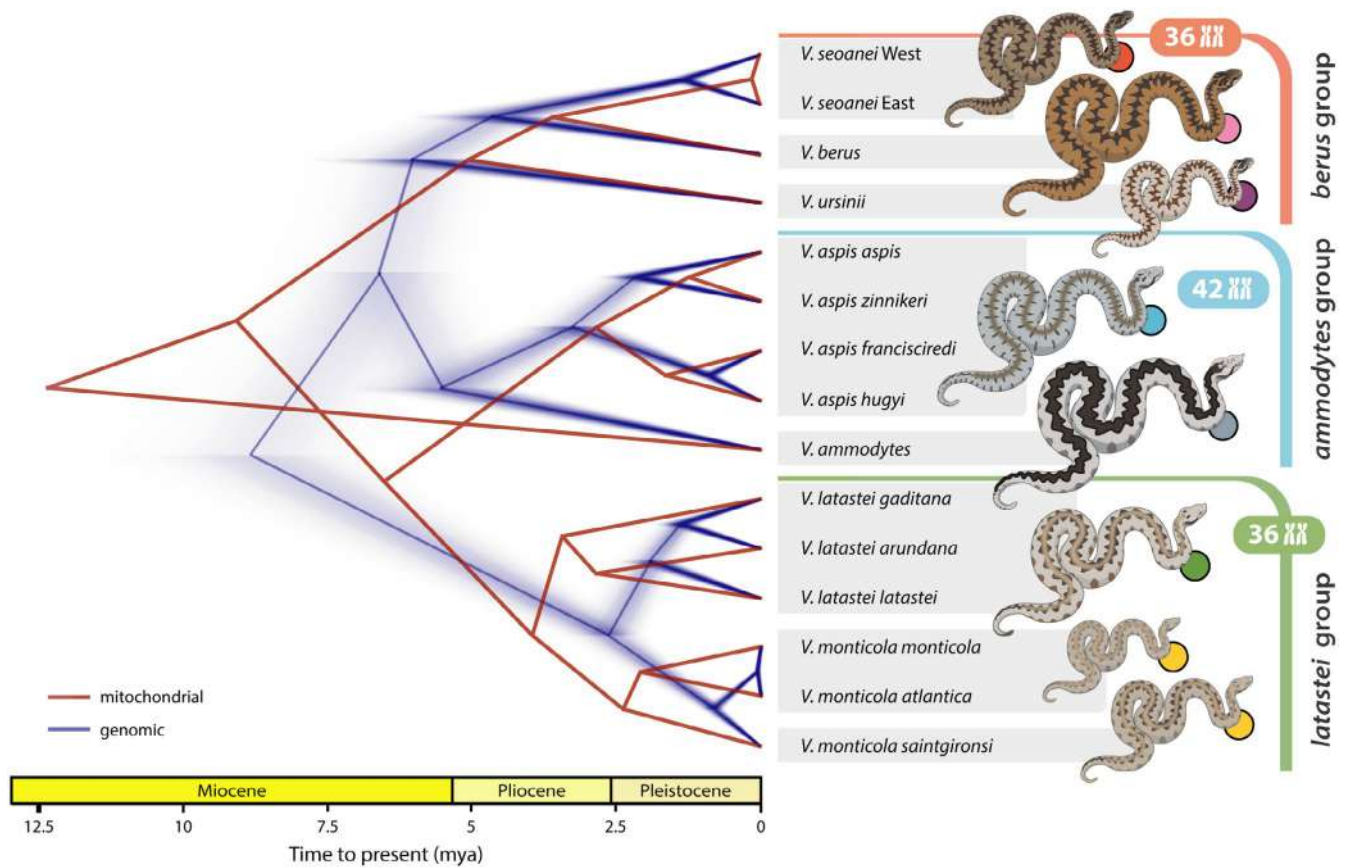


FIGURE 2 | Time-calibrated phylogenomic and mitochondrial species tree of European vipers, showing mito-nuclear discordance. In dark blue, genomic species tree from SNAPP, inferred from 126 k unlinked SNPs. In red, mitochondrial species tree from a 10.7-kbp alignment of all mitochondrial protein-coding genes. The mitochondrial tree was rooted with *Crotalus horridus* (see Figure S8), but excluded a posteriori for a better comparison. The genomic species tree recovers three main old species groups, namely: *berus*, *ammodytes*, and *latastei* groups. In disagreement with mitogenomes, *Vipera aspis* is sister to the *V. ammodytes* complex, which agrees with the neurotoxicity of their venoms and a higher chromosome number. The *latastei* group represents the earliest diverging branch, instead of *V. ammodytes*, as the mitogenomes estimate. Thus, representatives of the three old species groups are found in the Iberian Peninsula: *V. seoanei*, *V. aspis*, and *V. latastei*. Other *Vipera* species not included in this study are more closely related to *V. ursinii*.

3 | Results

We generated and annotated three chromosome-level genome assemblies for *V. aspis*, *V. latastei*, *V. seoanei*. Assembly sizes are comprised between 1.57 and 1.60 Gb, scaffold N50 between 114 and 236 Mb, contig N50 between 35.3 and 75.9 Mb, and single-copy ortholog genetic completeness between 96.0% and 97.2% (Table S4 and Figures S1–S3). 99.30% of the assembly is anchored to the expected number of chromosomes in *V. latastei* (17 autosomes and Z), 99.74% in *V. aspis* (20 autosomes, with Z partially assembled into 3 scaffolds) and 100% in *V. seoanei* (17 autosomes and Z) (Apréa et al. 2006; Saint Girons and Duguy 1976; Spangenberg et al. 2022).

3.1 | Genomic and Mitochondrial Phylogenies

In order to disentangle the nuclear history of the European vipers of the genus *Vipera*, we inferred the phylogenomic relationships among seven *Vipera* species, which represent all main lineages within the genus. The only two European species not included, *V. graeca* and *V. renardi*, are closely related

to *V. ursinii* (Dufresnes et al. 2024). We inferred Bayesian species trees from autosomal (SNAPP) and mitochondrial data to account for mito-nuclear discordance. We recovered concordant topologies for the genomic species trees in two SNAPP analyses, including or not intraspecific diversity (Figure 2 and Figures S6 and S7), although time calibrations of all nodes within *Vipera* were estimated to be much older when including a rattlesnake (subfamily Crotalinae) as an outgroup. Genomic phylogenies supported the existence of three old lineages within *Vipera*, namely, *latastei*, *ammodytes* and *berus* species groups. The *berus* group comprises all Eurosiberian species, that is, *V. berus*, *V. seoanei*, *V. ursinii*, as well as all the species not included in the study but more closely related to *V. ursinii* (Dufresnes et al. 2024). The *ammodytes* group, comprising the *V. ammodytes* species complex and *V. aspis*, is sister to the *berus* group, whereas the early-diverging *latastei* group comprises *V. latastei* and *V. monticola*. In this scenario, the Iberian Peninsula emerges as the only place where the three old lineages or species groups meet. The mitochondrial phylogeny is concordant with previous studies (Šmíd and Tolley 2019), although incongruent with the nuclear species trees inferred in the present study (Figure 2 and Figure S8).

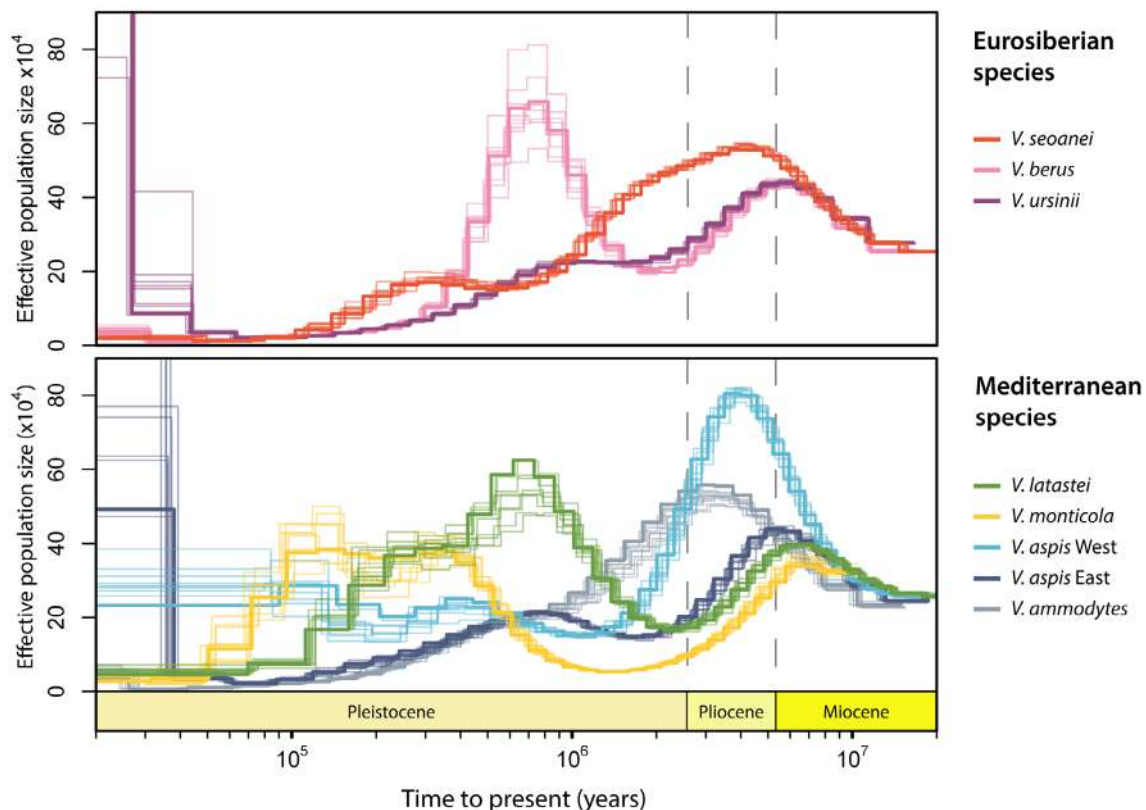


FIGURE 3 | Demographic history of the studied European vipers, sorted by ecological affinities. Thick and thin lines of the same colour represent the median and 10 bootstrap rounds, respectively. All species demographic inferences coalesce around 8 mya, in accordance with the time-calibrated genomic species tree (Figure 2 and Figure S7). Eurosiberian and Mediterranean sets of species do not show higher resemblance among themselves than between ecological groups. Whereas *V. seoanei*, Western *V. aspis*, and *V. ammodytes* seem to have thrived during the Pliocene, *V. berus*, *V. latastei*, and *V. monticola* underwent population expansions during the Pleistocene.

Mainly, *V. latastei* + *V. monticola* represent the first group to branch out within the genus; *V. aspis* is sister to *V. ammodytes* rather than to *V. latastei*, and some intraspecific lineage divergences are shallower with genomic than with mitochondrial data, for example, within *V. monticola*.

3.2 | Demographic Inference, ROHs, and Heterozygosity

We used high-coverage WGS data of the eight oldest lineages studied in this work to infer past demographic oscillations along the evolutionary history of European vipers. The analyses revealed that Eurosiberian and Mediterranean sets of species do not follow similar trends based on their biogeographic affinities. Whereas *V. aspis* West, *V. ammodytes*, and *V. seoanei* seem to have thrived during the Pliocene, *V. berus* and the *latastei* group underwent population expansions during the Pleistocene (Figure 3). Regarding *V. berus*, this expansion matches the proposed recolonisation of the continent 1 mya (Ursenbacher et al. 2006), while the following contraction might be due to a predicted loss of suitable habitat during glacial maxima (Lucchini et al. 2023). The demographic trends of the eight lineages clearly coalesce around 8 mya, matching the estimated divergence times with SNAPP (Figure 2). Species with persistent population declines, like *V. ursinii*, show the highest proportion of its genome covered by ROHs (i.e., the inheritance of identical

haplotypes from a common ancestor), especially short ones (Figure S9). Long ROHs arise from recent inbreeding over a few generations, which may reflect finer-scale, recent demographic trends of the population of origin of each sample. In contrast, shorter ROHs are indicative of ancient bottlenecks, as recombination events over time break ROHs apart (Ceballos et al. 2018). Finally, the highest mean genome-wide heterozygosity values appear to be found in the Iberian lineage of each species group (Figure S11). However, in some cases only a single or few individuals were sequenced, so these values may reflect the idiosyncrasies of the particular populations they belong to—for example, *V. berus* from Great Britain, which represents an island bottlenecked population.

3.3 | Introgression Analyses

We examined introgression events at the subspecific level with *f*-branch implemented on Dsuite, and backed these results with TreeMix estimates, which provide insights into the directionality of the events. We found introgression to be pervasive across the genus, with seven events estimated as optimal to explain the history of the studied *Vipera* (Figures S12–S15), and extraordinary *f*-branch values, with up to 23% of the genome in *V. seoanei* shared with *V. latastei* West (see “Population genomics” results for lineage delimitation; Figure 4A,B; Figure S16; Tables S5–S7). The detection of excess allele-sharing in *V. monticola* also

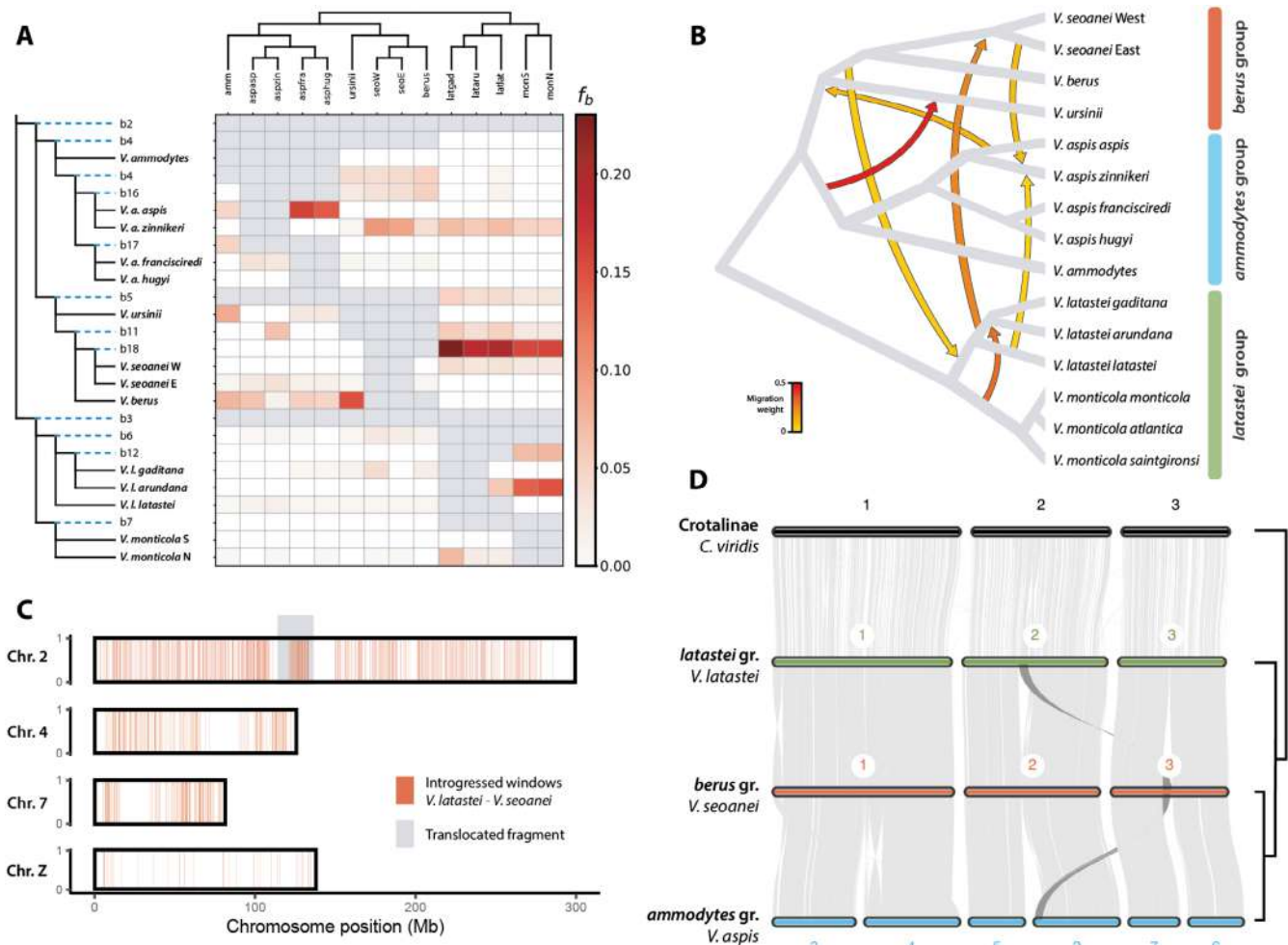


FIGURE 4 | Pervasive patterns of introgression across the genus *Vipera* and macrosynteny among the three largest chromosomes across major lineages. (A) Heatmap of f -branch (f_b) statistics, representing excess allele-sharing between branches of Y- and X-axes. Y-axis shows both terminal branches (in bold) and ancestral branches. Point estimates can be found in Table S6. Statistics have been corrected by a sensitive clustering threshold to avoid false positives due to homoplasy instead of introgression (see results with a robust threshold in Figure S16 and Table S7). Grey cells indicate comparisons that cannot be made due to tree topology. The most notable event of introgression seems to have happened between *V. seoanei* and *V. latastei*. (B) Most likely number of introgression events within the species tree of the studied European vipers, calculated with TreeMix and OptM (Figures S12–S15). Tree topology corresponds with Figure 2, and inferred introgression events are consistent with f -branch statistics, but in this case, the direction of gene flow is inferred as well. Colour of the migration edges corresponds to the intensity of the gene flow event. (C) Introgession landscape of the introgression event from *V. latastei* to *V. seoanei*. Windows with excess allele-sharing between both species are highlighted in *V. latastei*'s reference genome across chromosomes 2, 4, 7 and Z (see Figure S17 for all chromosomes). Introgression between these species is pervasive, although some areas display higher intensity than others, including the whole Z chromosome. An area affected by an interchromosomal translocation is highlighted in darker grey (see (D)). (D) Macrosynteny between the three *Vipera* species groups, represented by the Iberian species of each old lineage, and a rattlesnake as outgroup (*Crotalus viridis*, *Crotalinae*). Only the three major chromosomes are depicted; see Figure S18 for the whole comparison. Chromosome architecture is greatly conserved between *berus* and *latastei* groups ($2n = 36$) with a fragment being translocated from chr. 2 to chr. 3. Nevertheless, the *ammodytes* group ($2n = 42$), underwent three events of chromosomal fission, which might hamper hybridisation with other *Vipera* species groups.

suggests an ancient onset of this introgression, while an even higher excess between the Western lineages of both *V. latastei* and *V. seoanei* indicates that the introgression spread over time until or near the present. To delve into the introgression from *V. latastei* into *V. seoanei*, we looked at the landscape of this introgression in 25-kb windows across *V. latastei*'s genome. Although introgression is pervasive, certain areas show increased clustering, such as in chromosomes 2 and 7, whereas others, including the Z chromosome in its entirety, are all but devoid of introgression (Figure 4C and Figure S17). Other remarked introgression is displayed between Northern *V. monticola* and Western *V.*

latastei, implying overseas dispersal, which had been already reported in some islands (Torres-Roig et al. 2021).

3.4 | Macrosynteny Analysis

We investigated chromosomal rearrangements between the three old species groups (represented by the three Iberian species), which may explain hybridisation and introgression patterns among them. Macrosynteny analyses confirm the retention of the ancestral snake bimodal karyotype ($2n = 36$) (Beçak and

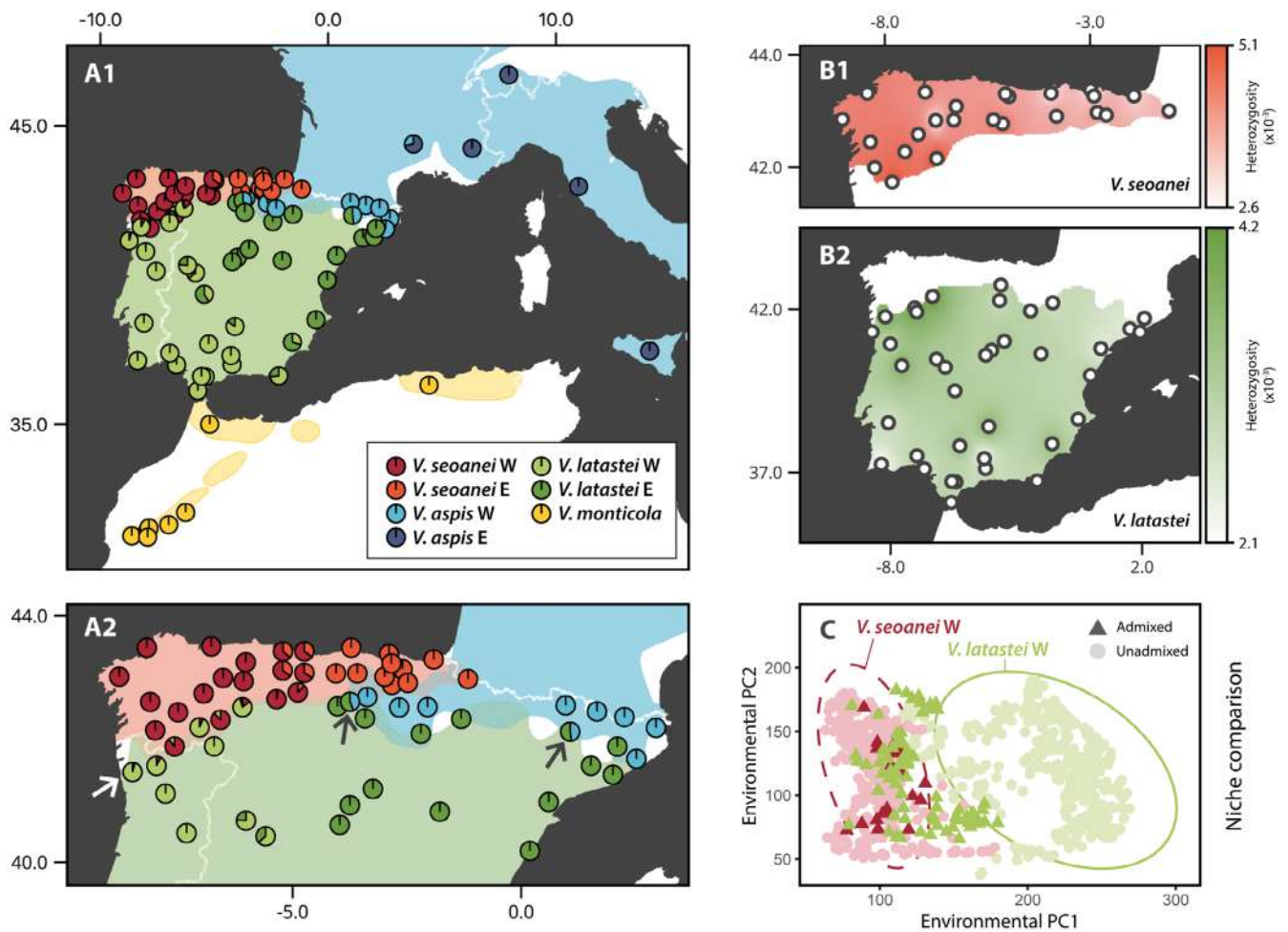


FIGURE 5 | Population genomics of vipers in the Western Mediterranean, a crucible for the three *Vipera* species groups. (A1, A2) Admixture results for $K=7$, 1.59 M unlinked SNPs and 90 low-coverage whole genomes, covering all subspecific diversity of these species. *Vipera seoanei*, *V. latastei* and *V. aspis* show Western and Eastern clades and exhibit different inter-specific admixture patterns in Northern Iberia (A2). In NW Iberia, the Western lineages of *V. seoanei* and *V. latastei* ($2n=36$) show admixture despite their marked ecological differences. The white arrow points to a *V. latastei* individual with *V. seoanei* admixture that was sampled 70 km away from *V. seoanei*'s distribution limit. On the other hand, on the contact zones of *V. aspis* West and *V. latastei* East, only pure or F1 generation hybrids (highlighted with dark arrows) are found despite the Mediterranean affinities of both species, which, in addition, differ in chromosome number. (B1, B2) Genome-wide heterozygosity interpolations across the distributions of *V. seoanei* and *V. latastei*. Darker colours represent higher mean heterozygosity values per individual. The highest values, for both species, are reached around the contact area in North Portugal and South Galicia (Spain). (C) Climatic niche comparison of the Western *V. seoanei* and *V. latastei* lineages. Ellipses represent the dispersion with 95% CI of individuals from unadmixed populations of those species (circles). Individuals from recently-admixed populations are depicted by triangles. While putatively admixed *V. seoanei* individuals inhabit similar localities than unadmixed conspecifics, admixed *V. latastei* specimens occupy more Eurosiberian environments than expected for their lineage.

Beçak 2008; Gorman and Gress 1970; Singh 1972) in *V. latastei* and *V. seoanei*, but reveal three fission events in the three largest chromosomes of *V. aspis*, rendering $2n=42$ (Figure 4D and Figure S18). Despite *V. latastei* and *V. seoanei* karyotypes being almost identical, an inner fragment from *V. latastei* chromosome 2 has been translocated to chromosome number 3 of *V. seoanei*.

3.5 | Population Genomics

We investigated at a population level the Western Mediterranean, and specifically the Iberian Peninsula, as a potential crucible of ancient lineages: only in Iberia do the three old species groups co-occur, with two species, *V. seoanei* and *V. latastei*, sharing a staggering excess of alleles. We used a dataset of 90 genomes

at low coverage, covering all species and subspecies in the Western Mediterranean, and yielding 1.59 M uSNPs that we subsequently analysed by PCAs (Figures S19–S24), Admixture (Figure 5A1,A2 and Figure S25), mean heterozygosity per individual (Figure 5B1,B2 and Figure S26) and Isolation-by-Distance (IBD) models (Figures S27 and S28).

We found that *V. seoanei* and *V. latastei* are structured into Western and Eastern lineages across their respective ranges (Figure 5A). In *V. seoanei*, the high summits of the Cantabrian Mountains, particularly at Picos de Europa, separate a genetically more diverse Western lineage from a relatively impoverished Eastern one (Figure 5B1), a pattern that is partially explained by isolation-by-distance (Figure S28). *Vipera seoanei*'s type locality is comprised within the Western range,

which encompasses as well the “cantabrica” ecotype, previously considered a subspecies (Martínez-Freiría et al. 2015). For *V. latastei*, its Eastern lineage mostly corresponds to *V. l. latastei*, whereas the Western one comprises *V. l. gaditana*, the recently described *V. l. arundana* (Martínez-Freiría et al. 2021) and some SE populations traditionally assigned to *V. l. latastei* (Figure 5A1 and Figure S21). However, genomic differences are fully explained by IBD (Figure S27). *Vipera aspis* seems better resolved by the PCA than Admixture due to sample biases, and the Alps split an impoverished Eastern lineage (*V. a. francisciredi* + *V. a. hugyi*) from two more diverged subspecies in the West: *V. a. zinnikeri* and *V. a. aspis* (Figure S23). *Vipera monticola*, in contrast, is more homogeneous, and PC1 splits it into a Northern lineage, partially corresponding to *V. m. saintgironi* (Martínez-Freiría et al. 2021), and a Southern lineage, including Eastern High Atlas populations of *V. m. saintgironi*, *V. m. monticola* and *V. m. atlantica* (Martínez-Freiría et al. 2021) (Figure S22).

In the northern sector of Iberia, the three old species groups meet with uneven outcomes (Figure 5A2). Between the Mediterranean species *V. aspis* and *V. latastei*, recent admixture is barely detected, although two F1-hybrids have been identified, with half ancestry of each species (Figure S25) and soared heterozygosity (Figure S11). However, in NW Iberia, the ecologically dissimilar *V. latastei* and *V. seoanei* admix in a broad area of secondary contact encompassing South Galicia and North Portugal, in which the highest heterozygosity values are found for both species (Figure 5B1,B2). Admixture even reaches a locality next to the city of Porto, ~70 km from the current limit of *V. seoanei*'s distribution (Figure 5A2).

To ascertain the adaptive character of this admixture, we used publicly available occurrence data to estimate ecological niche overlap between *V. seoanei* and *V. latastei* (Table S8) and depicted it by environmental PCAs (Figures S29–S34, Table S9). Whereas these species and their respective lineages barely overlap (OI = 9%–20%, Table S8), when we include the admixed populations, their niches converge (32%). PC1 demonstrates that a Eurosiberian–Mediterranean gradient separates the niches of these species. Remarkably, individuals of *V. latastei* admixed with *V. seoanei* tend to occupy Eurosiberian niches typical of *V. seoanei* (Figure 5C and Figure S32). Western and Eastern intraspecific lineages were found to be similar in *V. seoanei* (62%, Figure S34), but unexpectedly different in *V. latastei* (12%), with the Eastern lineage inhabiting areas of greater seasonality and/or drier winters (Figure S33).

3.6 | Venom Characterisation and Evolution

We used both transcriptomic and proteomic (new and published) data from the three Iberian vipers, and their reference genomes, to explore how toxin-encoding genes have diversified in the genus *Vipera*. Differential gene expression analyses (Figure 6A) revealed 40 genes exclusively upregulated in the venom gland, 18 of them being toxin-encoding genes. Other upregulated genes are related to muscle contraction (13), protein synthesis (3), or vacuolar fusion (1), one is an immunoglobulin and four correspond to transposable elements (Figure S35). For *V. latastei*, genes identified through transcriptomics and through proteomics greatly overlap (Table S10), although proteomic results were

more complete, validating proteomic results for the other two species (Tables S11 and S12). Microsynteny of the primary and secondary toxin families is depicted in Figure S36. We found one PLA₂ gene in *V. aspis* to be responsible for the neurotoxicity of its venom reported in some populations (Chowdhury et al. 2022). PLA₂s and other primary families, more expressed in venom, do not necessarily have more gene copies than secondary families, like CTLs/snaclecs (Figure 6B). Indeed, some families supposed to bear many copies, like SVSPs and SVMPS, do instead express a multitude of isoforms out of just one gene copy (e.g., P-I, P-II, P-III and disintegrins produced by alternative splicing of the same SVMPS gene). For some families, genome-wide blasts detected candidate genes to be ancestral to each family gene cluster, such as *adam28* for SVMPS (Giorgianni et al. 2020). In addition, next to some CRiSPs (Figure S36), we found *serotriflin* genes, a carrier of an endogenous inhibitor (SSP-2) in serum involved in suppressing the toxic activity of CRiSPs in case of self-venomation (Aoki et al. 2008). In total, we found 28, 16 and 41 toxin-encoding genes in the genomes of *V. latastei*, *V. seoanei* and *V. aspis*, respectively, although with similar copy numbers within main families (Figure 6C and Tables S10–S12).

Finally, selection analyses revealed episodic selection in all main toxin families but 5'-nucleotidases, svVEGF, and svNGF, the latter two being involved in fostering existing prey's metabolic pathways, instead of disrupting them (Ferreira et al. 2021; Sunagar et al. 2013) (Table S13). Regarding pervasive selection, the toxin-encoding gene families that are more expressed undergo a stronger purifying selection (Figure 6D; adjusted $r^2 = 0.47$; $p_{\text{value}} = 0.012$), whereas diversifying selection does not depend on the importance of the family ($p_{\text{value}} = 0.692$). Indeed, diversifying selection increases with the number of copies within a family (Figure 6D; adjusted $r^2 = 0.54$; $p_{\text{value}} = 0.006$), whereas this feature does not correlate with purifying selection ($p_{\text{value}} = 0.730$).

4 | Discussion

In the present work we have generated three genome assemblies with unprecedented levels of contiguity in Viperidae. Although scaffold N50 values in *V. latastei* and *V. seoanei* are similar to other snake assemblies (Table S4), this parameter, widely used to compare chromosome-level assemblies' contiguity (Rao et al. 2022; Suryamohan et al. 2020), is dependent on genome architecture. Thus, these two species, with three macrochromosomes accounting for half of their genome, hold unsurprisingly higher values than *V. aspis*, in which these chromosomes are fragmented due to fission. However, contig N50 better depicts contiguity in chromosome-level assemblies. While *Crotalus tigris* was probably the most contiguous viperid genome prior to this work (Margres et al. 2021), these three *Vipera* genomes are 16–36 times more contiguous (Table S4).

4.1 | The Evolutionary History of European Vipers

Here we have first estimated the phylogenomic relationships among the European vipers of the genus *Vipera* from WGS data, revealing the monophyly of *V. aspis* + *V. ammodytes*. This is supported by a karyotypic synapomorphy: three extra chromosome

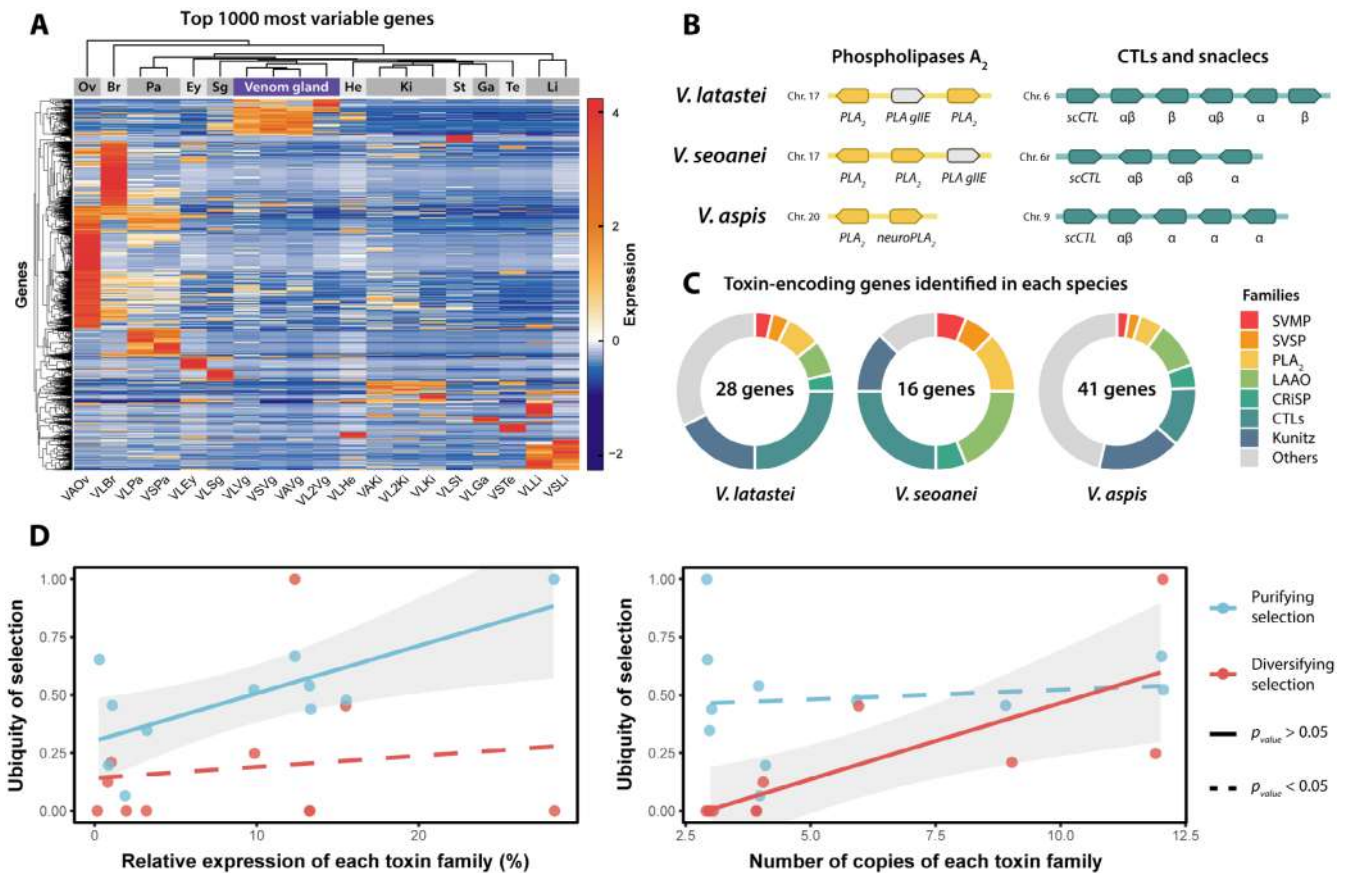


FIGURE 6 | Characterisation of toxin-encoding genes of *V. latastei*, *V. seoanei* and *V. aspis*, and venom selection analysis. (A) Heatmap of differential gene expression analyses, from different organs of the mentioned species (Figure S5). Y-axis is composed by the top 1000 most variable genes, and X-axis shows RNA-Seq samples from different organs and tissues. Samples from same organs cluster together, such as the venom glands coming from four specimens of the three species. Red hues indicate upregulation, whereas blue colours downregulation. Ov stands for ovary, Br for brain, Pa for pancreas, Ey for eye, Sg for salivary gland, He for heart, Ki for kidney, St for stomach, Ga for gall bladder, Te for testes and Li for liver. (B) Microsynteny of two selected toxin families in the three Iberian species. Here, only Phospholipases A₂ and C-type lectins (CTLs) + snakelets are shown, but all main toxin families are depicted in Figure S36. PLA₂ gIIIE are non-toxic, but related to toxic PLA₂. PLA₂ is a main family in the venom of vipers, but only two copies are found in the genome of these species (one of them producing neurotoxic PLA₂, in *V. aspis*). CTLs, albeit composing a secondary family in terms of total expression in the venom, is richer in gene copies. scCTL stands for the ancestral (albeit venomous) single-chain CTLs, whereas α, β and αβ represent snakelet genes producing either one or both subunits of the snakelet heterodimer. (C) Pie charts of all characterised venom-encoding genes in each of the three reference genomes by toxin family (Tables S9–S11). (D) Selection analyses on the identified toxin-encoding genes. Toxin families with a higher expression undergo a significantly stronger purifying selection. On the other hand, the diversifying selection only increases with the number of gene copies within the family.

pairs (Spangenberg et al. 2022). In addition, *V. aspis* and *V. ammodytes* share significantly neurotoxic venoms, although the presence of neurotoxicity in other *Vipera* lineages makes it difficult to determine whether this trait is a synapomorphy or the result of convergent evolution (Chowdhury et al. 2022). This new topology is incongruent with the only previous genus-level study using genomic data, which used 1.1 Mbp to estimate phylogenomic relationships but incorrectly assumed that *V. ammodytes* was the root (Dufresnes et al. 2024), as shown by mitochondrial data.

The discovered mito-nuclear discordance in the phylogeny can arise from multiple sources, such as (1) incomplete lineage sorting (ILS), (2) mitochondrial capture, or (3) sex-biased asymmetries (Toews and Brelsford 2012). (1) A phylogeny is more prone to ILS when a rapid succession of speciation events occurs while ancestral population sizes are large, and both circumstances are met according to our demographic inference results (Figure 3),

which could explain a closer mitochondrial relationship of *V. aspis* to the *latastei* group than to the *berus* group. However, the largest mito-nuclear disagreement is the placement of *V. ammodytes* as sister taxon to all other *Vipera* species included in the analysis, with a remarkably deep mitochondrial divergence that is not compatible with the context that promotes ILS.

(2) The so-called mitochondrial capture occurs when mtDNA from one species/lineage is completely replaced by another through—usually adaptive—introgression (Toews and Brelsford 2012). When the donor lineage is extinct, this process is referred to as ghost introgression. Although previously overlooked, this phenomenon might be relatively widespread in nature (Ottenburghs 2020; Zhang, Tang, et al. 2019), and has even been described within some species of the *berus* group (Doniol-Valcroze et al. 2021). Regarding the case of *V. ammodytes*, a plausible hypothesis is that a proto-*V. ammodytes* lineage, already

diverged from the ancestor of *V. aspis*, hybridised with a now-extinct deeply divergent lineage of *Vipera*, expanding into its range and capturing its mtDNA. Further research should aim to validate this hypothesis, and address whether any nuclear fraction from this ghost lineage persists in *V. ammodytes*.

Moreover, mitochondrial capture could as well have caused the shallow mitochondrial structure between Western-Eastern *V. seoanei*, despite important genomic differences (Figure 2). This species seems to have re-expanded from West to East during the Last Glacial Maximum (Martínez-Freiría et al. 2015), but larger genomic differences would suggest that relict populations in the East contributed somewhat with their genomic material while capturing Western mtDNA.

The opposite scenario—high mitochondrial divergence despite low genomic differentiation—is referred to as deep mitochondrial divergence (DMD) and reflects long periods of geographical isolation followed by secondary contacts (Hogner et al. 2012). DMD is fostered in the face of high levels of geneflow and sex-biased asymmetries (3), that is, strong male-biased dispersal and strict female philopatry (Zhang, Tang, et al. 2019), as described in the European vipers (François et al. 2021; Zwahlen et al. 2021). This is likely the case for *V. monticola* and, to a lesser extent, *V. latastei* subspecific lineages, whose ancient mitochondrial divergences are not congruent with a relatively shallower nuclear structure, a byproduct of IBD. Thus, during harsh climatic conditions, the distributions of the ancestors of these species became fragmented, and in those small populations, mtDNA was easily fixed a long time ago. On the other hand, during climate amelioration periods, populations expanded and reconnected, with mainly male-mediated geneflow re-homogenising the nuclear genomes.

4.2 | Pervasive Introgression Across European Vipers

We have found introgression to pervade the phylogeny of European vipers at multiple levels (Figures 4A–C and 5A). Nevertheless, hybridisation among *Vipera* spp. seems to be avoided by (a) niche segregation and (b) chromosomal rearrangements. On the one hand, *Vipera* ranges in Europe are mainly parapatric (Figure 1), and in the narrow borders of sympatry between species, they can segregate by selecting slightly different habitats and microhabitats, having non-overlapping mating periods or different thermoregulatory strategies (Luiselli et al. 2007; Martínez-Freiría et al. 2010; Mebert et al. 2015, 2017). However, latitudinal differences in Europe used to be weaker in the past, with the Mediterranean regime only beginning in the Late Pliocene (3.2 mya; Suc 1984). This could have led to broad areas of sympatry, promoting ancient introgressions, and is illustrated by the lack of common demographic oscillations within the Eurosiberian and Mediterranean sets of species (Figure 3). On the other hand, niche convergence instead of segregation has been observed only between sympatric *V. latastei* and *V. aspis* (Freitas et al. 2023; Martínez-Freiría et al. 2008), two species with differing karyotypes. Individuals that are heterozygous for chromosomal rearrangements are partially or completely sterile, quickly generating a strong reproductive isolation (de Vos et al. 2020; Faria and Navarro 2010), in line with F1-hybrids co-existing with pure parental individuals in sympatric localities

(Figure 5A) (although see Freitas et al. 2024; Guiller et al. 2017; Tarroso et al. 2014 for putative F₂ specimens).

If reproductive isolation mechanisms have arisen among most *Vipera* spp., it is striking how the situation differs between *V. latastei* and *V. seoanei*. After diverging from *V. berus*, *V. seoanei* possibly underwent a rapid niche shift towards a less-cold Atlantic habitat in contrast to its boreal sister species, and even adapted to an aestival spermatogenesis, typical in Mediterranean vipers like *V. latastei* (Lucchini et al. 2023). Adaptation to new environments can be achieved by introgressive hybridization or new mutations if such variation does not pre-exist in the species. Introgression, though, leads to faster adaptation than mutation, providing genetic variants that have already been tested by natural selection (Hedrick 2013). Indeed, the highest genomic variance values per individual in both species have been found in their secondary contact zone in NW Iberia (Figure 5B1,B2 and Figure S25). Higher heterozygosity values are often expected in genome regions involved in adaptive introgression, as both correlate with higher recombination rates (Jensen et al. 2023).

Different areas across the genome yield differential effects on fitness when introgressed, shaping what is called the introgression landscape (Martin and Jiggins 2017). Whereas some areas seem to be more prone to accept the movement of genes, others appear devoid of foreign material (“introgression deserts” *sensu* Jensen et al. 2023). For *V. seoanei*, the Z chromosome remains almost free from *V. latastei* introgression (Figure 4C; Figure S17), agreeing with multiple genomic studies that have found evidence of reduced introgression on sex chromosomes (Garrigan et al. 2012; Martin et al. 2013; Sankararaman et al. 2014; Swaegers et al. 2022), suggesting a major role of the Z chromosome in the reproductive isolation of *Vipera* spp. However, some weaker areas of introgression could be simply linked to lower recombination rates (Sankararaman et al. 2014) (e.g., around the interchromosomal translocated fragment, Figure 4C,D), and some introgression deserts found in many autosomes might correspond to centromeres (e.g., chr. 4 and 7 in Figure 4C).

Finally, compelling evidence for the adaptive nature of introgression can come from cases of niche expansion following hybridisation (Wellenreuther et al. 2018). It seems that *V. latastei* populations recently admixed with *V. seoanei* have been able to occupy areas with Atlantic climates that are more typical of *V. seoanei* (Figure 5C and Figure S32). This pattern may support the adaptive nature of this recent introgression event for *V. latastei*. Although we cannot fully rule out the possibility that admixture is just a consequence of the niche expansion, rather than its catalyst, recently admixed *V. seoanei* populations do not appear to occupy a distinct niche compared to their conspecifics.

4.3 | A Fine Balance of Opposing Forces Drives Venom Evolution

The genetic mechanisms behind the origin of an evolutionary novelty like venom are not well understood. Although diversifying selection (increased mutation rate), expression changes, and gene duplication are regarded as the drivers of toxin evolution in snakes, the order of events that lead from a pre-existing protein towards a novel function is unclear (Giorgianni et al. 2020).

Some of the most important toxin families, like PLA₂s and SVMPs, are thought to have been recruited as venom genes after undergoing ancestral duplications (Almeida et al. 2021), which is supported too in *Vipera* by non-venomous paralogs next to venomous copies (Figure S36). Most of the duplicated genes often become nonfunctional, but some of the retained duplicates are thought to be preserved through a wide range of mechanisms: neo-functionalisation, sub-functionalisation, dosage amplification, and back-up compensation (Kuzmin et al. 2022). In neo-functionalisation, one of the copies assumes a new role while the other preserves the ancestral function (Ohno 1970). In contrast, in subfunctionalisation, the two copies divide the ancestral roles between them (also referred as duplication-degeneration-complementation model, by Force et al. 1999). In the context of venom evolution, neofunctionalisation suggests that a gene not originally expressed in the venom gland undergoes duplication and is subsequently recruited by the venom gland. On the other hand, subfunctionalisation proposes that a broadly expressed gene is duplicated, and the expression of one copy becomes restricted to the venom gland—a scenario that seems to be more likely in nature (Hargreaves et al. 2014).

Both neo- and subfunctionalisation rely on gene duplication events preceding the emergence of novel mutations and roles. However, alternative theoretical models have also been proposed, in which innovation occurs prior to duplication, or even in the absence of duplications (Giorgianni et al. 2020). For instance, some toxin genes arose by direct co-option without ancestral duplications, like many secondary families or the SVSPs, whose co-option was facilitated by the activity of transposable elements (Almeida et al. 2021; Perry et al. 2022). Interestingly, four transposable elements were found to be differentially upregulated in the *Vipera* venom gland, with one located next to an SVSP gene, in line with this hypothesis. Overall, different models seem to explain the origin of each toxin family. However, since venom works as an integrated phenotype (Aird et al. 2015), the evolution of their toxin families should be expected to follow some common patterns once they have joined the toxic assembly.

The genus *Vipera* represents a paradigm shift in our understanding of venom evolution, demonstrating that the most abundant toxin families are not necessarily those that have undergone the highest rates of gene duplication or the most accelerated evolutionary change. While it remains unclear whether certain families entirely lacked gene duplications or instead experienced ancestral gene losses, as reported in some rattlesnakes (Margres et al. 2021), this new perspective allows us to more clearly disentangle the respective roles of gene duplication and diversifying selection in shaping venom composition. Hence, evolution accelerates once gene duplication has occurred, that is, there are spare copies to allow innovation (back-up compensation, Kuzmin et al. 2022). Otherwise, purifying selection governs the evolution of these genes, especially the most important ones (Figure 6D), whose maintenance is in turn vital for the individual's fitness. In contrast to diversifying selection, purifying selection has received little attention in snakes, and has only been hypothesised to constrain venom evolution in ancient lineages like cnidarians and scorpions (Sunagar and Moran 2015). However, purification stages might be more widespread than expected, even in relatively “young” lineages like snakes, and especially in the absence of ongoing ecological specialisation.

Author Contributions

A.T., T.M.-B. and S.C. conceived the research, A.T., F.M.-F., A.M. and S.C. designed the research, A.T., F.M.-F., G.M.-R., B.B.-C., M.E., D.F.-G., A.C., S.U., J.V., B.H. and S.C. collected samples, A.T., D.P. and J.J.C. performed the laboratory work, A.T. and M.P.-F. assembled the genomes, A.T. and F.M.-F. analyzed the data, A.T. wrote the paper, all authors reviewed the manuscript and approved its final version.

Acknowledgements

We must thank Juan Pablo González de la Vega, Juan Timms, Rodrigo Bustos, Emiliano Mori, María Casero (Centro de Recuperação e Investigação de Animais Selvagens (RIAS)–Associação ALDEIA (Portugal)), Ricardo Brandão (Centro de Ecologia, Recuperação e Vigilância de Animais Selvagens (CERVAS)–Associação ALDEIA (Portugal)), Luis Guilherme Sousa (HerpEborá (Portugal)), Alberto Sánchez-Vialas and Marta Calvo (Museo Nacional de Ciencias Naturales—CSIC (Madrid, Spain)), Inês Freitas, Nahla Lucchini, Soumia Fahd, Abdellah Bouazza, Jon Buldain, Ismael Espasandín, Fèlix Amat, David Donaire, Neftali Sillero, João Campos, Víctor Ortega, Marc Albiac, Egoitz Ikaza, Aritz Ibarzabal, and Paco Moreno for their help in sample collection. We thank as well Jonathan Wood for his thorough comments on genome assembly manual curation, Klara Eleftheriadi for her kind help in genome assembly and annotation, Ignazio Avella for providing the *V. seoanei*'s proteome, and Ignacio O'Mullony for his graphic design advice. Also, we thank the CNAG genome assembly team, Tyler Alioto, Jessica Gómez and Fernando Cruz for their comments on genome assembly, and Sandra Ruibal and Laia Llovera for their help in WGS library preparation.

Ethics Statement

Sampling Permits: Samplings were authorised in: Morocco by the Haut Commissariat aux Eaux et Forêts et à la Lutte Contre la Désertification of Morocco, with permission numbers HCEFLCD/DLCPDN/DPRN/CFF N° 20/2013, 19/2015, 35/2018 and 12/2022; Portugal by the Instituto da Conservação da Natureza e das Florestas with permission references 537/2018, 295/2020 and 326/2022; Andalucía by Consejería de Agricultura, Ganadería, Pesca y Desarrollo Sostenible from Junta de Andalucía, with permission reference GYB/DBP; Álava by Departamento de Medio Ambiente y Urbanismo from Arabako Foru Aldundia/Diputación Foral de Álava, with permission number 2021/017; Aragón by Departamento de Agricultura, Ganadería y Medio Ambiente from Gobierno de Aragón, with verification code CSV3-6PK6D-0KYB4-FWREG; Principado de Asturias and Picos de Europa National Park by Consejería de Medio Rural y Cohesión Territorial from Gobierno del Principado de Asturias, with permission numbers DECO/2021/9850 and CO/09/0230/2021, respectively; Bizkaia by Departamento de Sostenibilidad y Medio Ambiente from Diputación Foral de Bizkaia with permission number AM-30-2021; Cantabria by Consejería de Desarrollo Rural, Ganadería, Pesca, Alimentación y Medio Ambiente from Gobierno de Cantabria, with registration number 2021GA001S023262; Castilla-La Mancha by Dirección General de Medio Natural y Biodiversidad from Consejería de Desarrollo Sostenible, with permission number DGMNB/SEN/avp_21_204; Castilla y León by Consejería de Fomento y Medio Ambiente from Junta de Castilla y León, with permission numbers EP/CyL/94/2016, 56/2018, 192/2020 and AUES/CyL/198/2021; Catalonia and Els Ports Natural Park by Direcció General de Polítiques Ambientals i Medi Natural from Generalitat de Catalunya, with permission numbers SF/0118/2019 and SF/761, respectively; Extremadura by Consejería para la Transición Ecológica y Sostenibilidad from Junta de Extremadura, with permission number CN0034/21/ACA; Galicia by Consellería de Medio Ambiente, Territorio e Vivenda from Xunta de Galicia, with permission numbers EB-017/2019, 018/2020, 015/2021, and 048/2021; Comunidad de Madrid by Consejería de Medio Ambiente, Vivienda y Agricultura from Gobierno de la Comunidad de Madrid, with permission number

10/283514.9/21; Comunitat Valenciana by Conselleria de Agricultura, Desarrollo Rural, Emergencia Climática y Transición Ecológica from Generalitat Valenciana, with permission number 926/2021-VS (FAU 31_21); Comunidad Foral de Navarra by Departamento de Desarrollo Rural y Medio Ambiente from Gobierno de Navarra, with reference 177E/2021; Región de Murcia by Consejería de Agua, Agricultura, Ganadería Pesca y Medio Ambiente from Gobierno de la Región de Murcia, with permission number AUF/2021/045. Euthanasia of selected individuals was authorised by Consejería de Medio Ambiente, Vivienda y Ordenación del Territorio from Junta de Castilla y León with code AUES/CYL/228/2023 for *Vipera latastei* and *V. seoanei*, and Direcció General de Polítiques Ambientals i Medi Natural from Generalitat de Catalunya with code SF/0199/22 for *V. aspis*.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data from this study is available as of the date of publication in the NCBI BioProjects PRJNA838200 for *V. latastei* genome assembly, and PRJNA1284094 for other assemblies and raw data. All scripts used for the analyses described in this study, and the *V. latastei* proteomes, are publicly available at: <https://github.com/adtalave/EuropeanVipersGenomics>.

References

- Aird, S. D., S. Aggarwal, A. Villar-Briones, M. M.-Y. Tin, K. Terada, and A. S. Mikhayev. 2015. "Snake Venoms Are Integrated Systems, but Abundant Venom Proteins Evolve More Rapidly." *BMC Genomics* 16, no. 1: 647. <https://doi.org/10.1186/s12864-015-1832-6>.
- Alexander, D. H., J. Novembre, and K. Lange. 2009. "Fast Model-Based Estimation of Ancestry in Unrelated Individuals." *Genome Research* 19, no. 9: 1655–1664. <https://doi.org/10.1101/gr.094052.109>.
- Allio, R., A. Schomaker-Bastos, J. Romiguier, F. Prosdocimi, B. Nabholz, and F. Delsuc. 2020. "MitoFinder: Efficient Automated Large-Scale Extraction of Mitogenomic Data in Target Enrichment Phylogenomics." *Molecular Ecology Resources* 20, no. 4: 892–905. <https://doi.org/10.1111/1755-0998.13160>.
- Almeida, D. D., V. L. Viala, P. G. Nachtigall, et al. 2021. "Tracking the Recruitment and Evolution of Snake Toxins Using the Evolutionary Context Provided by the *Bothrops jararaca* Genome." *Proceedings of the National Academy of Sciences of the United States of America* 118, no. 20: e2015159118. <https://doi.org/10.1073/pnas.2015159118>.
- Aoki, N., A. Sakiyama, K. Kuroki, et al. 2008. "Serotriffin, a CRISP Family Protein With Binding Affinity for Small Serum Protein-2 in Snake Serum." *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1784, no. 4: 621–628. <https://doi.org/10.1016/j.bbapap.2007.12.010>.
- Apra, G., A. Gentilli, M. Zuffi, and G. Odierna. 2006. "The Karyology of *Vipera Aspis*, *V. Atrata*, *V. Hugi*, and *Cerastes Vipera*." *Amphibia-Reptilia* 27, no. 1: 113–119. <https://doi.org/10.1163/156853806776052209>.
- Avella, I., M. Damm, I. Freitas, et al. 2023. "One Size Fits All—Venomics of the Iberian Adder (*Vipera Seoanei*, Latase 1878) Reveals Low Levels of Venom Variation Across Its Distributional Range." *Toxins* 15, no. 6: 371. <https://doi.org/10.3390/toxins15060371>.
- Barbanera, F., M. A. L. Zuffi, M. Guerrini, et al. 2009. "Molecular Phylogeography of the Asp Viper *Vipera aspis* (Linnaeus, 1758) in Italy: Evidence for Introgressive Hybridization and Mitochondrial DNA Capture." *Molecular Phylogenetics and Evolution* 52, no. 1: 103–114. <https://doi.org/10.1016/j.ympev.2009.02.006>.
- Beçak, W., and M. L. Beçak. 2008. "Cytotaxonomy and Chromosomal Evolution in Serpents." *Cytogenetics* 8, no. 4: 247–262. <https://doi.org/10.1159/000130037>.
- Blonder, B., C. Lamanna, C. Violle, and B. J. Enquist. 2014. "The n-Dimensional Hypervolume." *Global Ecology and Biogeography* 23, no. 5: 595–609. <https://doi.org/10.1111/geb.12146>.
- Bouckaert, R., J. Heled, D. Kühnert, et al. 2014. "BEAST 2: A Software Platform for Bayesian Evolutionary Analysis." *PLoS Computational Biology* 10, no. 4: e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>.
- Brüna, T., K. J. Hoff, A. Lomsadze, M. Stanke, and M. Borodovsky. 2021. "BRAKER2: Automatic Eukaryotic Genome Annotation With GeneMark-EP+ and AUGUSTUS Supported by a Protein Database." *NAR Genomics and Bioinformatics* 3, no. 1: lqaa108. <https://doi.org/10.1093/nargab/lqaa108>.
- Calvete, J. J., C. Marcinkiewicz, D. Monleón, et al. 2005. "Snake Venom Disintegrins: Evolution of Structure and Function." *Toxicon* 45, no. 8: 1063–1074. <https://doi.org/10.1016/j.toxicon.2005.02.024>.
- Carøe, C., S. Gopalakrishnan, L. Vinner, et al. 2018. "Single-Tube Library Preparation for Degraded DNA." *Methods in Ecology and Evolution* 9, no. 2: 410–419. <https://doi.org/10.1111/2041-210X.12871>.
- Casewell, N. R., S. C. Wagstaff, R. A. Harrison, C. Renjifo, and W. Wüster. 2011. "Domain Loss Facilitates Accelerated Evolution and Neofunctionalization of Duplicate Snake Venom Metalloproteinase Toxin Genes." *Molecular Biology and Evolution* 28, no. 9: 2637–2649. <https://doi.org/10.1093/molbev/msr091>.
- Ceballos, F. C., P. K. Joshi, D. W. Clark, M. Ramsay, and J. F. Wilson. 2018. "Runs of Homozygosity: Windows Into Population History and Trait Architecture." *Nature Reviews Genetics* 19, no. 4: 220–234. <https://doi.org/10.1038/nrg.2017.109>.
- Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. "fastp: An Ultra-Fast All-in-One FASTQ Preprocessor." *Bioinformatics* 34, no. 17: i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Chowdhury, A., C. N. Zdenek, and B. G. Fry. 2022. "Diverse and Dynamic Alpha-Neurotoxicity Within Venoms From the Palearctic Viperid Snake Clade of *Daboia*, *Macrovipera*, *Montivipera*, and *Vipera*." *Neurotoxicity Research* 40, no. 6: 1793–1801. <https://doi.org/10.1007/s12640-022-00572-w>.
- Damm, M., B.-F. Hempel, and R. D. Süßmuth. 2021. "Old World Vipers—A Review About Snake Venom Proteomics of Viperinae and Their Variations." *Toxins* 13, no. 6: 427. <https://doi.org/10.3390/toxin13060427>.
- Danecek, P., A. Auton, G. Abecasis, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27, no. 15: 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- de Vos, J. M., H. Augustijnen, L. Batscher, and K. Lucek. 2020. "Speciation Through Chromosomal Fusion and Fission in Lepidoptera." *Philosophical Transactions of the Royal Society B* 375, no. 1806: 20190539. <https://doi.org/10.1098/rstb.2019.0539>.
- Doniol-Valcroze, P., S. Ursenbacher, K. Mebert, et al. 2021. "Conflicting Relationships of *Vipera walser* Inferred From Nuclear Genes Sequences and Mitochondrial DNA." *Journal of Zoological Systematics and Evolutionary Research* 59, no. 8: 2307–2320. <https://doi.org/10.1111/jzs.12543>.
- Dufresnes, C., S. Ghielmi, B. Halpern, et al. 2024. "Phylogenomic Insights Into the Diversity and Evolution of Palearctic Vipers." *Molecular Phylogenetics and Evolution* 197: 108095. <https://doi.org/10.1016/j.ympev.2024.108095>.
- Faria, R., and A. Navarro. 2010. "Chromosomal Speciation Revisited: Rearranging Theory With Pieces of Evidence." *Trends in Ecology & Evolution* 25, no. 11: 660–669. <https://doi.org/10.1016/j.tree.2010.07.008>.
- Ferreira, I. G., M. B. Pucca, I. S. de Oliveira, F. A. Cerni, B. d. C. da Silva Jacob, and E. C. Arantes. 2021. "Snake Venom Vascular Endothelial Growth Factors (svVEGFs): Unravelling Their Molecular Structure,

- Functions, and Research Potential." *Cytokine & Growth Factor Reviews* 60: 133–143. <https://doi.org/10.1016/j.cytogfr.2021.05.003>.
- Fick, S. E., and R. J. Hijmans. 2017. "Worldclim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37: 4302–4315.
- Fitak, R. R. 2021. "OptM: Estimating the Optimal Number of Migration Edges on Population Trees Using Treemix." *Biology Methods & Protocols* 6, no. 1: bpab017. <https://doi.org/10.1093/biomethods/bpab017>.
- Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. Yan, and J. Postlethwait. 1999. "Preservation of Duplicate Genes by Complementary, Degenerative Mutations." *Genetics* 151, no. 4: 1531–1545. <https://doi.org/10.1093/genetics/151.4.1531>.
- Formenti, G., L. Abueg, A. Brajuka, et al. 2022. "Gfastats: Conversion, Evaluation and Manipulation of Genome Sequences Using Assembly Graphs." *Bioinformatics* 38, no. 17: 4214–4216. <https://doi.org/10.1093/bioinformatics/btac460>.
- François, D., S. Ursenbacher, A. Boissinot, F. Ysnel, and O. Lourdais. 2021. "Isolation-By-Distance and Male-Biased Dispersal at a Fine Spatial Scale: A Study of the Common European Adder (*Vipera berus*) in a Rural Landscape." *Conservation Genetics* 22, no. 5: 823–837. <https://doi.org/10.1007/s10592-021-01365-y>.
- Freitas, I., P. Tarroso, Ó. Zuazo, et al. 2023. "Local Niches Explain Coexistence in Environmentally-Distinct Contact Zones Between Western Mediterranean Vipers." *Scientific Reports* 13, no. 1: 21113. <https://doi.org/10.1038/s41598-023-48204-3>.
- Freitas, I., S. Ursenbacher, K. Mebert, et al. 2020. "Evaluating Taxonomic Inflation: Towards Evidence-Based Species Delimitation in Eurasian Vipers (Serpentes: Viperinae)." *Amphibia-Reptilia* 41, no. 3: 285–311. <https://doi.org/10.1163/15685381-bja10007>.
- Freitas, I., G. Velo-Antón, A. Kaliontzopoulou, Ó. Zuazo, and F. Martínez-Freiria. 2024. "Association Between Genetic Admixture and Morphological Patterns in a Hybrid Zone Between the Two Iberian Vipers, *Vipera aspis* and *V. latastei*." *Journal of Zoological Systematics and Evolutionary Research* 2024, no. 1: 3800363. <https://doi.org/10.1155/2024/3800363>.
- Gabriel, L., K. J. Hoff, T. Brůna, M. Borodovsky, and M. Stanke. 2021. "TSEBRA: Transcript Selector for BRAKER." *BMC Bioinformatics* 22, no. 1: 566. <https://doi.org/10.1186/s12859-021-04482-0>.
- Garrigan, D., S. B. Kingan, A. J. Geneva, et al. 2012. "Genome Sequencing Reveals Complex Speciation in the *Drosophila simulans* Clade." *Genome Research* 22, no. 8: 1499–1511. <https://doi.org/10.1101/gr.130922.111>.
- Giorgianni, M. W., N. L. Dowell, S. Griffin, V. A. Kassner, J. E. Selegue, and S. B. Carroll. 2020. "The Origin and Diversification of a Novel Protein Family in Venomous Snakes." *Proceedings of the National Academy of Sciences* 117, no. 20: 10911–10920. <https://doi.org/10.1073/pnas.1920011117>.
- Giribaldi, J., T. Kazandjian, F. G. Amorim, et al. 2020. "Venomics of the Asp Viper *Vipera aspis aspis* From France." *Journal of Proteomics* 218: 103707. <https://doi.org/10.1016/j.jprot.2020.103707>.
- Göçmen, B., P. Heiss, D. Petras, A. Nalbantsoy, and R. D. Süßmuth. 2015. "Mass Spectrometry Guided Venom Profiling and Bioactivity Screening of the Anatolian Meadow Viper, *Vipera anatolica*." *Toxicon* 107: 163–174. <https://doi.org/10.1016/j.toxicon.2015.09.013>.
- Gorman, G. C., and F. Gress. 1970. "Chromosome Cytology of Four Boid Snakes and a Varanid Lizard, With Comments on the Cytosystematics of Primitive Snakes." *Herpetologica* 26, no. 3: 308–317. <http://www.jstor.org/stable/3891257>.
- Guiller, G., O. Lourdais, and S. Ursenbacher. 2017. "Hybridization Between a Euro-Siberian (*Vipera berus*) and a Para-Mediterranean Viper (*V. aspis*) at Their Contact Zone in Western France." *Journal of Zoology* 302, no. 2: 138–147. <https://doi.org/10.1111/jzo.12431>.
- Gutiérrez, J. M., J. J. Calvete, A. G. Habib, R. A. Harrison, D. J. Williams, and D. A. Warrell. 2017. "Snakebite Envenoming." *Nature Reviews Disease Primers* 3, no. 1: 17063. <https://doi.org/10.1038/nrdp.2017.63>.
- Hargreaves, A. D., M. T. Swain, M. J. Hegarty, D. W. Logan, and J. F. Mulley. 2014. "Restriction and Recruitment—Gene Duplication and the Origin and Evolution of Snake Venom Toxins." *Genome Biology and Evolution* 6, no. 8: 2088–2095. <https://doi.org/10.1093/gbe/evu166>.
- Hedrick, P. W. 2013. "Adaptive Introgression in Animals: Examples and Comparison to New Mutation and Standing Variation as Sources of Adaptive Variation." *Molecular Ecology* 22, no. 18: 4606–4618. <https://doi.org/10.1111/mec.12415>.
- Hennig, C., and B. Hausdorf. 2019. Package 'Prabclus', Version 2.3–1. Functions for Clustering of Presence-Absence, Abundance and Multilocus Genetic Data. <http://cran.r-project.org/web/packages/prabclus/>.
- Hoff, K. J., S. Lange, A. Lomsadze, M. Borodovsky, and M. Stanke. 2016. "BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation With GeneMark-ET and AUGUSTUS." *Bioinformatics* 32, no. 5: 767–769. <https://doi.org/10.1093/bioinformatics/btv661>.
- Hogner, S., T. Laskemoen, J. T. Lifjeld, et al. 2012. "Deep Sympatric Mitochondrial Divergence Without Reproductive Isolation in the Common Redstart *Phoenicurus phoenicurus*." *Ecology and Evolution* 2, no. 12: 2974–2988. <https://doi.org/10.1002/ece3.398>.
- Jensen, A., F. Swift, D. de Vries, et al. 2023. "Complex Evolutionary History With Extensive Ancestral Gene Flow in an African Primate Radiation." *Molecular Biology and Evolution* 40, no. 12: msad247. <https://doi.org/10.1093/molbev/msad247>.
- Jin, J.-J., W.-B. Yu, J.-B. Yang, et al. 2020. "GetOrganelle: A Fast and Versatile Toolkit for Accurate de Novo Assembly of Organelle Genomes." *Genome Biology* 21, no. 1: 241. <https://doi.org/10.1186/s13059-020-02154-5>.
- Juárez, P., I. Comas, F. González-Candelas, and J. J. Calvete. 2008. "Evolution of Snake Venom Disintegrins by Positive Darwinian Selection." *Molecular Biology and Evolution* 25, no. 11: 2391–2407. <https://doi.org/10.1093/molbev/msn179>.
- Keilwagen, J., M. Wenk, J. L. Erickson, M. H. Schattat, J. Grau, and F. Hartung. 2016. "Using Intron Position Conservation for Homology-Based Gene Prediction." *Nucleic Acids Research* 44, no. 9: e89. <https://doi.org/10.1093/nar/gkw092>.
- Kerkkamp, H. M. I., R. M. Kini, A. S. Pospelov, F. J. Vonk, C. V. Henkel, and M. K. Richardson. 2016. "Snake Genome Sequencing: Results and Future Prospects." *Toxins* 8, no. 12: 360. <https://doi.org/10.3390/toxin8120360>.
- Kielbasa, S. M., R. Wan, K. Sato, P. Horton, and M. C. Frith. 2011. "Adaptive Seeds Tame Genomic Sequence Comparison." *Genome Research* 21, no. 3: 487–493. <https://doi.org/10.1101/gr.113985.110>.
- Kosakovsky Pond, S. L., and S. D. W. Frost. 2005. "Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection." *Molecular Biology and Evolution* 22, no. 5: 1208–1222. <https://doi.org/10.1093/molbev/msi105>.
- Kuzmin, E., J. S. Taylor, and C. Boone. 2022. "Retention of Duplicated Genes in Evolution." *Trends in Genetics* 38, no. 1: 59–72. <https://doi.org/10.1016/j.tig.2021.06.016>.
- Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2017. "PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses." *Molecular Biology and Evolution* 34, no. 3: 772–773. <https://doi.org/10.1093/molbev/msw260>.
- Li, H. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs With BWA-MEM." *ArXiv*: 1303.3997v2. <https://doi.org/10.48550/arXiv.1303.3997>.
- Love, M., S. Anders, and W. Huber. 2014. "Differential Analysis of Count Data—the DESeq2 Package." *Genome Biology* 15, no. 550: 10–1186. <https://doi.org/10.1186/s13059-014-0550-8>.

- Lucchini, N., A. Kaliontzopoulou, O. Lourdaïs, and F. Martínez-Freiría. 2023. "Climatic Adaptation Explains Responses to Pleistocene Oscillations and Diversification in European Vipers." *Journal of Biogeography* 50, no. 11: 1838–1851. <https://doi.org/10.1111/jbi.14694>.
- Luiselli, L., E. Filippi, and E. Di Lena. 2007. "Ecological Relationships Between Sympatric *Vipera Aspis* and *Vipera ursinii* in High-Altitude Habitats of Central Italy." *Journal of Herpetology* 41, no. 3: 378–384. [https://doi.org/10.1670/0022-1511\(2007\)41\[378:ERBSVA\]2.0.CO;2](https://doi.org/10.1670/0022-1511(2007)41[378:ERBSVA]2.0.CO;2).
- Lynch, V. J. 2007. "Inventing an Arsenal: Adaptive Evolution and Neofunctionalization of Snake Venom Phospholipase A₂ Genes." *BMC Evolutionary Biology* 7, no. 1: 2. <https://doi.org/10.1186/1471-2148-7-2>.
- Malinsky, M., M. Matschiner, and H. Svoldal. 2021. "Dsuite-Fast D-Statistics and Related Admixture Evidence From VCF Files." *Molecular Ecology Resources* 21, no. 2: 584–595. <https://doi.org/10.1111/1755-0998.13265>.
- Margres, M. J., R. M. Rautsaw, J. L. Strickland, et al. 2021. "The Tiger Rattlesnake Genome Reveals a Complex Genotype Underlying a Simple Venom Phenotype." *PNAS* 118, no. 4: e2014634118. <https://doi.org/10.1073/pnas.2014634118>.
- Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, et al. 2013. "Genome-Wide Evidence for Speciation With Gene Flow in *Heliconius* Butterflies." *Genome Research* 23, no. 11: 1817–1828. <https://doi.org/10.1101/gr.159426.113>.
- Martin, S. H., and C. D. Jiggins. 2017. "Interpreting the Genomic Landscape of Introgression." *Current Opinion in Genetics & Development* 47: 69–74. <https://doi.org/10.1016/j.gde.2017.08.007>.
- Martin, S. H., and S. M. Van Belleghem. 2017. "Exploring Evolutionary Relationships Across the Genome Using Topology Weighting." *Genetics* 206, no. 1: 429–438. <https://doi.org/10.1534/genetics.116.194720>.
- Martínez-Freiría, F., I. Freitas, G. Velo-Antón, et al. 2021. "Integrative Taxonomy Reveals Two Species and Intraspecific Differentiation in the *Vipera latastei*–*Monticola* Complex." *Journal of Zoological Systematics and Evolutionary Research* 59, no. 8: 2278–2306. <https://doi.org/10.1111/jzs.12534>.
- Martínez-Freiría, F., I. Freitas, M. A. L. Zuffi, P. Golay, S. Ursenbacher, and G. Velo-Antón. 2020. "Climatic Refugia Boosted Allopatric Diversification in Western Mediterranean Vipers." *Journal of Biogeography* 47, no. 8: 1698–1713. <https://doi.org/10.1111/jbi.13861>.
- Martínez-Freiría, F., M. Lizana, J. P. do Amaral, and J. C. Brito. 2010. "Spatial and Temporal Segregation Allows Coexistence in a Hybrid Zone Among Two Mediterranean Vipers (*Vipera Aspis* and *V. Latastei*)." *Amphibia-Reptilia* 31, no. 2: 195–212. <https://doi.org/10.1163/156853810791069001>.
- Martínez-Freiría, F., N. Sillero, M. Lizana, and J. C. Brito. 2008. "GIS-Based Niche Models Identify Environmental Correlates Sustaining a Contact Zone Between Three Species of European Vipers." *Diversity and Distributions* 14, no. 3: 452–461. <https://doi.org/10.1111/j.1472-4642.2007.00446.x>.
- Martínez-Freiría, F., G. Velo-Antón, and J. C. Brito. 2015. "Trapped by Climate: Interglacial Refuge and Recent Population Expansion in the Endemic Iberian Adder *Vipera seoanei*." *Diversity and Distributions* 21, no. 3: 331–344. <https://doi.org/10.1111/ddi.12265>.
- Mebert, K., T. Jagar, R. Grželj, et al. 2015. "The Dynamics of Coexistence: Habitat Sharing Versus Segregation Patterns Among Three Sympatric Montane Vipers." *Biological Journal of the Linnean Society* 116, no. 2: 364–376. <https://doi.org/10.1111/bij.12582>.
- Mebert, K., L. Luiselli, V. Cafuta, P. Golay, S. Dubey, and S. Ursenbacher. 2017. "A Home for Three: Analyzing Ecological Correlates of Body Traits in a Triple Contact Zone of Alpine Vipers." *North-Western Journal of Zoology* 13, no. 2: 251–261.
- Mizsei, E., D. Jablonski, S. A. Roussos, et al. 2017. "Nuclear Markers Support the Mitochondrial Phylogeny of *Vipera Ursinii*–*Renardi* Complex (Squamata: Viperidae) and Species Status for the Greek Meadow Viper." *Zootaxa* 4227, no. 1: 75–88. <https://doi.org/10.11646/zootaxa.4227.1.4>.
- Mochales-Riaño, G., B. Burriel-Carranza, M. I. Barros, et al. 2024. "Hidden in the Sand: Phylogenomics Unravel an Unexpected Evolutionary History for the Desert-Adapted Vipers of the Genus *Cerastes*." *Molecular Phylogenetics and Evolution* 191: 107979. <https://doi.org/10.1016/j.ympev.2023.107979>.
- Mochales-Riaño, G., C. Fontseré, M. de Manuel, et al. 2023. "Genomics Reveals Introgression and Purging of Deleterious Mutations in the Arabian Leopard (*Panthera pardus nimr*)." *IScience* 26, no. 9: 107481. <https://doi.org/10.1016/j.isci.2023.107481>.
- Mochales-Riaño, G., S. R. Hirst, A. Talavera, et al. 2024. "Chromosome-Level Reference Genome for the Medically Important Arabian Horned Viper (*Cerastes Gasperettii*)." *BioRxiv* 14: 2024.07.29.605543. <https://doi.org/10.1101/2024.07.29.605543>.
- Murrell, B., S. Weaver, M. D. Smith, et al. 2015. "Gene-Wide Identification of Episodic Selection." *Molecular Biology and Evolution* 32, no. 5: 1365–1371. <https://doi.org/10.1093/molbev/msv035>.
- Myers, E. A., J. L. Strickland, R. M. Rautsaw, et al. 2022. "De Novo Genome Assembly Highlights the Role of Lineage-Specific Gene Duplications in the Evolution of Venom in Fea's Viper (*Azemiops Feae*)." *Genome Biology and Evolution* 14, no. 7: evac082. <https://doi.org/10.1093/gbe/evac082>.
- Narasimhan, V., P. Danecek, A. Scally, Y. Xue, C. Tyler-Smith, and R. Durbin. 2016. "BCFtools/RoH: A Hidden Markov Model Approach for Detecting Autozygosity From Next-Generation Sequencing Data." *Bioinformatics* 32, no. 11: 1749–1751. <https://doi.org/10.1093/bioinformatics/btw044>.
- Ohno, S. 1970. *Evolution by Gene Duplication*. Springer Verlag.
- Ottensburghs, J. 2020. "Ghost Introgression: Spooky Gene Flow in the Distant Past." *BioEssays* 42, no. 6: 2000012. <https://doi.org/10.1002/bies.202000012>.
- Pasquesi, G. I. M., R. H. Adams, D. C. Card, et al. 2018. "Squamate Reptiles Challenge Paradigms of Genomic Repeat Element Evolution Set by Birds and Mammals." *Nature Communications* 9, no. 1: 2774. <https://doi.org/10.1038/s41467-018-05279-1>.
- Perry, B. W., S. S. Gopalan, G. I. M. Pasquesi, et al. 2022. "Snake Venom Gene Expression Is Coordinated by Novel Regulatory Architecture and the Integration of Multiple Co-Opted Vertebrate Pathways." *Genome Research* 32, no. 6: 1058–1073. <https://doi.org/10.1101/gr.276251.121>.
- Pickrell, J., and J. Pritchard. 2012. "Inference of Population Splits and Mixtures From Genome-Wide Allele Frequency Data." *PLoS Genetics* 8, no. 11: e1002967. <https://doi.org/10.1371/journal.pgen.1002967>.
- Poplin, R., V. Ruano-Rubio, M. A. DePristo, et al. 2018. "Scaling Accurate Genetic Variant Discovery to Tens of Thousands of Samples." *BioRxiv*: 201178. <https://doi.org/10.1101/201178>.
- Rao, W., K. Kalogeropoulos, M. E. Allentoft, et al. 2022. "The Rise of Genomics in Snake Venom Research: Recent Advances and Future Perspectives." *GigaScience* 11: giac024. <https://doi.org/10.1093/gigascience/giac024>.
- Saint Girons, H., and R. Duguy. 1976. "Écologie et Position Systématique de *Vipera seoanei* Lataste, 1879." *Bulletin de la Société Zoologique de France* 101, no. 2: 325–339.
- Sajevic, T., A. Leonardi, and I. Križaj. 2014. "An Overview of Hemostatically Active Components of *Vipera Ammodytes Ammodytes* Venom." *Toxin Reviews* 33, no. 1–2: 33–36. <https://doi.org/10.3109/15569543.2013.835827>.
- Sankararaman, S., S. Mallick, M. Dannemann, et al. 2014. "The Genomic Landscape of Neanderthal Ancestry in Present-Day Humans." *Nature* 507, no. 7492: 354–357. <https://doi.org/10.1038/nature12961>.

- Schild, D. R., D. C. Card, N. R. Hales, et al. 2019. "The Origins and Evolution of Chromosomes, Dosage Compensation, and Mechanisms Underlying Venom Regulation in Snakes." *Genome Research* 29, no. 4: 590–601.
- Schild, D. R., B. W. Perry, R. H. Adams, et al. 2022. "The Roles of Balancing Selection and Recombination in the Evolution of Rattlesnake Venom." *Nature Ecology & Evolution* 6, no. 9: 1367–1380. <https://doi.org/10.1038/s41559-022-01829-5>.
- Shibata, H., T. Chijiwa, N. Oda-Ueda, et al. 2018. "The Habu Genome Reveals Accelerated Evolution of Venom Protein Genes." *Scientific Reports* 8, no. 1: 11300. <https://doi.org/10.1038/s41598-018-28749-4>.
- Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness With Single-Copy Orthologs." *Bioinformatics* 31, no. 19: 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Singh, L. 1972. "Evolution of Karyotypes in Snakes." *Chromosoma* 38, no. 2: 185–236. <https://doi.org/10.1007/BF00326193>.
- Šmíd, J., and K. A. Tolley. 2019. "Calibrating the Tree of Vipers Under the Fossilized Birth-Death Model." *Scientific Reports* 9, no. 1: 5510. <https://doi.org/10.1038/s41598-019-41290-2>.
- Spangenberg, V., I. Redekop, S. A. Simanovsky, and O. Kolomiets. 2022. "Cytogenetic Analysis of the Bimodal Karyotype of the Common European Adder, *Vipera Berus* (Viperidae)." *Animals* 12, no. 24: 3563. <https://doi.org/10.3390/ani12243563>.
- Suc, J.-P. 1984. "Origin and Evolution of the Mediterranean Vegetation and Climate in Europe." *Nature* 307, no. 5950: 429–432. <https://doi.org/10.1038/307429a0>.
- Sunagar, K., B. G. Fry, T. N. W. Jackson, et al. 2013. "Molecular Evolution of Vertebrate Neurotrophins: Co-Option of the Highly Conserved Nerve Growth Factor Gene Into the Advanced Snake Venom Arsenal." *PLoS One* 8, no. 11: e81827. <https://doi.org/10.1371/journal.pone.0081827>.
- Sunagar, K., W. E. Johnson, S. J. O'Brien, V. Vasconcelos, and A. Antunes. 2012. "Evolution of CRISPs Associated With Toxicoforan-Reptilian Venom and Mammalian Reproduction." *Molecular Biology and Evolution* 29, no. 7: 1807–1822. <https://doi.org/10.1093/molbev/mss058>.
- Sunagar, K., and Y. Moran. 2015. "The Rise and Fall of an Evolutionary Innovation: Contrasting Strategies of Venom Evolution in Ancient and Young Animals." *PLoS Genetics* 11, no. 10: e1005596. <https://doi.org/10.1371/journal.pgen.1005596>.
- Suryamohan, K., S. P. Krishnankutty, J. Guillory, et al. 2020. "The Indian Cobra Reference Genome and Transcriptome Enables Comprehensive Identification of Venom Toxins." *Nature Genetics* 52, no. 1: 106–117. <https://doi.org/10.1038/s41588-019-0559-8>.
- Swaegers, J., R. A. Sánchez-Guillén, P. Chauhan, M. Wellenreuther, and B. Hansson. 2022. "Restricted X Chromosome Introgression and Support for Haldane's Rule in Hybridizing Damselflies." *Proceedings of the Royal Society B: Biological Sciences* 289, no. 1979: 20220968. <https://doi.org/10.1098/rspb.2022.0968>.
- Tang, H., J. E. Bowers, X. Wang, R. Ming, M. Alam, and A. H. Paterson. 2008. "Synteny and Collinearity in Plant Genomes." *Science* 320, no. 5875: 486–488. <https://doi.org/10.1126/science.1153917>.
- Tang, S., A. Lomsadze, and M. Borodovsky. 2015. "Identification of Protein Coding Regions in RNA Transcripts." *Nucleic Acids Research* 43, no. 12: e78. <https://doi.org/10.1093/nar/gkv227>.
- Tarroso, P., R. J. Pereira, F. Martínez-Freiria, R. Godinho, and J. C. Brito. 2014. "Hybridization at an Ecotone: Ecological and Genetic Barriers Between Three Iberian Vipers." *Molecular Ecology* 23, no. 5: 1108–1123.
- Taylor, S. A., and E. L. Larson. 2019. "Insights From Genomes Into the Evolutionary Importance and Prevalence of Hybridization in Nature." *Nature Ecology & Evolution* 3, no. 2: 170–177. <https://doi.org/10.1038/s41559-018-0777-y>.
- Thanou, E., D. Jablonski, and P. Kornilios. 2023. "Genome-Wide Single Nucleotide Polymorphisms Reveal Recurrent Waves of Speciation in Niche-Pockets, in Europe's Most Venomous Snake." *Molecular Ecology* 32, no. 13: 3624–3640. <https://doi.org/10.1111/mec.16944>.
- Toews, D. P. L., and A. Brelsford. 2012. "The Biogeography of Mitochondrial and Nuclear Discordance in Animals." *Molecular Ecology* 21, no. 16: 3907–3930. <https://doi.org/10.1111/j.1365-294X.2012.05664.x>.
- Torres-Roig, E., K. J. Mitchell, J. A. Alcover, et al. 2021. "Origin, Extinction and Ancient DNA of a New Fossil Insular Viper: Molecular Clues of Overseas Immigration." *Zoological Journal of the Linnean Society* 192, no. 1: 144–168. <https://doi.org/10.1093/zoolinnean/zlaa094>.
- Ursenbacher, S., M. Carlsson, V. Helfer, H. Tegelström, and L. Fumagalli. 2006. "Phylogeography and Pleistocene Refugia of the Adder (*Vipera berus*) as Inferred From Mitochondrial DNA Sequence Data." *Molecular Ecology* 15, no. 11: 3425–3437. <https://doi.org/10.1111/j.1365-294X.2006.03031.x>.
- Vörös, J., S. Ursenbacher, D. Jelić, et al. 2022. "Well-Known Species, Unexpected Results: High Genetic Diversity in Declining *Vipera Ursinii* in Central, Eastern and Southeastern Europe." *Amphibia-Reptilia* 43, no. 4: 407–423. <https://doi.org/10.1163/15685381-bja10116>.
- Weaver, S., S. D. Shank, S. J. Spielman, M. Li, S. V. Muse, and S. L. Kosakovsky Pond. 2018. "Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes." *Molecular Biology and Evolution* 35, no. 3: 773–777. <https://doi.org/10.1093/molbev/msx335>.
- Wellenreuther, M., J. Muñoz, J. R. Chávez-Ríos, B. Hansson, A. Cordero-Rivera, and R. A. Sánchez-Guillén. 2018. "Molecular and Ecological Signatures of an Expanding Hybrid Zone." *Ecology and Evolution* 8, no. 10: 4793–4806. <https://doi.org/10.1002/ece3.4024>.
- Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Zhang, C., S.-S. Dong, J.-Y. Xu, W.-M. He, and T.-L. Yang. 2019. "PopLDdecay: A Fast and Effective Tool for Linkage Disequilibrium Decay Analysis Based on Variant Call Format Files." *Bioinformatics* 35, no. 10: 1786–1788. <https://doi.org/10.1093/bioinformatics/bty875>.
- Zhang, D., L. Tang, Y. Cheng, et al. 2019. "'Ghost Introgression' as a Cause of Deep Mitochondrial Divergence in a Bird Species Complex." *Molecular Biology and Evolution* 36, no. 11: 2375–2386. <https://doi.org/10.1093/molbev/msz170>.
- Zinenko, O., M. Sovic, U. Joger, and H. L. Gibbs. 2016. "Hybrid Origin of European Vipers (*Vipera magnifica* and *Vipera orlovi*) From the Caucasus Determined Using Genomic Scale DNA Markers." *BMC Evolutionary Biology* 16: 1–13. <https://doi.org/10.1186/s12862-016-0647-7>.
- Zwahlen, V., O. Lourda, S. Ursenbacher, and G. Guillier. 2022. "Rare Genetic Admixture and Unidirectional Gene Flow Between *Vipera aspis* and *Vipera berus* at Their Contact Zone in Western France." *Amphibia-Reptilia* 160, no. 1987: 1–14. <https://doi.org/10.1163/15685381-bja10091>.
- Zwahlen, V., S. Nanni-Geser, L. Kaiser, J. Golay, S. Dubey, and S. Ursenbacher. 2021. "Only Males Care About Their Environment: Sex-Biased Dispersal in the Asp Viper (*Vipera aspis*)." *Biological Journal of the Linnean Society* 132, no. 1: 104–115. <https://doi.org/10.1093/biolinnean/blaa177>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supplementary figures and tables

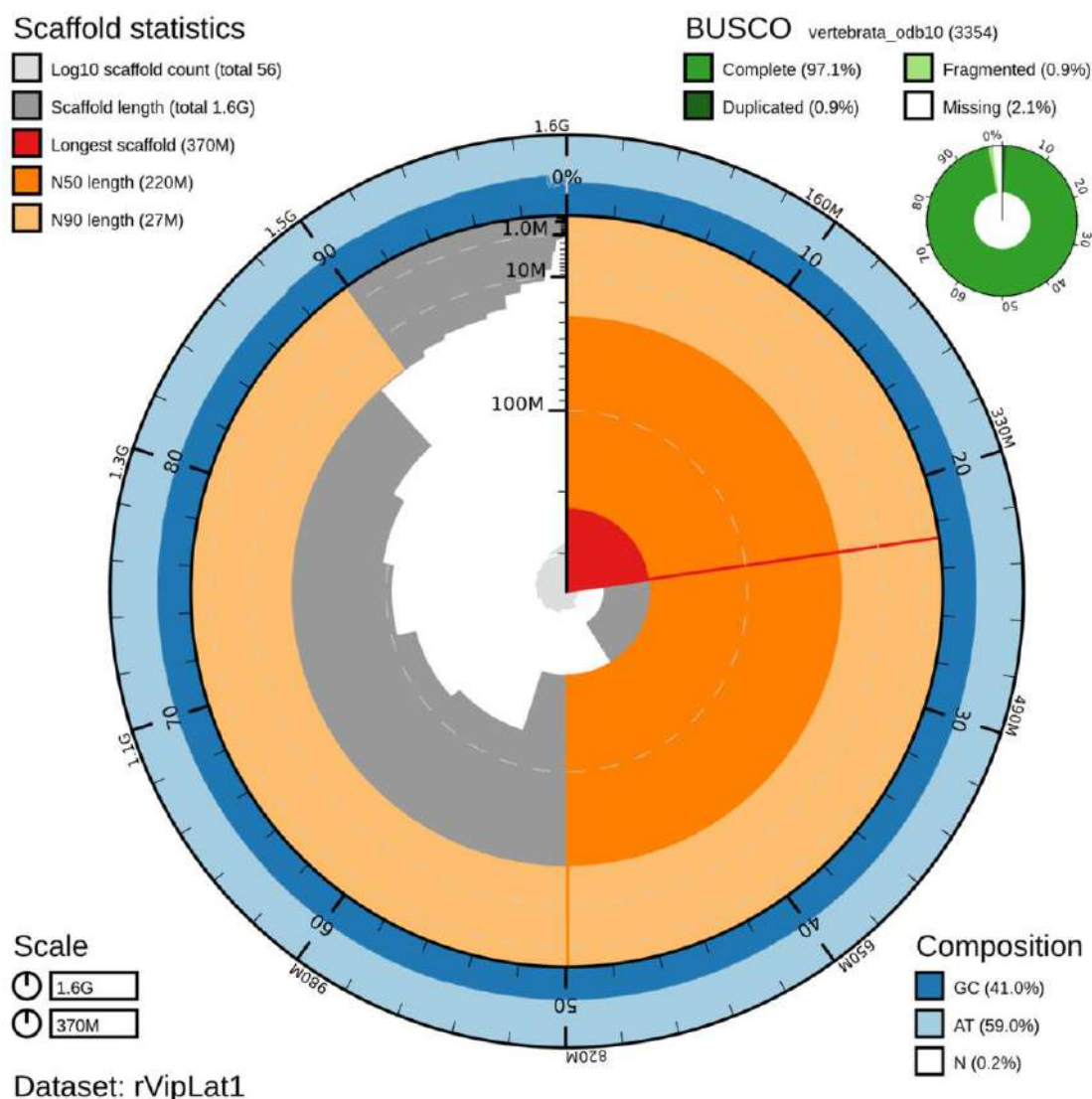


Fig. S1. Snail plot of *Vipera latastei* reference genome assembly presented in this study (rVipLat1), summarizing some contiguity and completeness metrics. M=Mega bases; G=Giga bases. The scaffolds contained in the assembly are shown ordered by length in the inner circle of the snail plot. The sequences within the N50 (220 Mb) and N90 (27 Mb) length marks are painted in dark or light orange, respectively. In the outer circle, the GC content of the sequences in represented in blue, whereas AT content in pale blue. Top left corner shows contiguity statistics like the longest scaffold (1.6 Gb) or Scaffold length (total 1.6 Gb) which indicates the assembly size. Top right corner represents the assembly genetic completeness using BUSCO scores with odb10 vertebrata database with 3354 genes, with 97.1% of the database genes found as Single-Copy Complete genes. Bottom left has a legend with the length in bases represented in the snail plot, the perimeter (1.6 Gb) and the radius (370 Mb, i.e., chromosome 1 length).

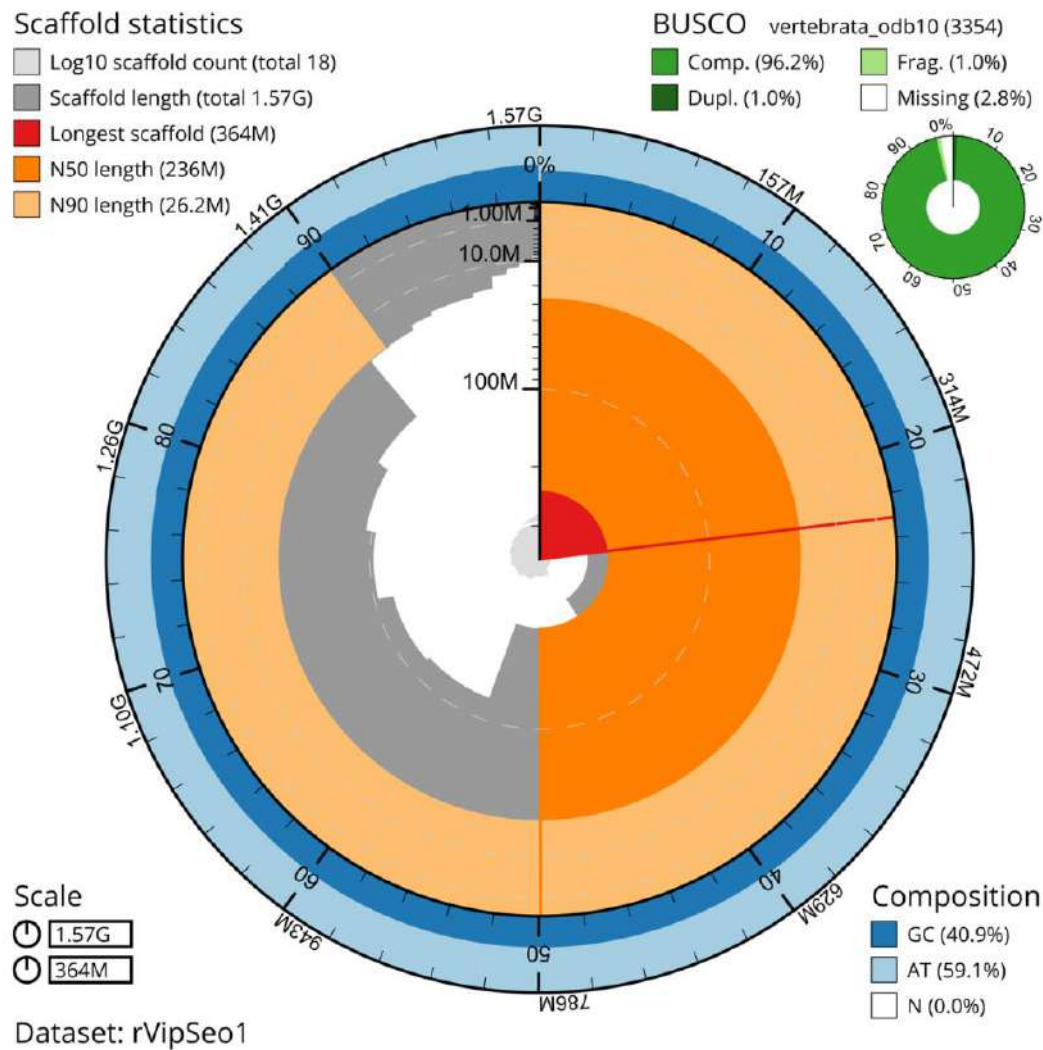


Fig. S2. Snail plot of *Vipera seoanei* reference genome assembly presented in this study (rVipSeo1), summarizing some contiguity and completeness metrics. M=Mega bases; G=Giga bases. The scaffolds contained in the assembly are shown ordered by length in the inner circle of the snail plot. The sequences within the N50 (236 Mb) and N90 (26.2 Mb) length marks are painted in dark or light orange, respectively. In the outer circle, the GC content of the sequences in represented in blue, whereas AT content in pale blue. Top left corner shows contiguity statistics like the longest scaffold (364 Mb) or Scaffold length (total 1.57 Gb) which indicates the assembly size. Top right corner represents the assembly genetic completeness using BUSCO scores with odb10 vertebrata database with 3354 genes, with 96.2% of the database genes found as Single-Copy Complete genes. Bottom left has a legend with the length in bases represented in the snail plot, the perimeter (1.57 Gb) and the radius (364 Mb, i.e., chromosome 1 length).

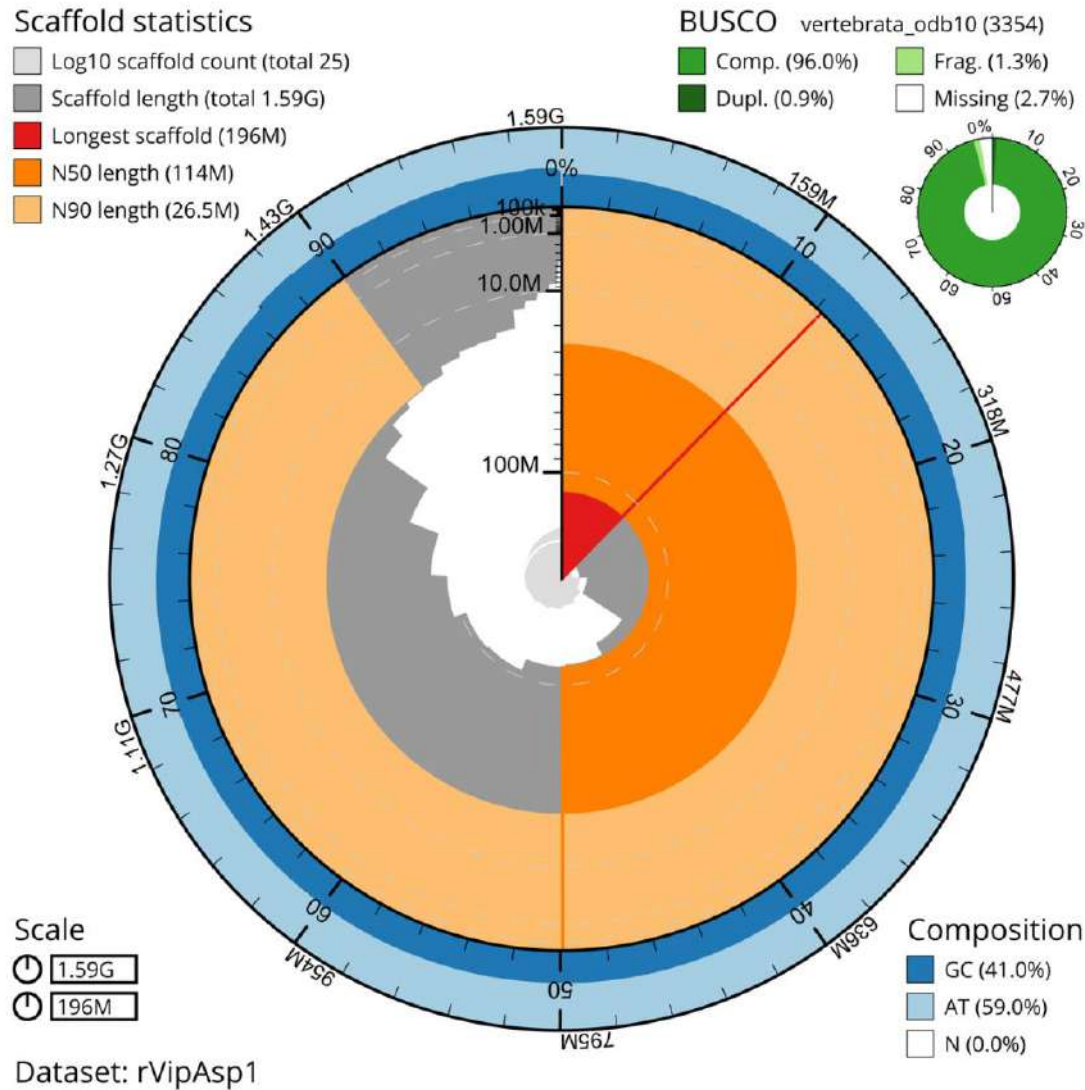


Fig. S3. Snail plot of *Vipera aspis* reference genome assembly presented in this study (rVipAsp1), summarizing some contiguity and completeness metrics. M=Mega bases; G=Giga bases. The scaffolds contained in the assembly are shown ordered by length in the inner circle of the snail plot. The sequences within the N50 (114 Mb) and N90 (26.5 Mb) length marks are painted in dark or light orange, respectively. In the outer circle, the GC content of the sequences in represented in blue, whereas AT content in pale blue. Top left corner shows contiguity statistics like the longest scaffold (196 Mb) or Scaffold length (total 1.59 Gb) which indicates the assembly size. Top right corner represents the assembly genetic completeness using BUSCO scores with odb10 vertebrata database with 3354 genes, with 96.0% of the database genes found as Single-Copy Complete genes. Bottom left has a legend with the length in bases represented in the snail plot, the perimeter (1.59 Gb) and the radius (196 Mb, i.e., chromosome 1 length).

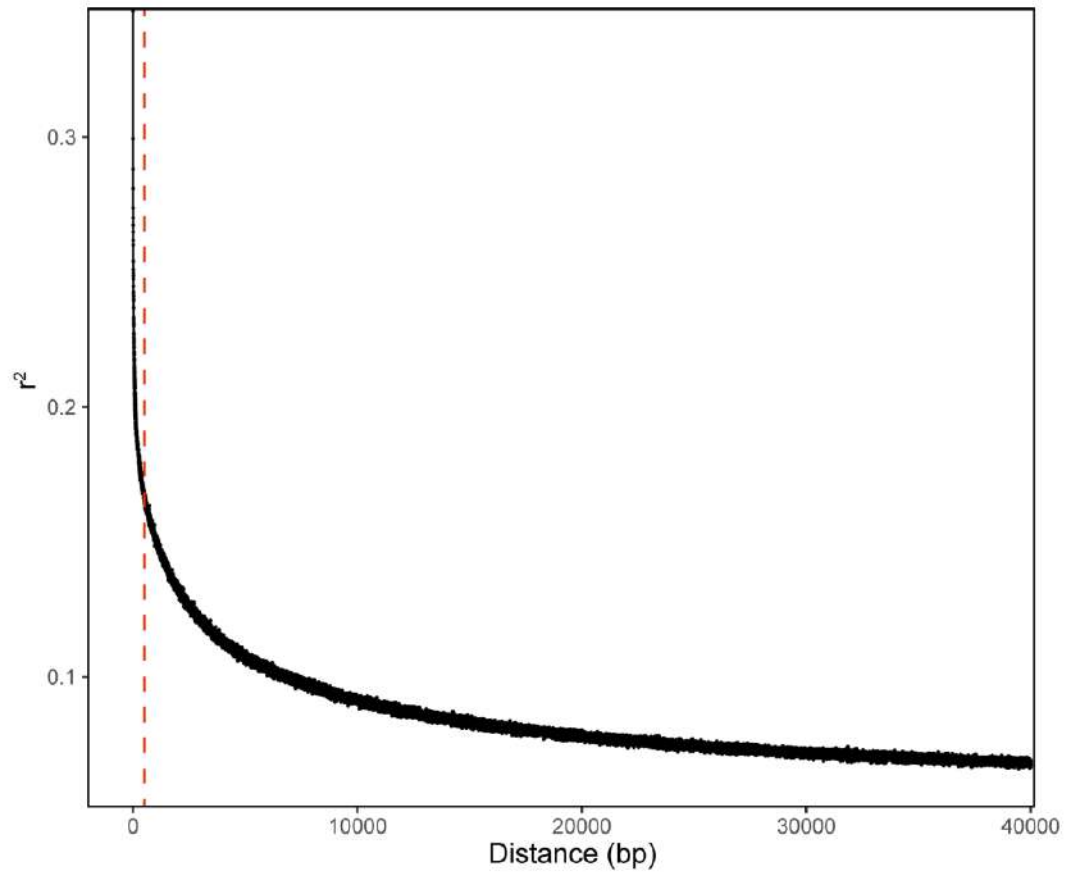


Fig. S4. Linkage disequilibrium (LD) decay in *Vipera seoanei* across sites. The vertical line marks LD at sites 0.5 kbp apart. Correlation between sites at this distance is lower than 20%.

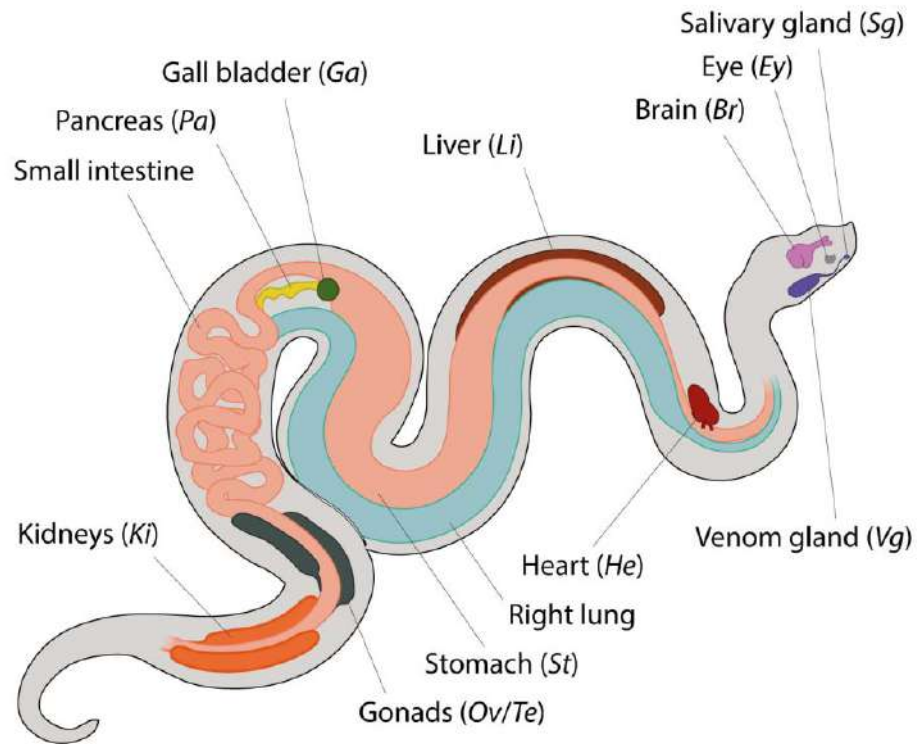


Fig. S5. Anatomical scheme of a viper. Tissues and organs with an abbreviation in between bracket have been used for the differential gene expression analysis shown in Fig 6A.

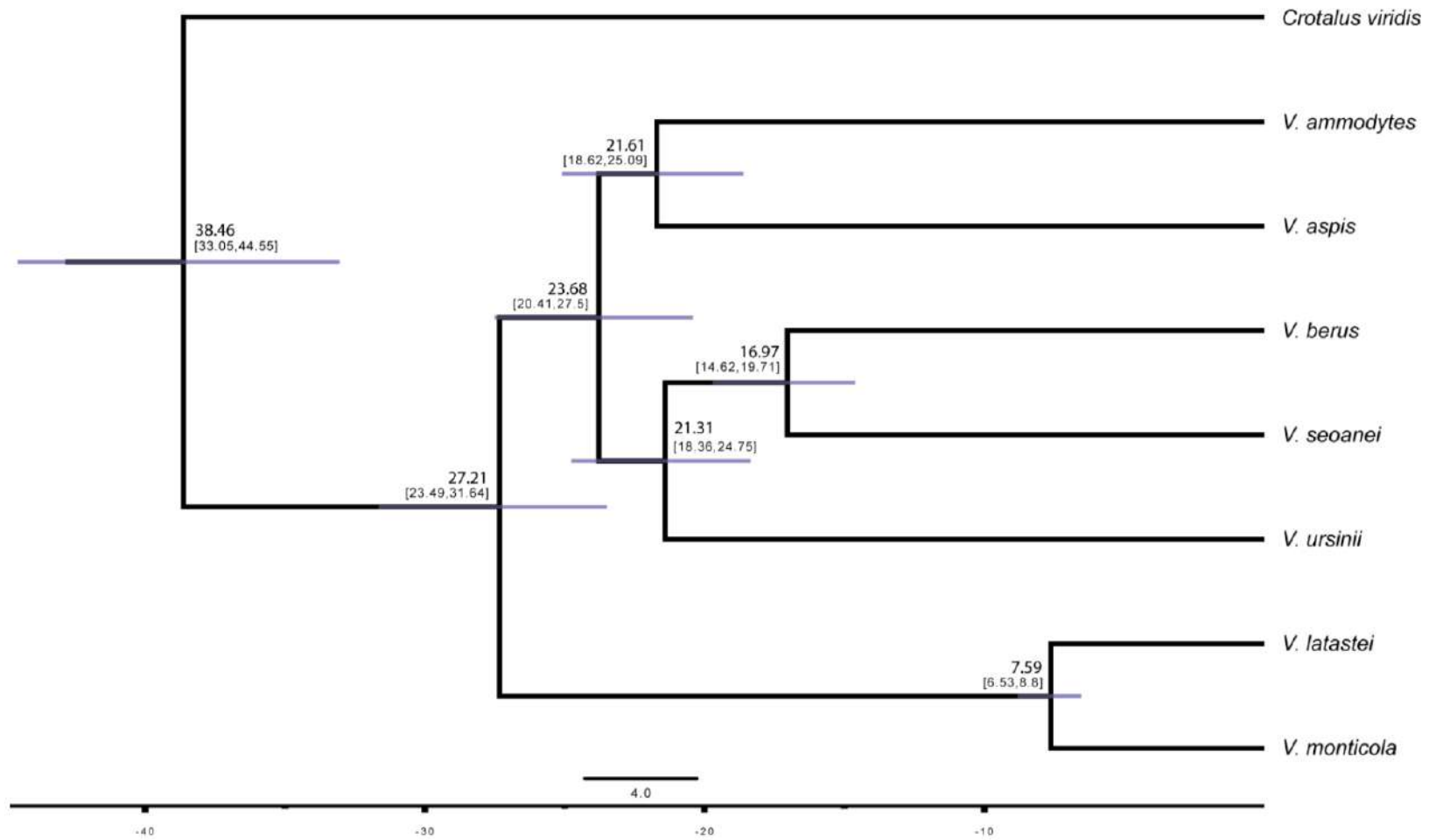


Fig. S6. Time-calibrated genomic species tree inferred with SNAPP, using 1.23M uSNPs and eight species. Each node is annotated with its median age and 95%CI between brackets, measured in million years ago (Mya). The posterior probabilities of all nodes are 1.

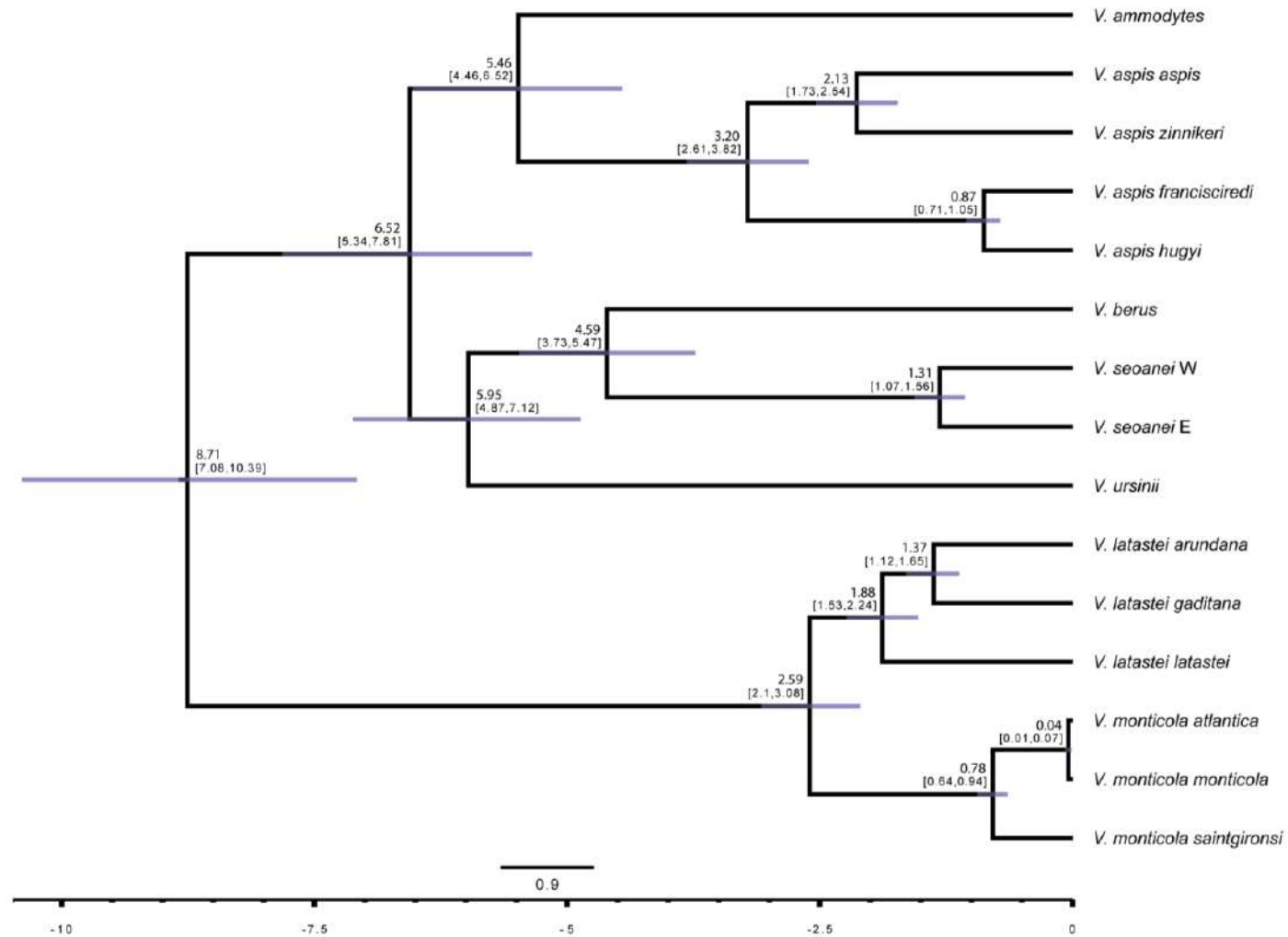


Fig. S7. Time-calibrated genomic species tree inferred with SNAPP, using 126k uSNPs and 15 taxa. Each node is annotated with its median age and 95%CI between brackets, measured in million years ago (Mya). The posterior probabilities of all nodes are 1.

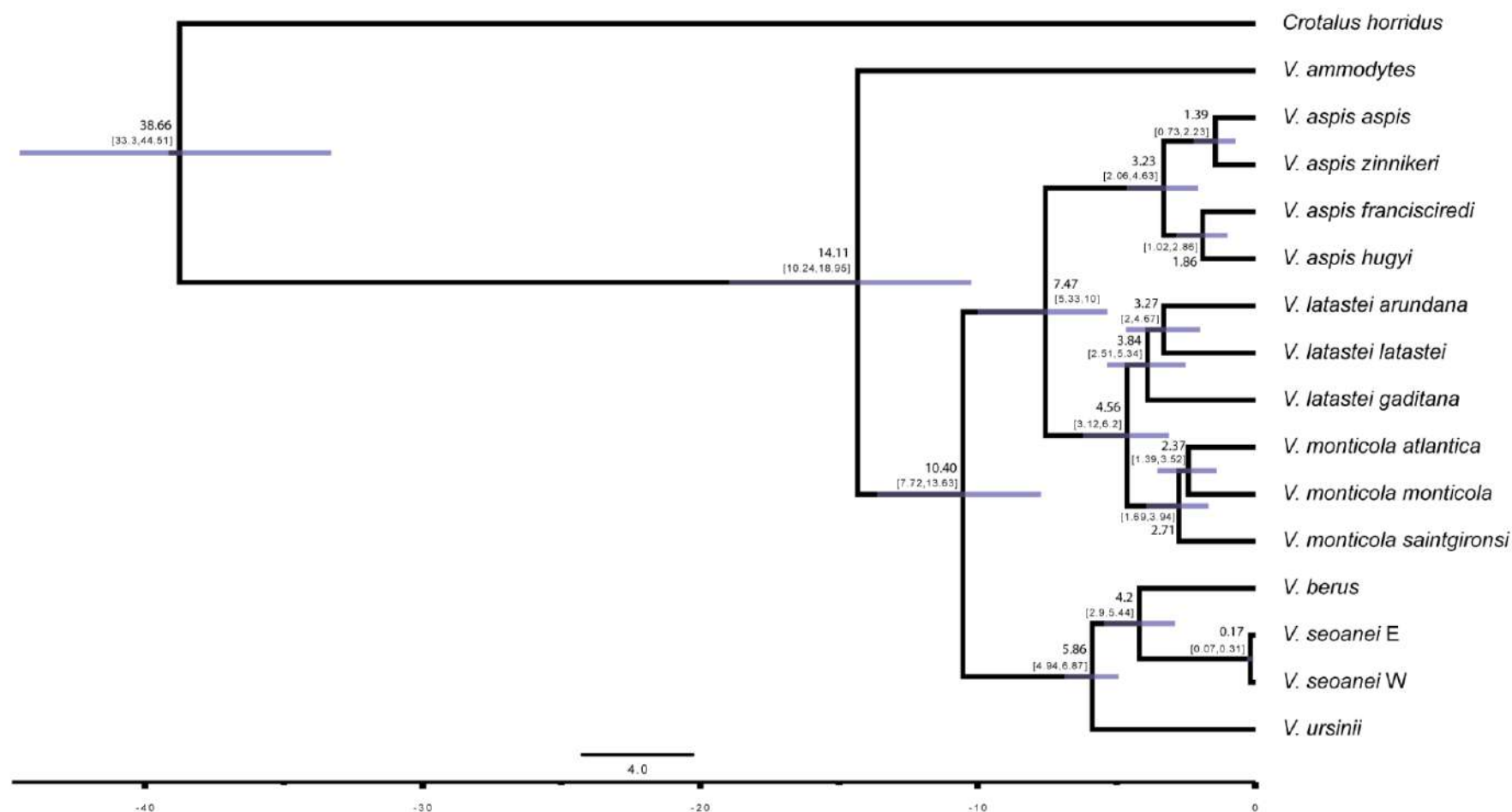


Fig. S8. Time-calibrated mitochondrial tree inferred with BEAST2, using a 10.7-kbp alignment, 18 partitions and 13 protein coding genes from 15 taxa. Each node is annotated with its median age and 95%CI between brackets. The posterior probabilities of all nodes are 1.

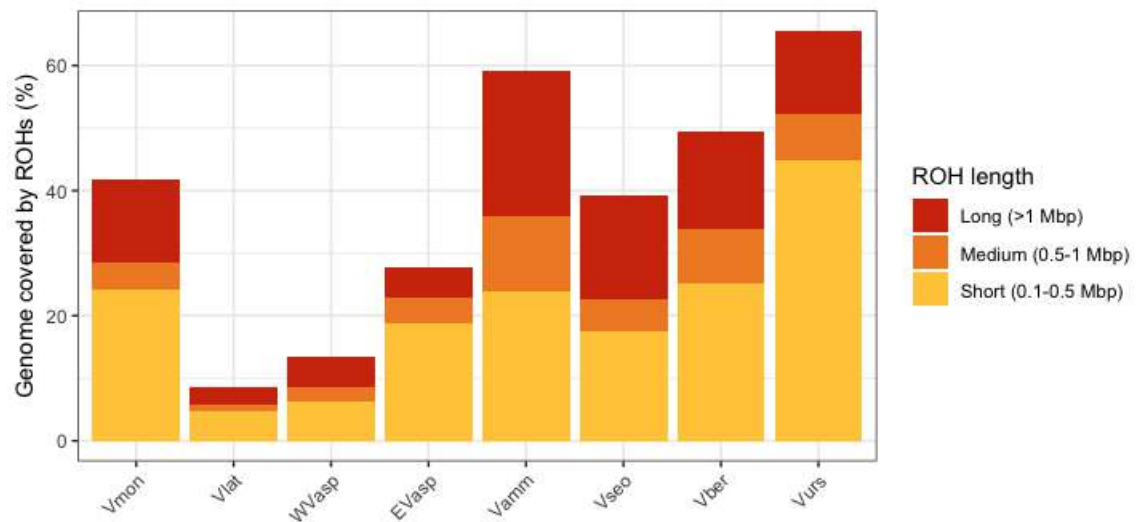


Fig. S9. Percentage of autosomes covered by Runs of Homozygosity (ROHs). Vmon stands for *V. monticola*, Vlat for *V. latastei*, WVasp for Western *V. aspis*, EVasp for Eastern *V. aspis*, Vamm for *V. ammodytes*, Vseo for *V. seoanei*, Vber for *V. berus* and Vurs for *V. ursinii*. Western *V. aspis* and *V. latastei* show the lowest percentage of their genomes covered by short ROHs, whereas *V. ursinii*, the highest.

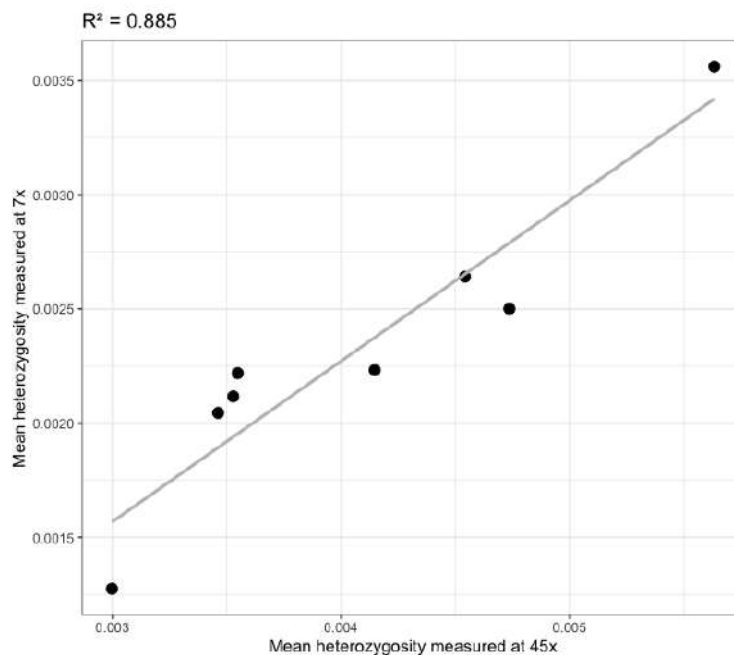


Fig. S10. Correlation between genome-wide heterozygosity estimates from low-coverage samples (7x) versus their high-coverage counterparts (45x). Only 8 samples were sequenced at high coverage, but the high correlation ($R^2=0.885$) indicates that, despite retrieving lower absolute estimates at 7x, measurements at different coverages are comparable.

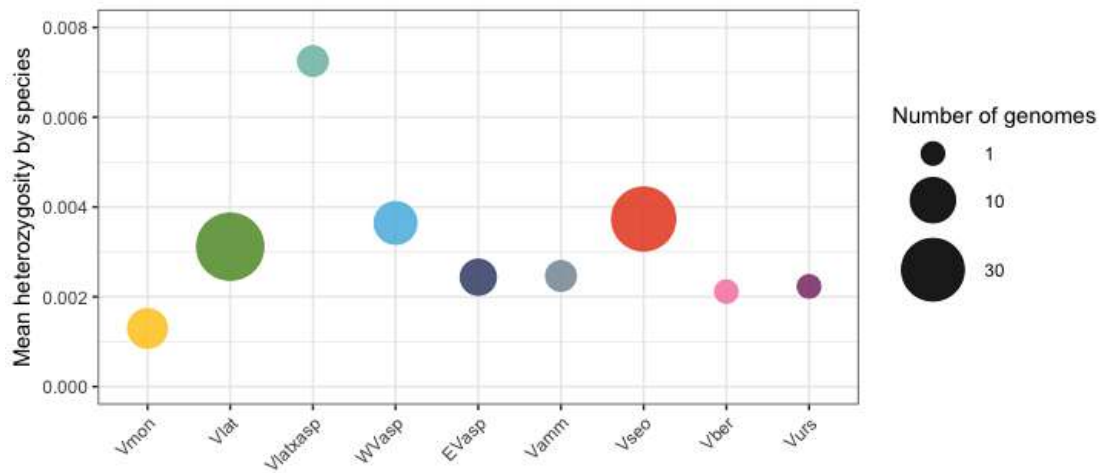


Fig. S11. Mean genome-wide heterozygosity values per species. Bubble sizes depict the number of samples for each group. Vmon stands for *V. monticola* (n=8), Vlat for *V. latastei* (n=37), Vlatxasp for *V. latastei* x *aspis* F1 hybrids (n=2), WVasp for Western *V. aspis* (n=8), EVasp for Eastern *V. aspis* (n=4), Vamm for *V. ammodytes* (n=2), Vseo for *V. seoanei* (n=32), Vber for *V. berus* (n=1), and Vurs for *V. ursinii* (n=1). *Vipera monticola* shows the lowest mean diversity, whereas the hybrids in between *V. latastei* x *aspis*, the highest. In general, Iberian lineages within each of the three old-diverging species groups show the highest diversity.

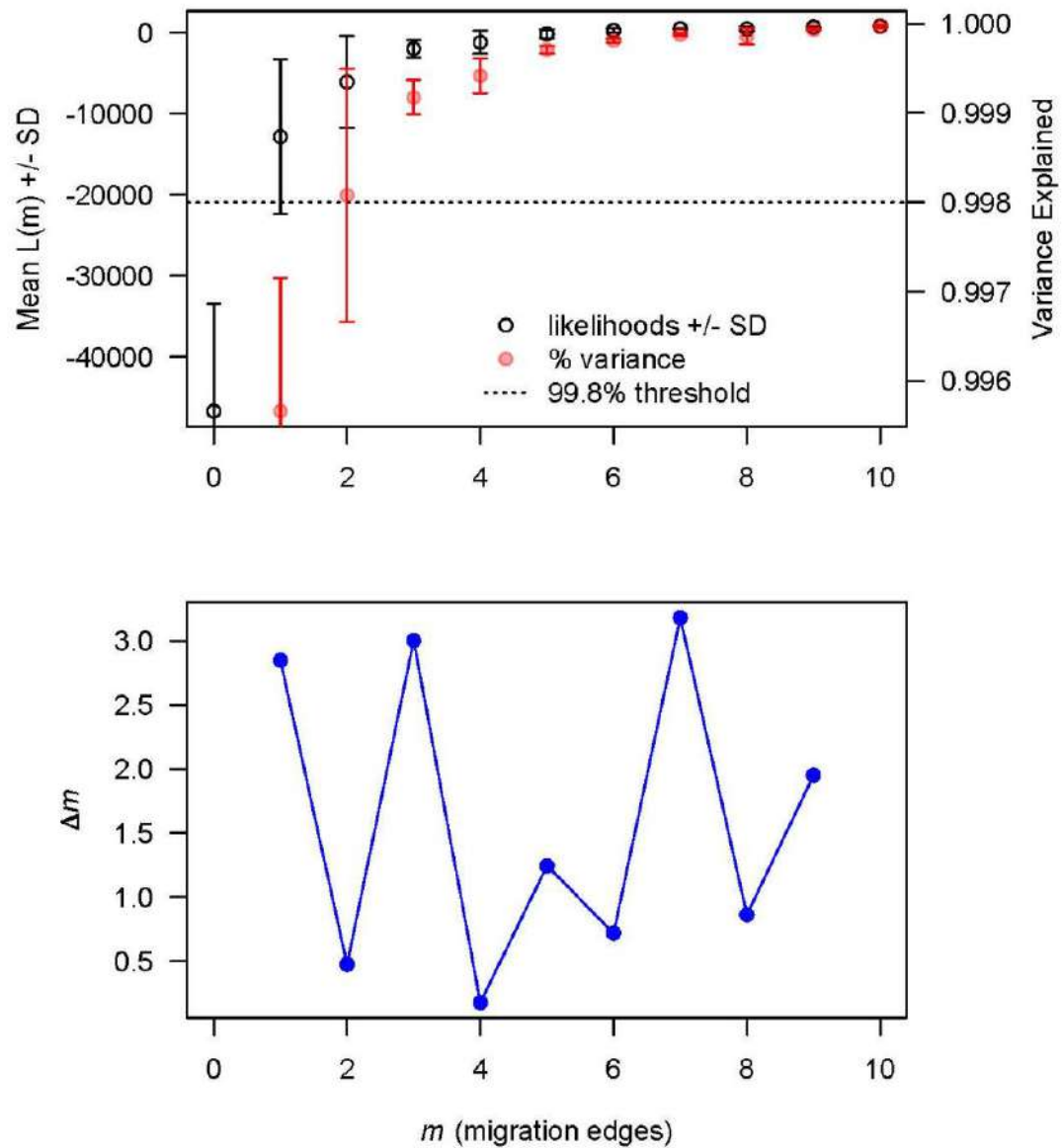


Fig. S12. Output produced by OptM to find the optimum number of migrations in TreeMix. At the top, the mean and standard deviation (SD) for the composite likelihood $L(m)$ (left axis, black circles) and proportion of the variance explained (right axis, red circles). The 99.8% threshold is that recommended by Pickrell and Pritchard (2012). On the bottom, the second-order rate of change (Δm) across values of m . The peak is reached at $m=7$ edges.

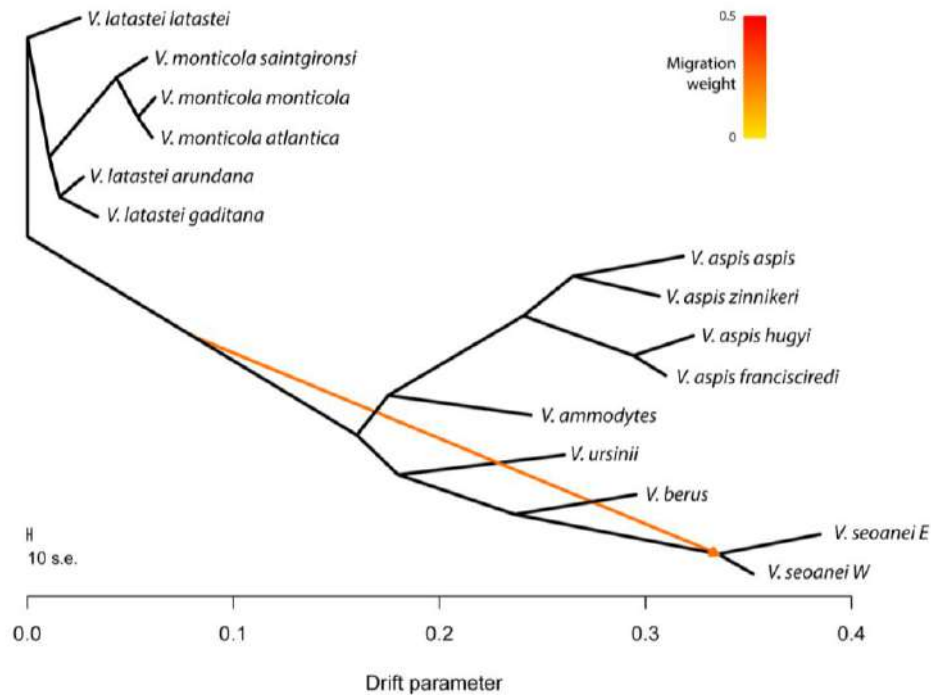


Fig S13. Tree with one migration edge inferred by TreeMix. $m=1$ is the third most likely value inferred by OptM. Color of the edge indicates the weight of migration. The drift parameter is a relative temporal measurement, and the scale bar indicates 10 times the average standard error of the relatedness among taxa based on the variance-covariance matrix of allele frequencies.

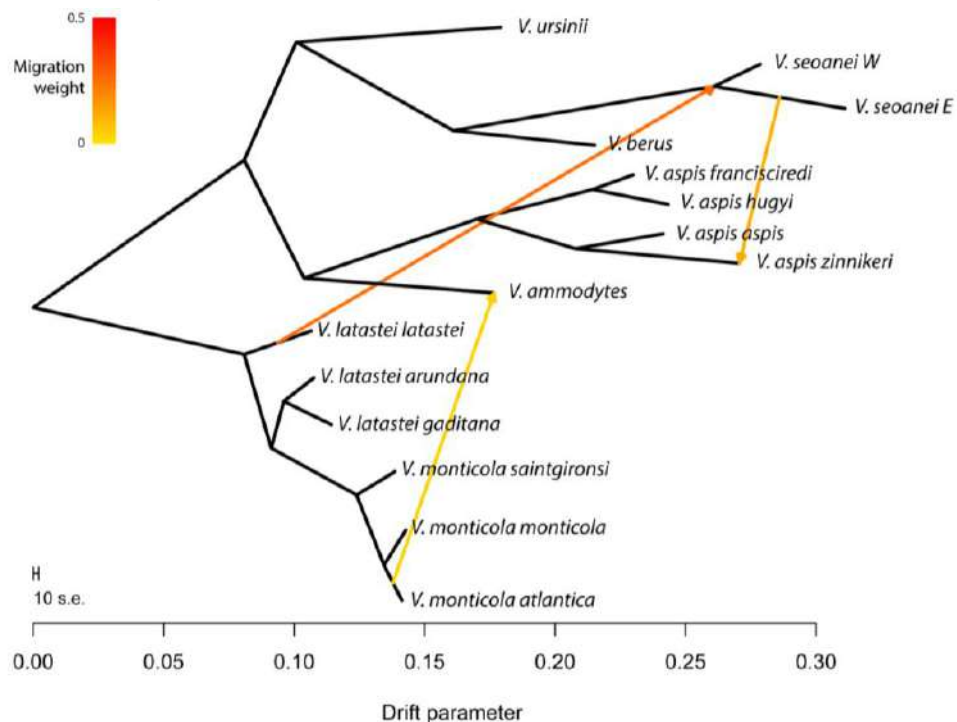


Fig S14. Tree with three migration edges inferred by TreeMix. $m=3$ is the second most likely value inferred by OptM. Color of the edges indicates the weight of migration. The drift parameter is a relative temporal measurement, and the scale bar indicates 10 times the average standard error of the relatedness among taxa based on the variance-covariance matrix of allele frequencies.

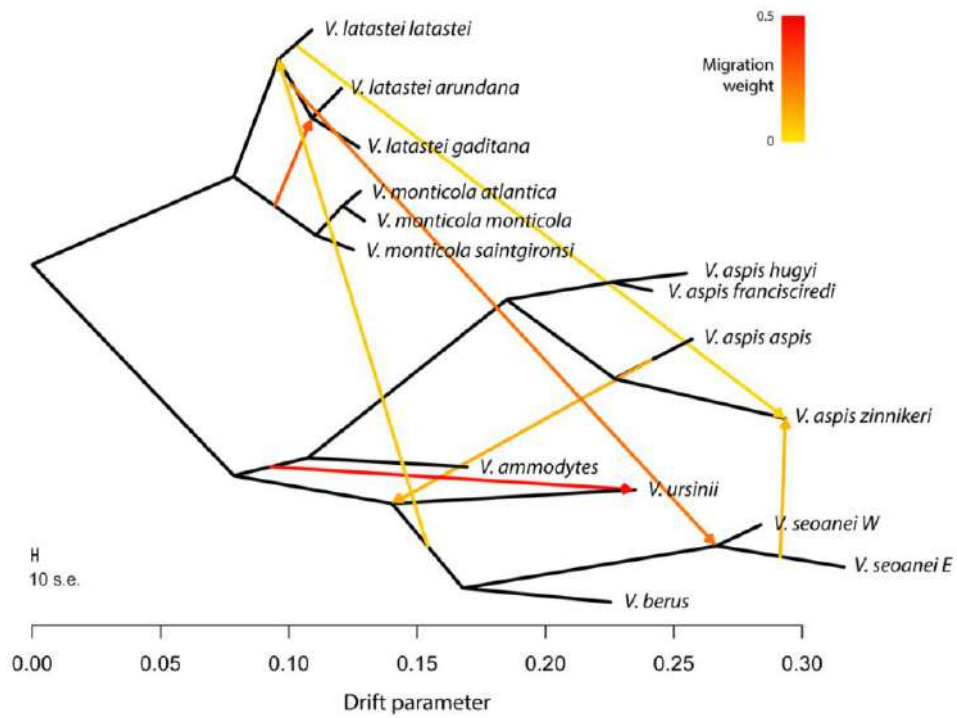


Fig S15. Tree with seven migration edges inferred by TreeMix. $m=7$ is the most likely value inferred by OptM. Color of the edges indicates the weight of migration. The drift parameter is a relative temporal measurement, and the scale bar indicates 10 times the average standard error of the relatedness among taxa based on the variance-covariance matrix of allele frequencies.

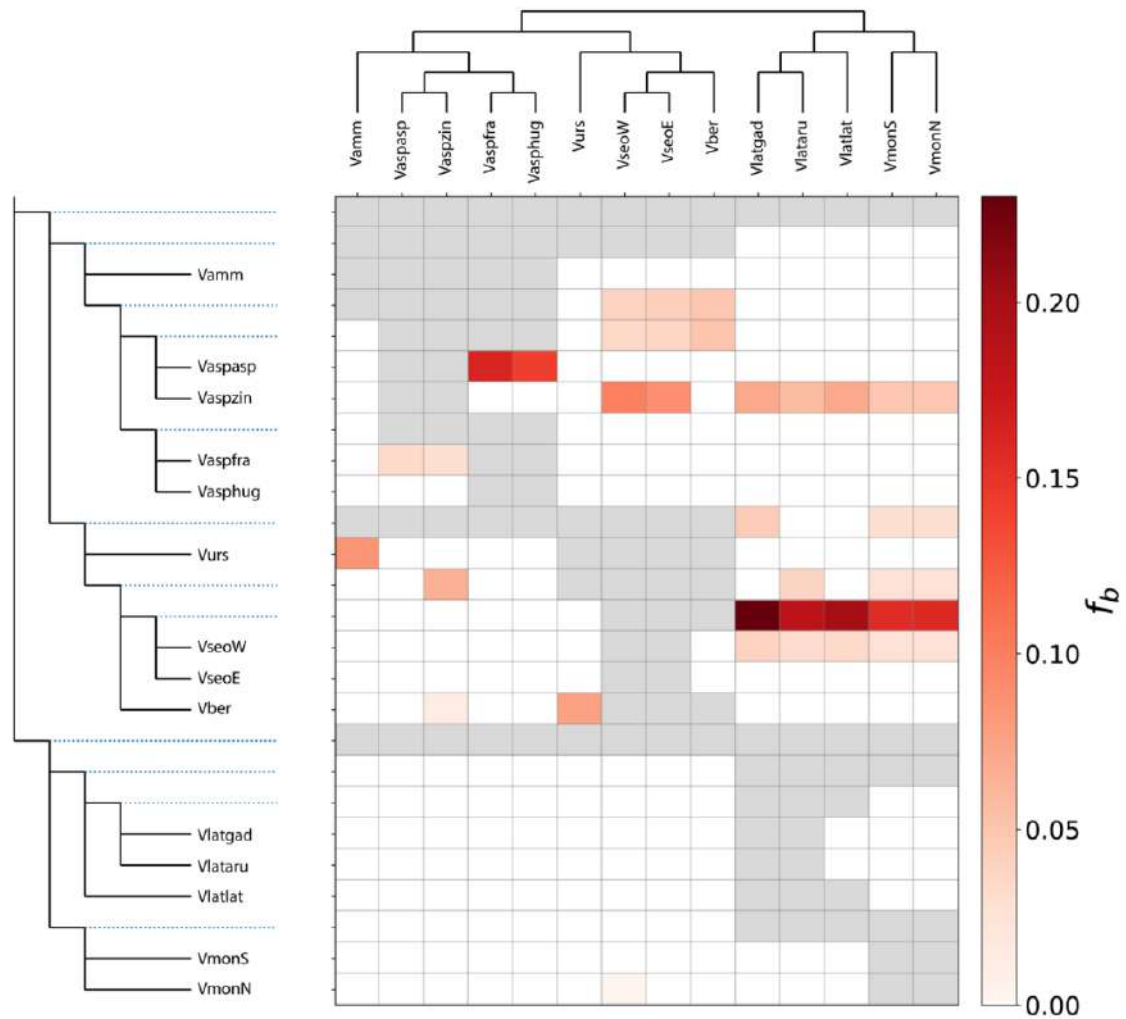


Fig. S16. Heatmap of f -branch (f_b) statistic, representing excess allele sharing between branches of y- and x-axes. Statistics have been corrected by a robust clustering threshold to avoid false positives due to homoplasy instead of introgression (see sensitive clustering threshold results in Fig. 4A). Blank cells depict values not statistically supported, whereas grey cells indicate comparisons that cannot be made due to tree topology. Exact f -branch values are shown in Table S4. Vamm stands *V. ammodytes*, Vaspasp for *V. aspis aspis*, Vaspzin for *V. aspis zinnikeri*, Vaspfra for *V. aspis francisciredi*, Vaspbug for *V. aspis hugyi*, Vurs for *V. ursinii*, VseoW for *V. seoanei* West, VseoE for *V. seoanei* East, Vber for *V. berus*, Vlatgad for *V. latastei gaditana*, Vlataru for *V. latastei arundana*, Vlatlat for *V. latastei latastei*, Vmon S for *V. monticola* South and VmonN for *V. monticola* North.

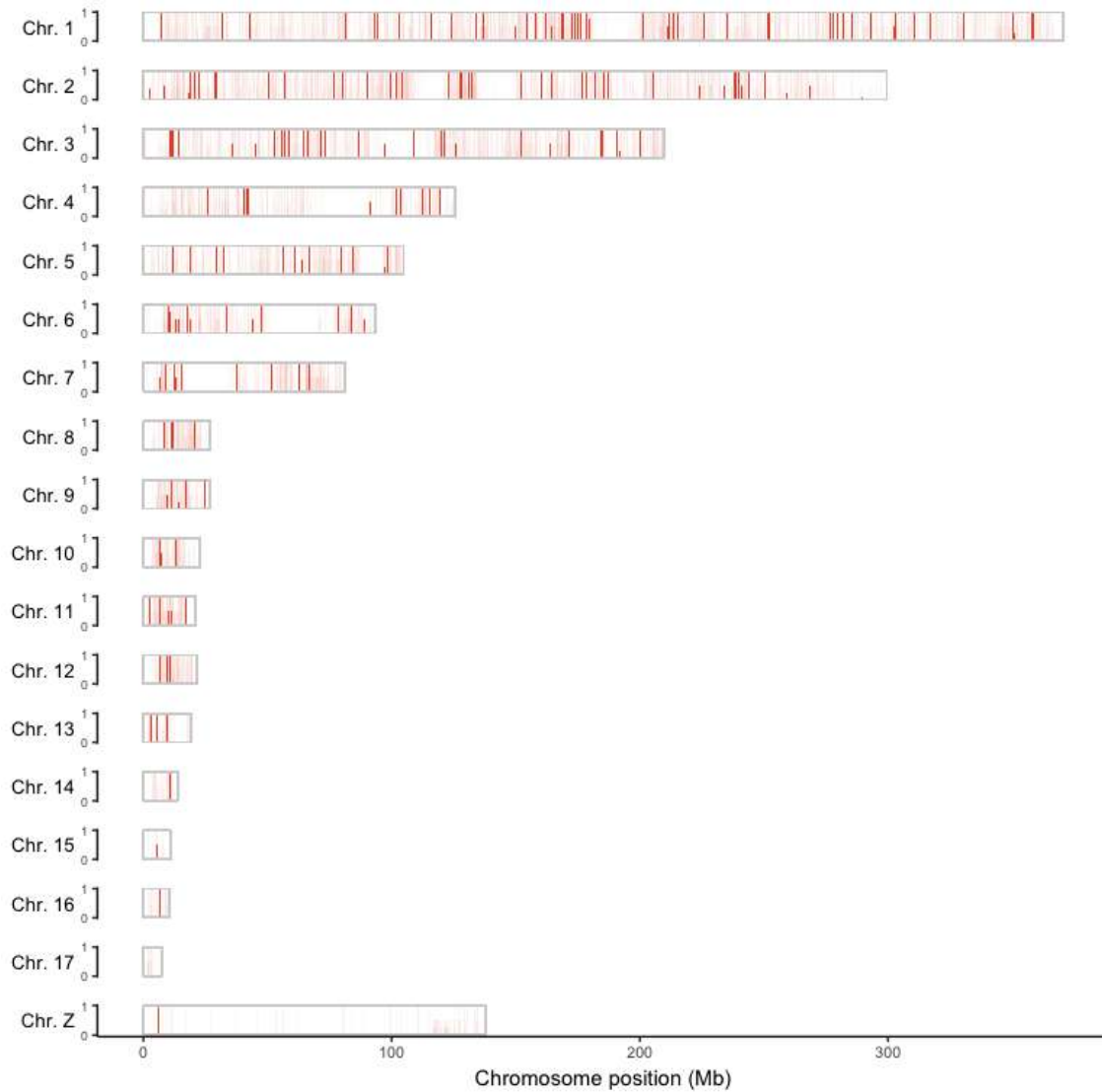


Fig. S17. Landscape of introgression from *V. latastei* to *V. seoanei*, shown in *V. latastei*'s reference genome highlighting windows with excess allele-sharing.

Introgression between these species seems to be pervasive, although some areas display higher intensity whereas others, including the whole Z chromosome, are less introgressed (i.e., “introgression deserts”).

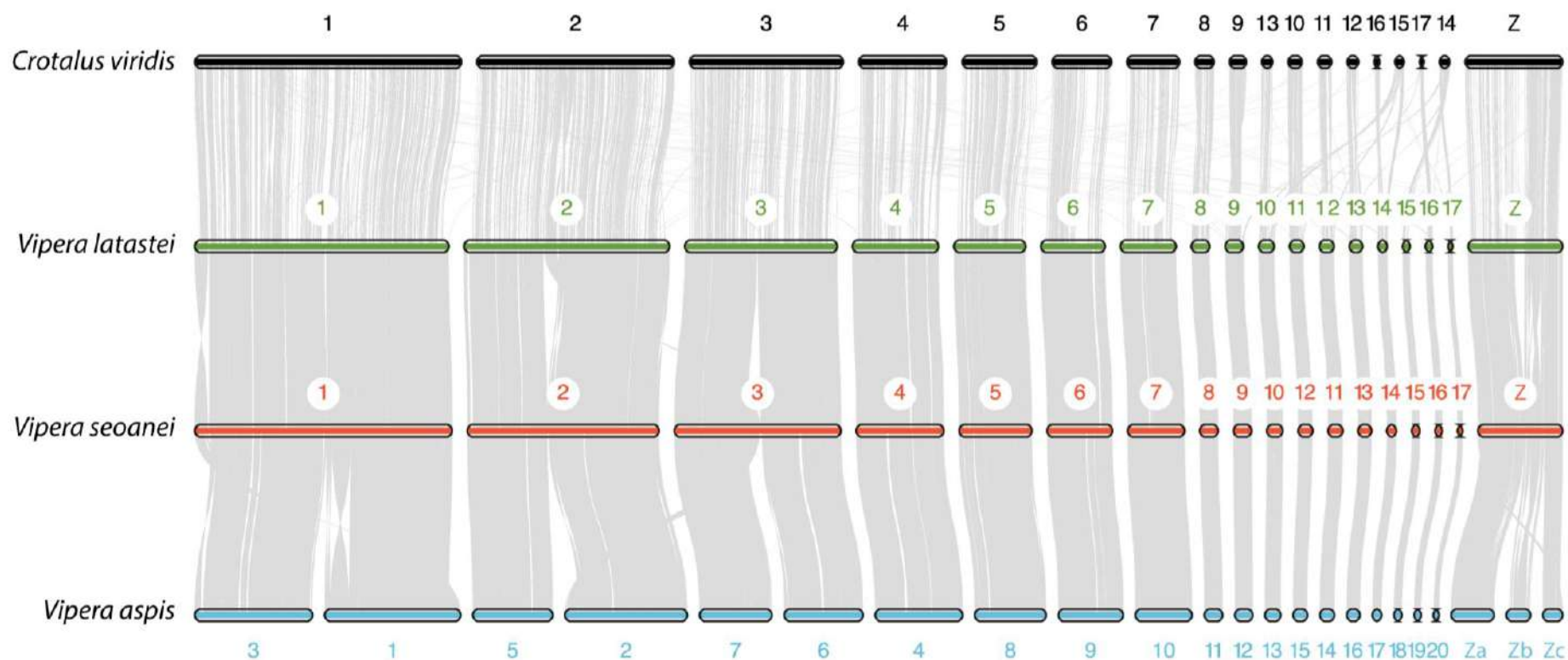


Fig. S18. Macrosynteny across the three old species groups in *Vipera*, represented by the three Iberian species, and a rattlesnake as outgroup.

Some microchromosomal rearrangements have occurred in between the sister subfamilies Crotalinae and Viperinae, while macrochromosomes have been mostly conserved. The ancestral chromosome number, i.e., $2n=36$ is conserved in *V. latastei* and *V. seoanei* but increased in *V. aspis* as a result of three fission events in the three major macrochromosomes. Z chromosome in *V. aspis* appears fragmented due to lower Omni-C coverage during the assembly, since it was sequenced from a female (ZW) and the other three assembled genomes correspond to males (ZZ). A fragment of *V. latastei* chromosome 2 is translocated to *V. seoanei* chromosome 3. Some chromosomes have been reoriented on the grounds of a better visualization.

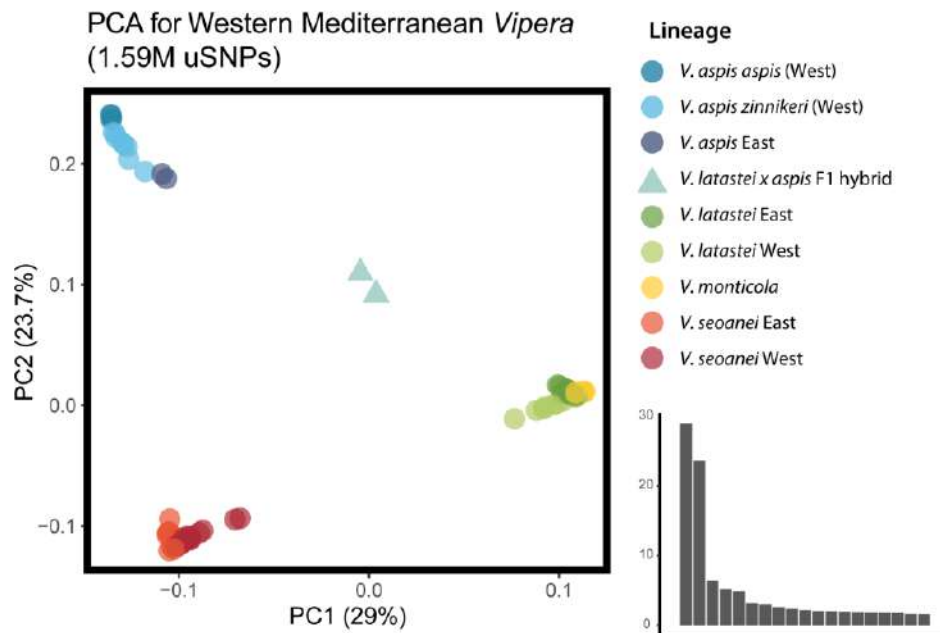


Fig. S19. Genomic PCA of Western Mediterranean *Vipera* spp. performed with 1.59M uSNPs and 90 whole genomes. PC1 (29% of explained variance) splits *Vipera latastei-monticola* complex from *V. seoanei* and *V. aspis*, whereas PC2 (23.7%) splits the latter two. Two individuals seem to be F1 hybrids of *V. latastei x aspis*, found in sympatry areas of N and NE Spain. The Western lineages of *V. seoanei* and *V. latastei* seem to be closer, probably due to higher admixture between them. More discretely, the Eastern lineages of *V. seoanei* and *V. latastei* seem to be relatively closer, but more slightly, to *V. aspis*. The bar plot on the right shows the explained variance by component.

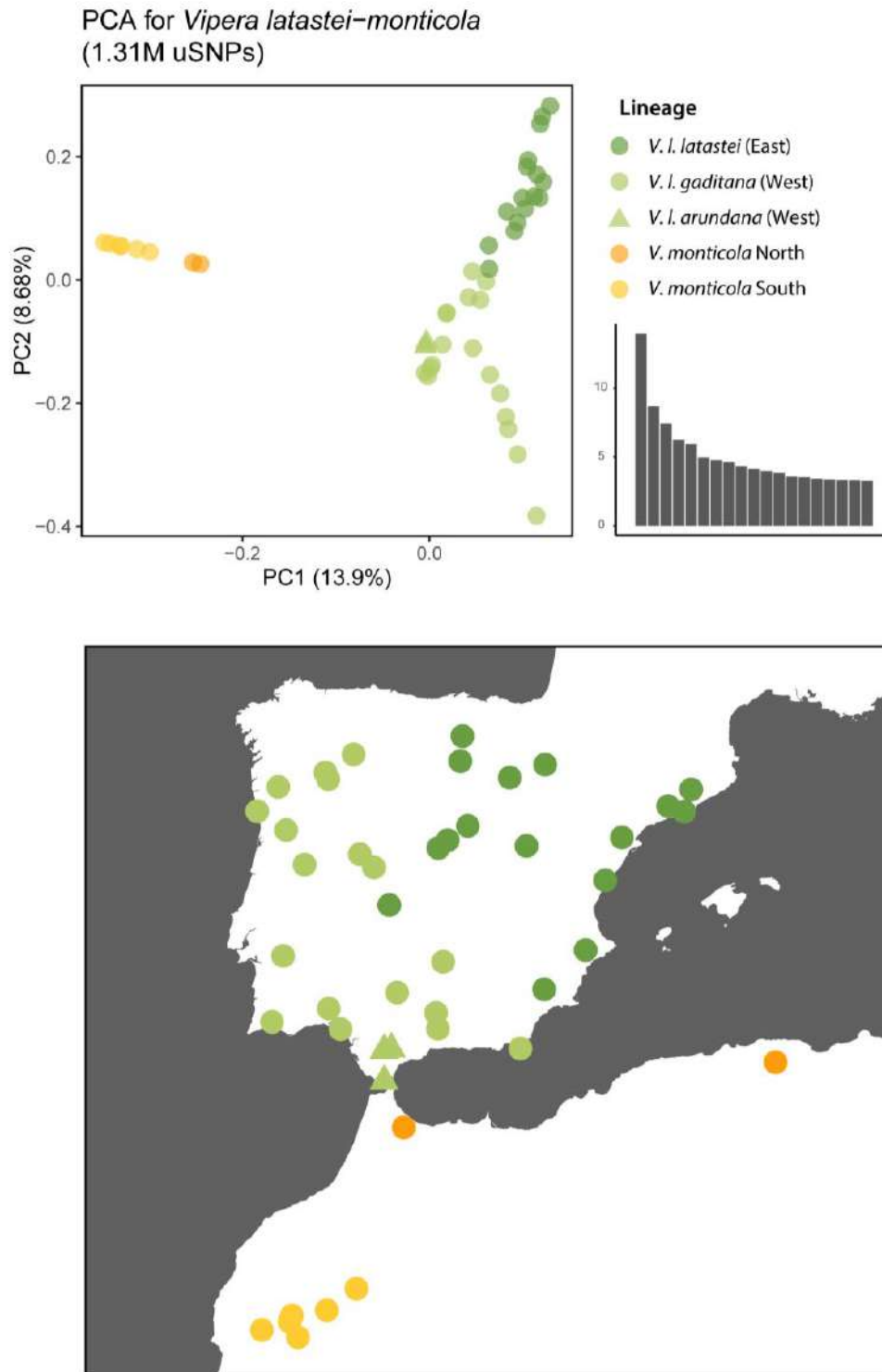


Fig. S20. Genomic PCA of the *latastei* group (top), performed with 1.31M uSNPs and 45 whole genomes and geographic origin of the samples (bottom). PC1 (13.9% of explained variance) splits *Vipera latastei* from *V. monticola*, as well as depicts a South-North gradient within each. PC2 (8.68%) explains an East-West gradient of *V. latastei* across the Iberian Peninsula. The individuals belonging to *V. latastei arundana* (triangles) fall within other Western *V. latastei*, which mostly corresponds to *V. latastei gaditana*. *Vipera latastei* East mostly corresponds to the nominotypical *V. latastei latastei*. The bar plot on the right shows the explained variance by component.

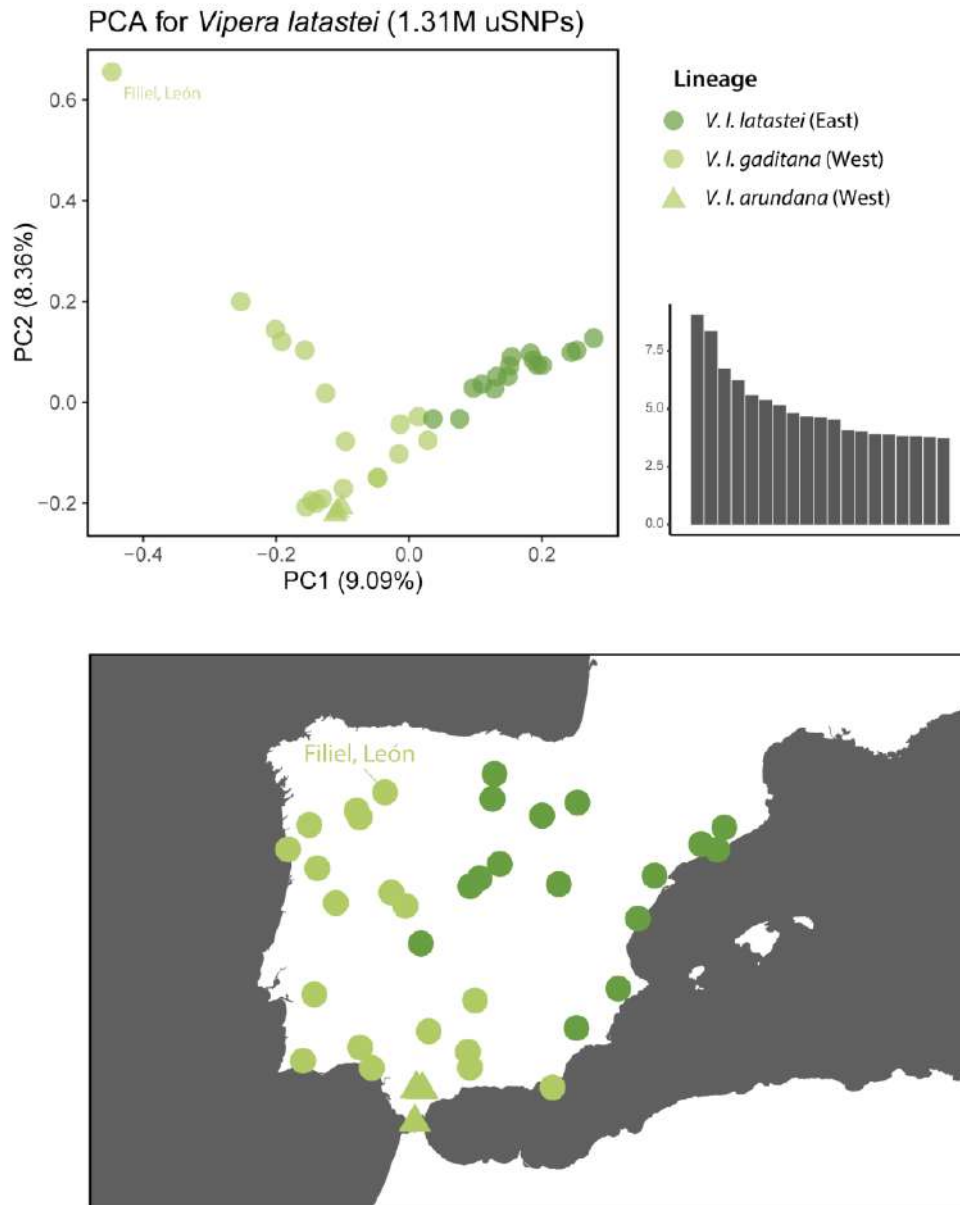


Fig. S21. Genomic PCA of *Vipera latastei* (top), performed with 1.31M uSNPs and 37 whole genomes and geographic origin of the samples (bottom). PC1 (9.09% of explained variance) depicts a West-East gradient, whereas PC2 (8.36%) explains North-South variation, including admixture with *V. seoanei* in the northwest, more pronounced in an individual from Filiel, León. The individuals belonging to *V. latastei arundana* (triangles) fall within other Western *V. latastei*, which mostly corresponds to *V. latastei gaditana*. *Vipera latastei* East mostly corresponds to *V. latastei latastei*. The bar plot on the right shows the explained variance by component.

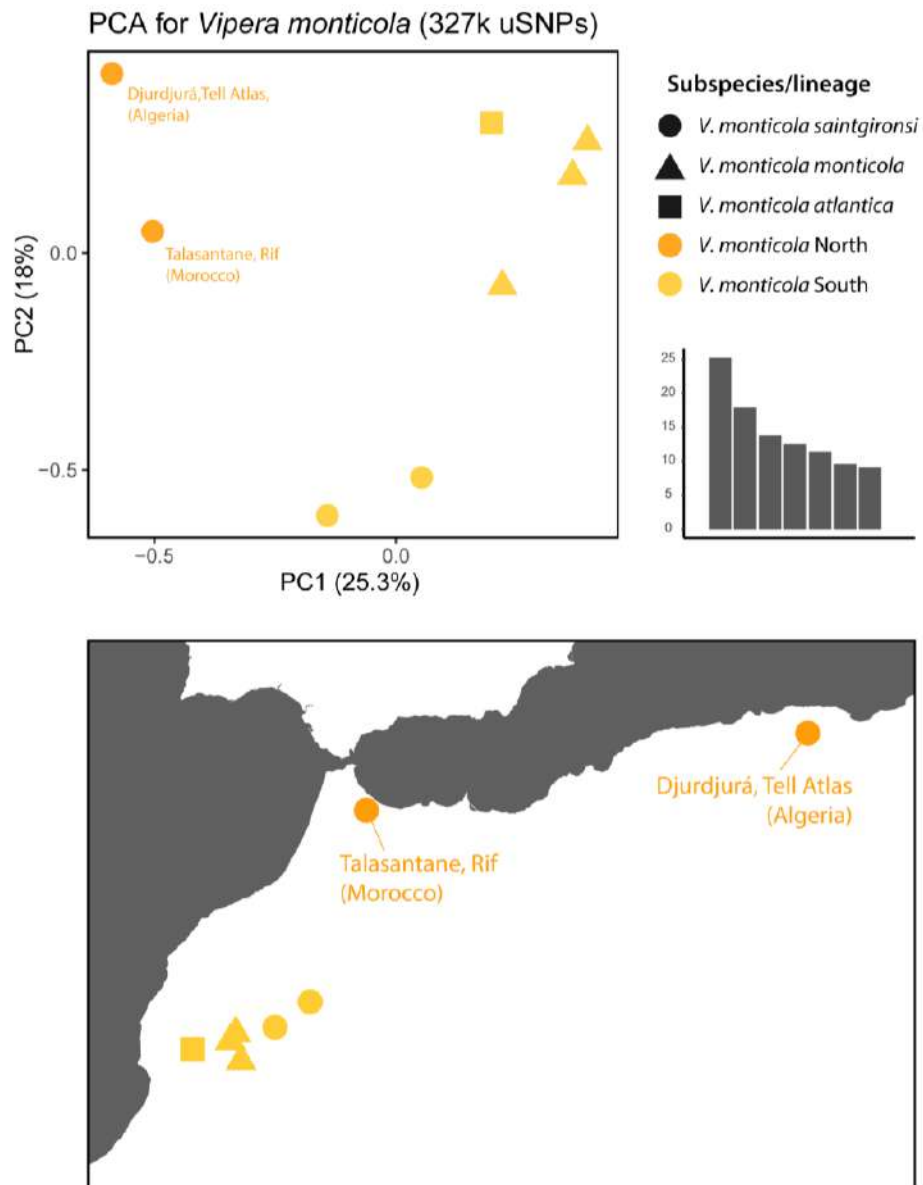


Fig. S22. Genomic PCA of *Vipera monticola* (top), performed with 327k uSNPs and eight whole genomes and geographic origin of the samples (bottom). PC1 (25.3% of explained variance) depicts a North-South gradient, from the northern populations attributed to *V. monticola saintgironsi* to the Atlas populations in the south, attributed to three different subspecies. Southern (Atlas) populations form a subtle cluster along PC1. Thus, we differentiate *V. monticola* into North and South (Atlas) lineages instead of using mitochondrial subspecies. PC2 (18%) explores mostly differences from the Western to the Eastern Atlas within the South lineage. The bar plot on the right shows the explained variance by component.

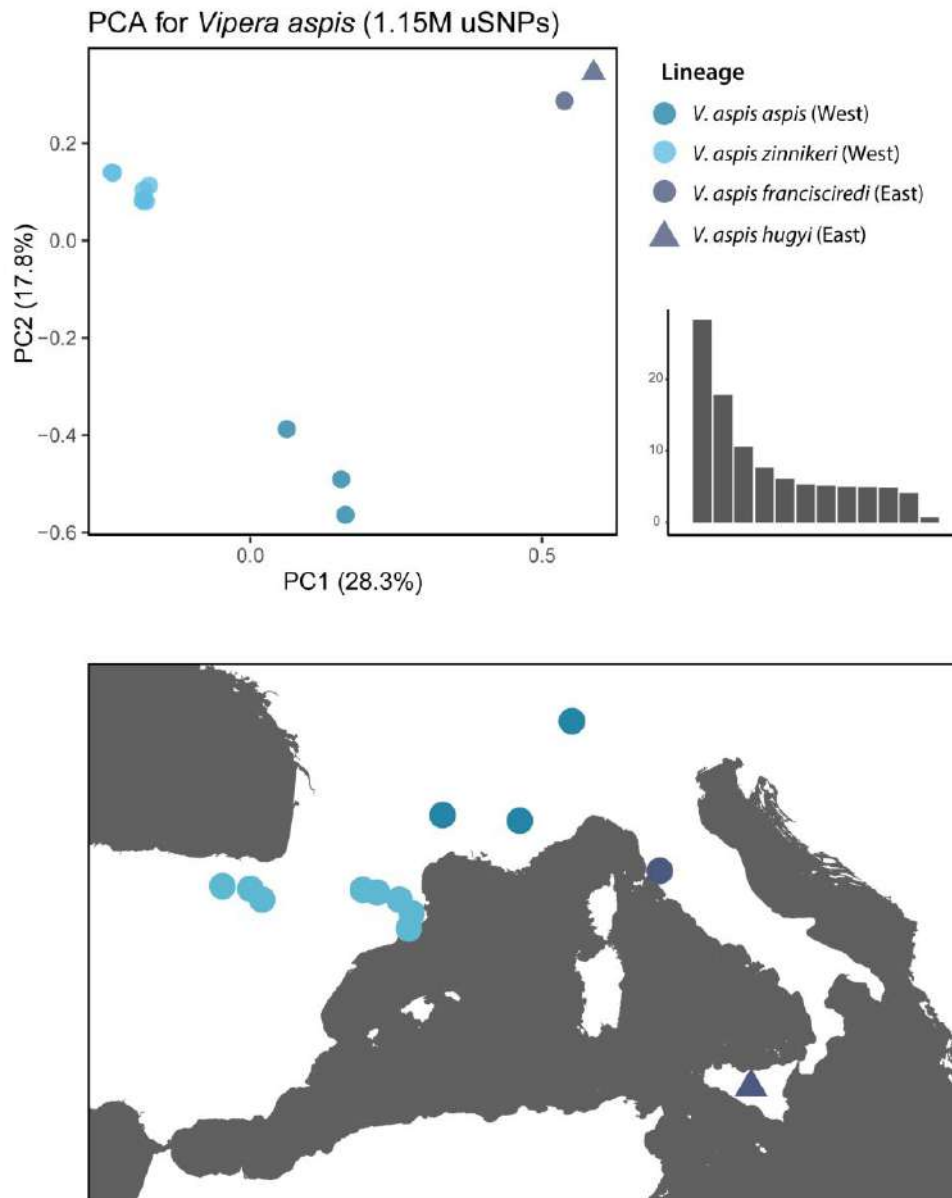


Fig. S23. Genomic PCA of *Vipera aspis* (top), performed with 1.15M uSNPs and 13 whole genomes and geographic origin of the samples (bottom). PC1 (28.3% of explained variance) splits *Vipera aspis* into three clusters: *V. a. zinnikeri* and *V. a. aspis*, closer between them, and an Eastern cluster composed by individuals from two subspecies: *V. a. francisciredi* and *V. a. hugyi*. PC2 (17.8%) separates *V. a. aspis* from the rest. The bar plot on the right shows the explained variance by component.

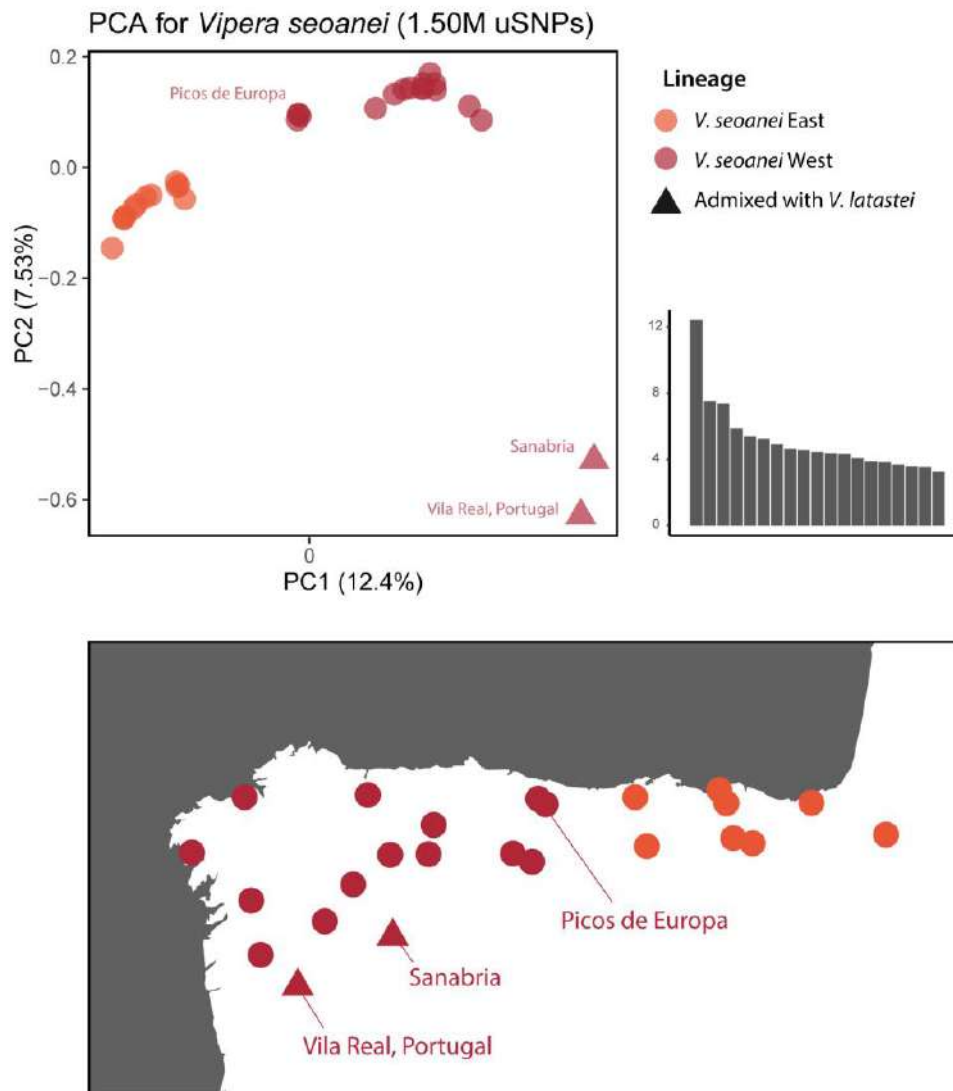


Fig. S24. Genomic PCA of *Vipera seoanei* (top), performed with 1.50M uSNPs and 30 whole genomes and geographic origin of the samples (bottom). PC1 (12.4% of explained variance) splits *Vipera seoanei* into Eastern and Western clades, with Picos de Europa populations exhibiting a more intermediate position. PC2 (7.53%) mainly separates two admixed Western individuals with *V. latastei* (triangles). The bar plot on the right shows the explained variance by component.

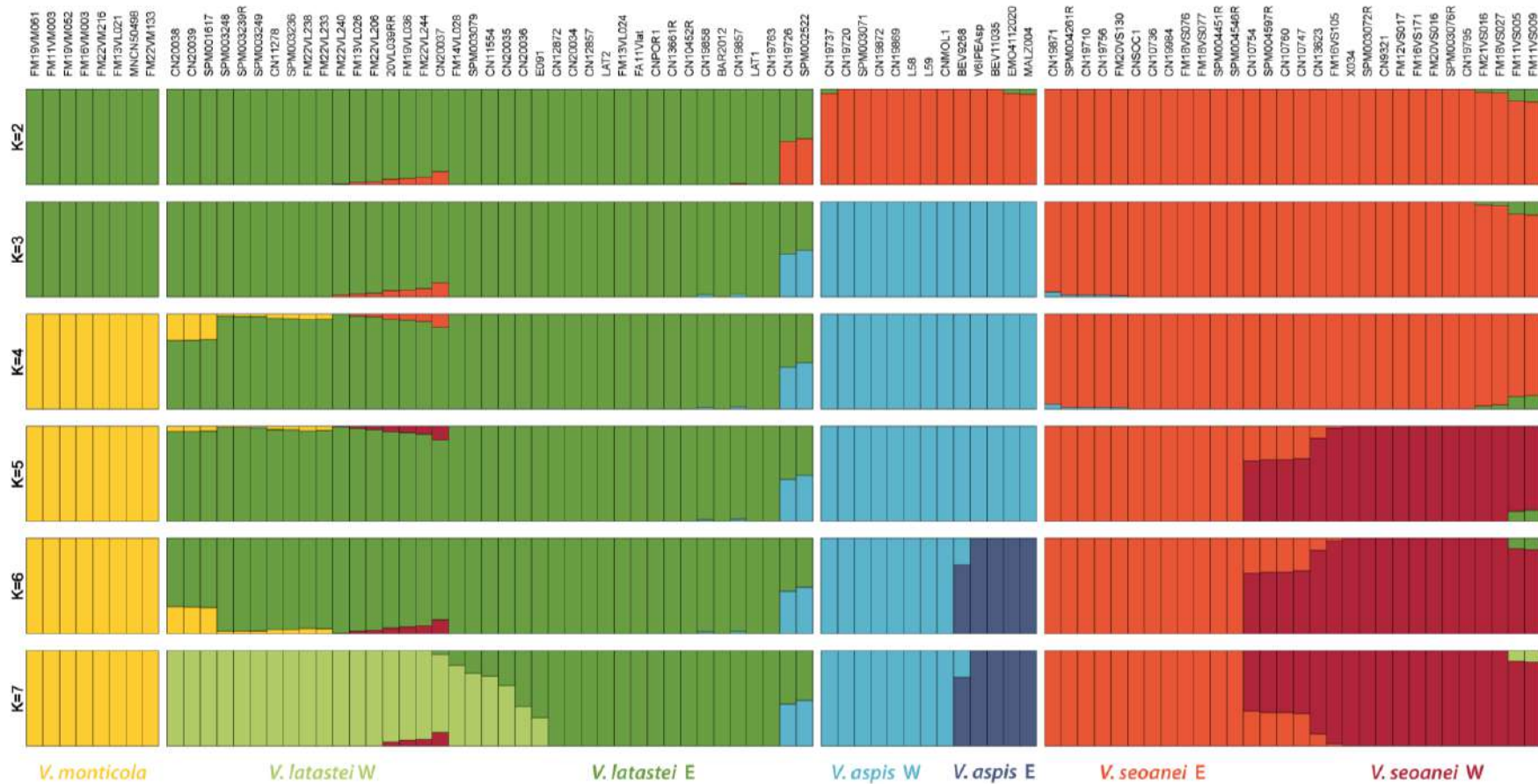


Fig. S25. Admixture results for Western Mediterranean *Vipera* spp., using 1.59M uSNPs and 90 whole genomes, from K=2 to K=7. Most likely K is 4, coinciding with the number of species, although K=7, which is depicted as well in Fig. 5A1-2, is more meaningful as it explores intraspecific structure within *V. latastei*, *V. seoanei* and *V. aspis*. Sample localities are shown in Table S1.

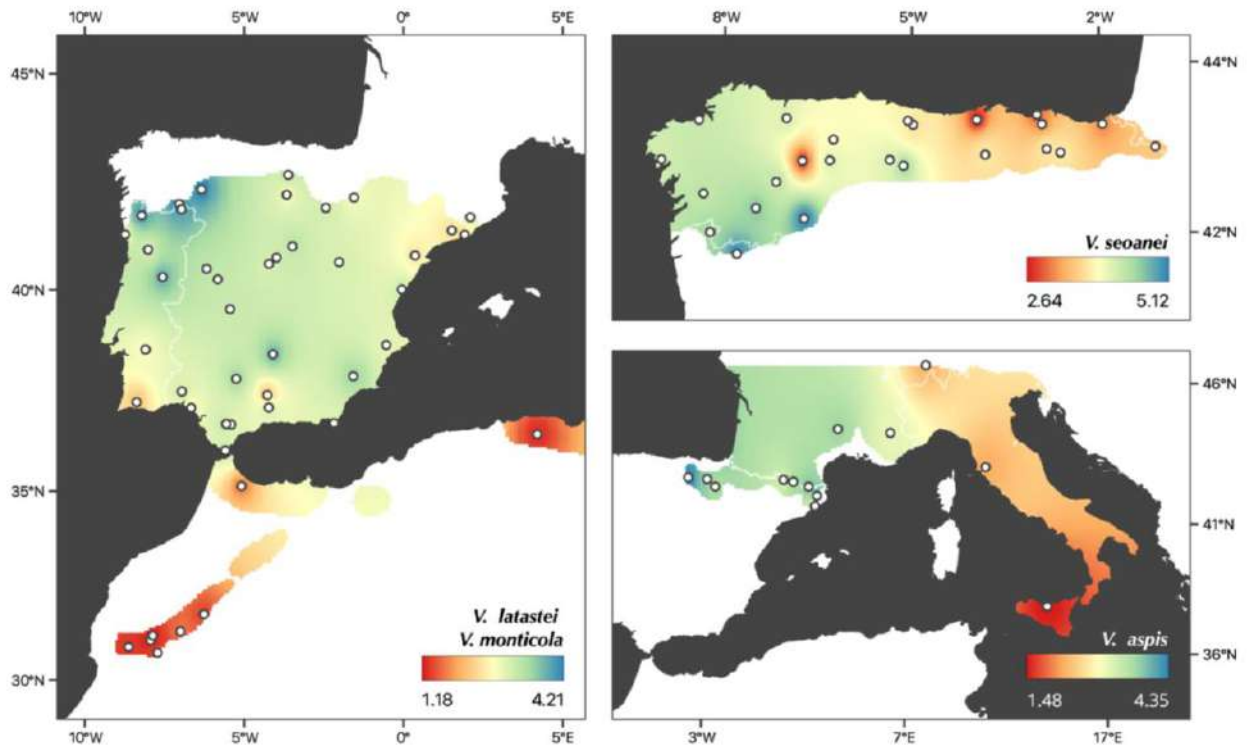


Fig. S26. Genome-wide heterozygosity interpolations per individual for *V. latastei-monticola* on the left, *V. seoanei* on the upper right corner and *V. aspis* on the bottom right. Diversity increases from south to north in *V. latastei-monticola*, and from East to West in *V. aspis* and *V. seoanei*. Highest heterozygosity values for the three Iberia species are found in sympatry areas.

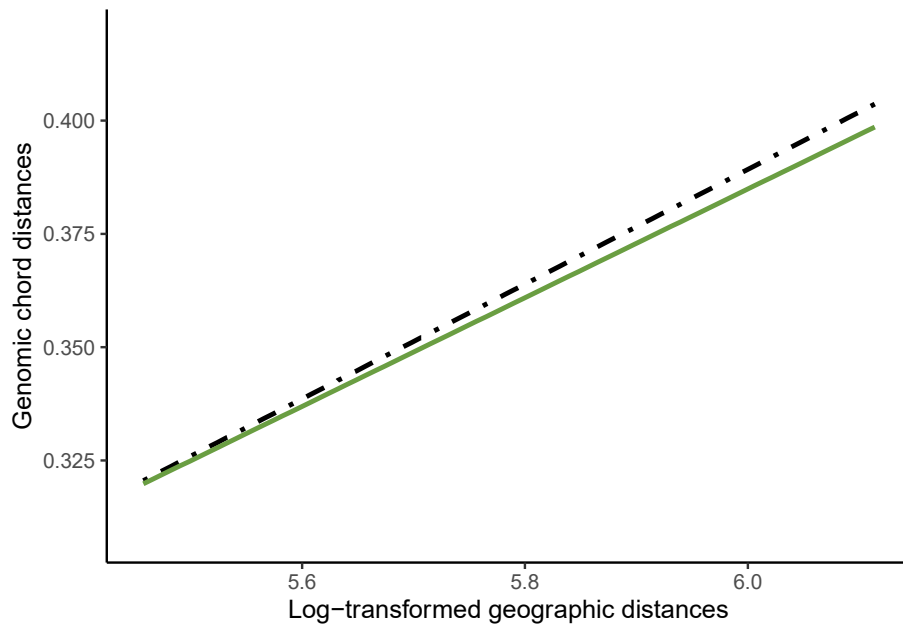


Fig. S27. Relationship between genomic and log-transformed geographical distances in pairs of individuals of the Western and Eastern lineages of *V. latastei* to explore isolation-by-distance (IBD) patterns. Green circles and light green triangles indicate distances between individuals belonging to Eastern and Western lineages, respectively. Black crosses indicate distances between individuals belonging to different intraspecific lineages. The solid green line is the regression line fitted only for within-lineage comparisons (i.e., circles and triangles) and its 95%CI in grey. Dashed black line is the regression fitted for all comparisons, either within- or between-lineage (circles, triangles, and crosses), i.e., the IBD pattern needed to explain lineage differences exclusively due to distance. Both regressions do not significantly differ, therefore, genomic differences between *V. latastei* lineages can be explain by IBD.

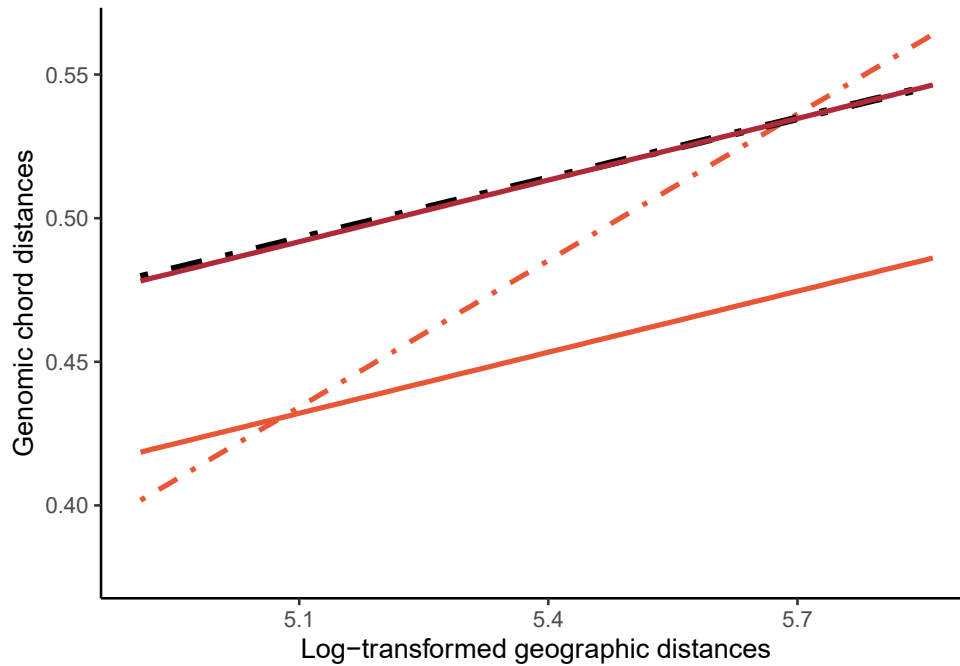


Fig. S28. Relationship between genomic and log-transformed geographical distances in pairs of individuals of the Western and Eastern lineages of *V. seoanei* to explore isolation-by-distance (IBD) patterns. Red circles and maroon triangles indicate distances between individuals belonging to Eastern and Western lineages, respectively. Black crosses indicate distances between individuals belonging to different intraspecific lineages. The solid red line is the regression line fitted only for within-East comparisons (i.e., circles) and its 95%CI in grey, whereas the solid maroon line is fitted for within-West comparisons (i.e., triangles). Dashed red and black lines are the regressions fitted for either within-East or within-West and between-lineage comparisons (i.e., either circles or triangles, plus crosses). The intercept of the red and maroon solid regressions is different, despite their similar slope. The slope of the black and red dashed regressions significantly differs. IBD in the Western lineage can explain differentiation of the whole range, but IBD in the Eastern cannot.

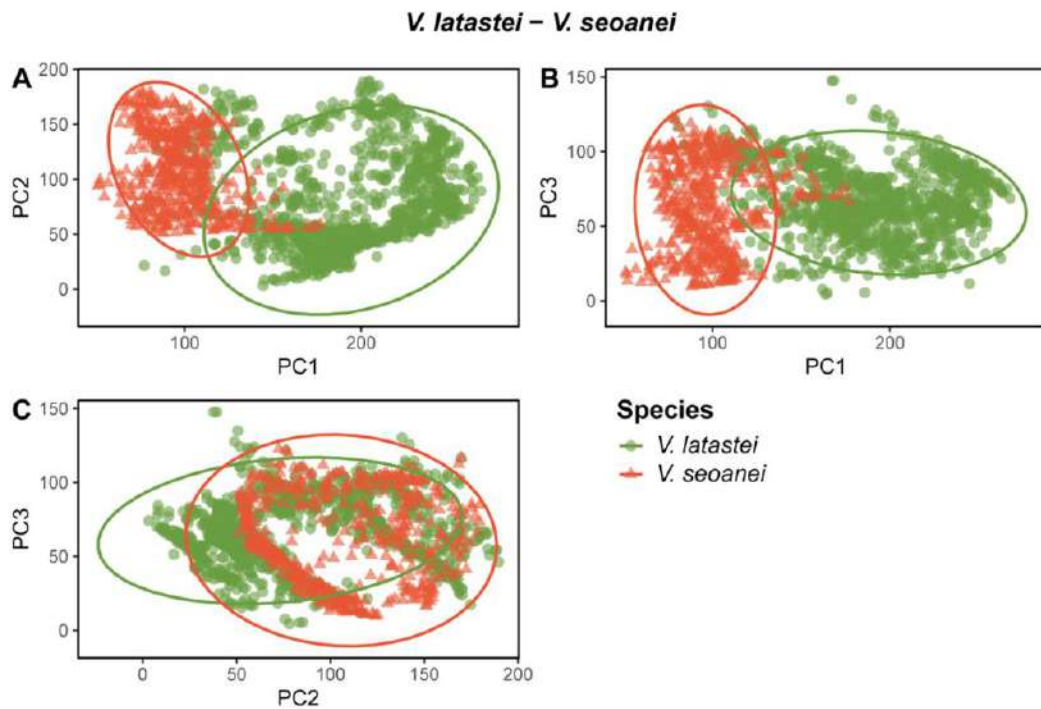


Fig. S29. Environmental PCA for all *V. latastei* (green circles, n=1,105) and all *V. seoanei* (red triangles, n=508). Ecological differences between these species are best explained by PC1, i.e., an Eurosiberian vs. Mediterranean gradient. Dispersion of the occurrences of each species with 95% CI are depicted.

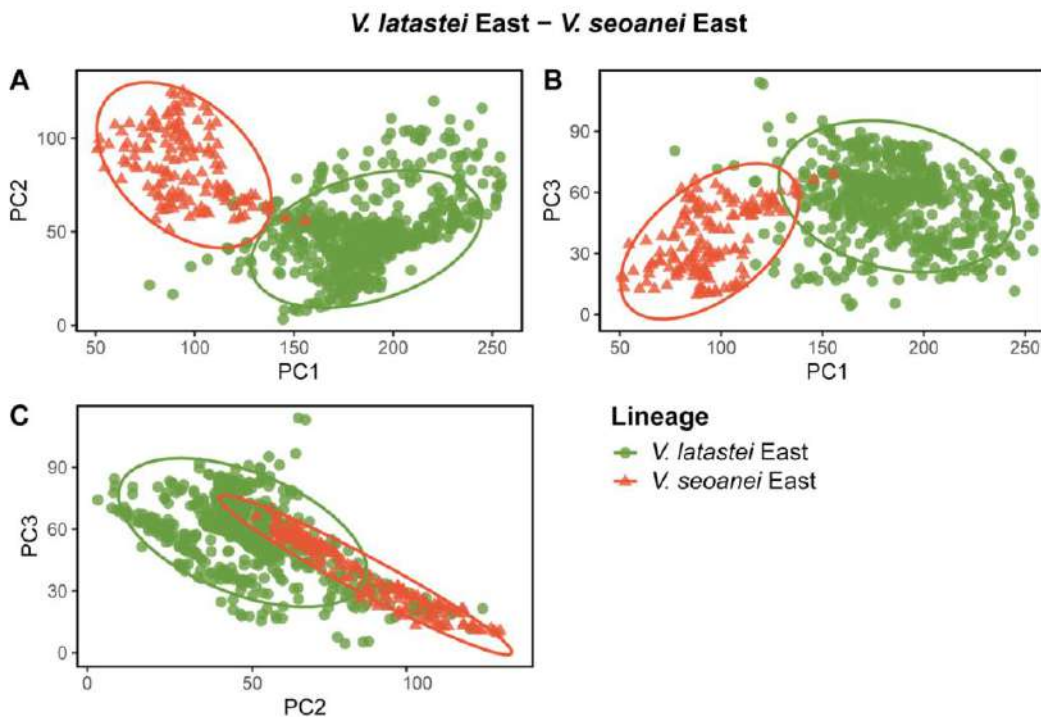


Fig. S30. Environmental PCA for Eastern *V. latastei* (green circles, n=617) and Eastern *V. seoanei* (red triangles, n=146). Ecological differences between these lineages are best explained by PC1, i.e., an Eurosiberian vs. Mediterranean gradient. Dispersion of the occurrences of each species with 95% CI are depicted.

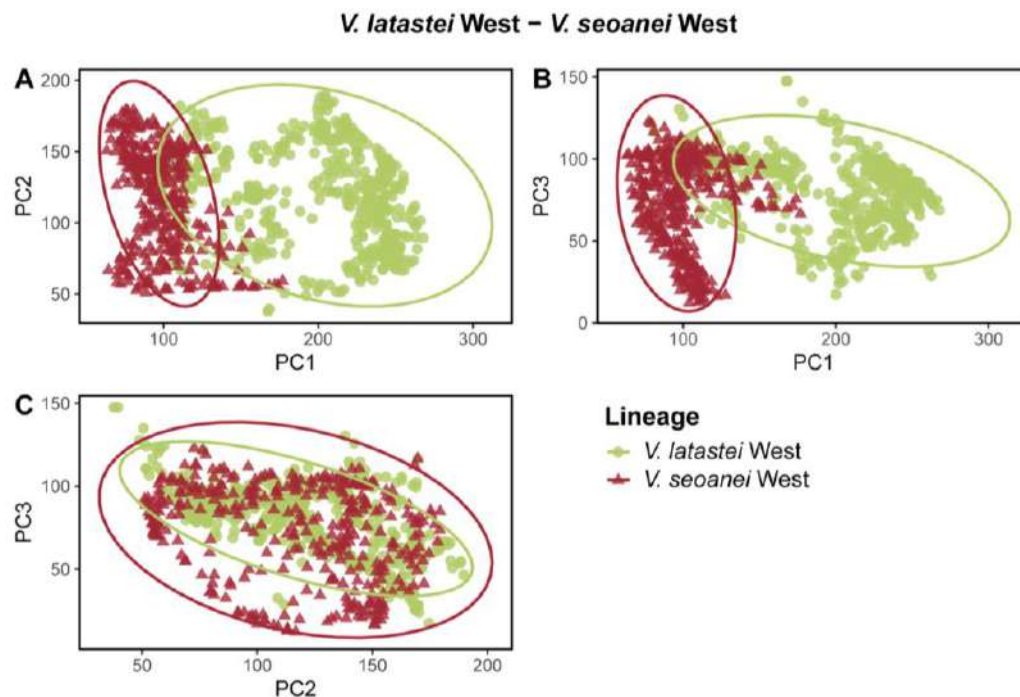


Fig. S31. Environmental PCA for Western *V. latastei* (light green circles, n=488) and Western *V. seoanei* (maroon triangles, n=362). Ecological differences between these lineages are best explained by PC1 vs. PC2. Dispersion of the occurrences of each species with 95% CI are depicted.

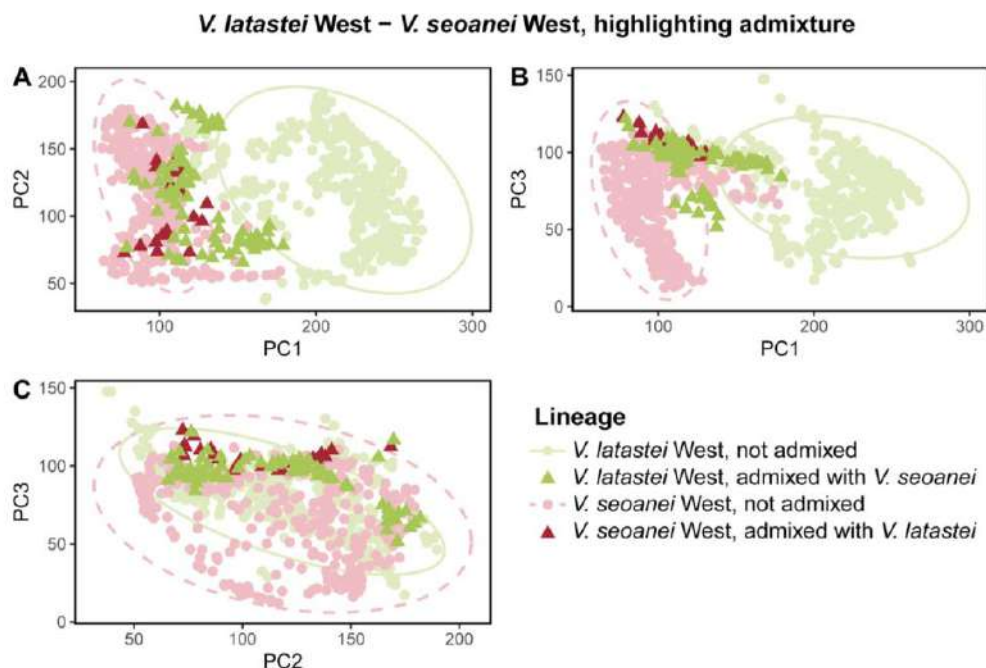


Fig. S32. Environmental PCA for Western *V. latastei* (light green, n=488) and Western *V. seoanei* (maroon, n=362), highlighting admixed individuals. Ecological differences between these lineages are best explained by PC1 vs. PC2. Dispersion of the occurrences of each species for unadmixed populations with 95% CI are depicted (dashed for Western *V. seoanei*). Circles represent occurrences of putatively unadmixed individuals, whereas triangles indicate admixed individuals. *Vipera latastei* individuals from admixed populations seems to inhabit more Eurosiberian areas than expected for its lineage.

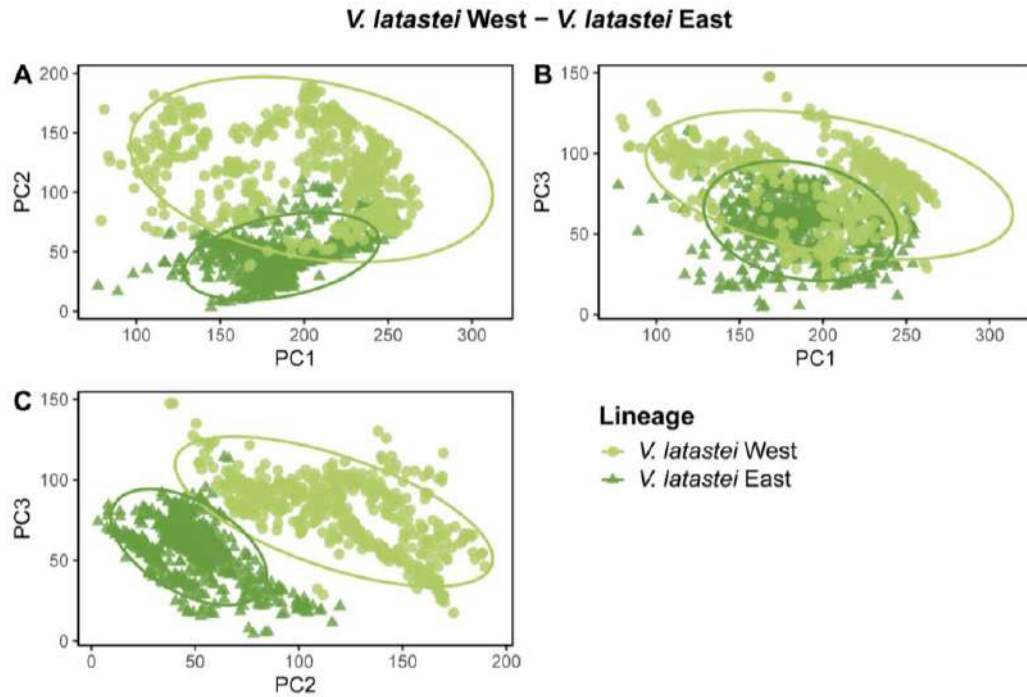


Fig. S33. Environmental PCA for the Western (light green circles, n=488) and Eastern (green triangles, n=617) lineages of *V. latastei*. Ecological differences between these lineages are best explained by PC2 vs. PC3. Dispersion of the occurrences of each species for unadmixed populations with 95% CI are depicted.

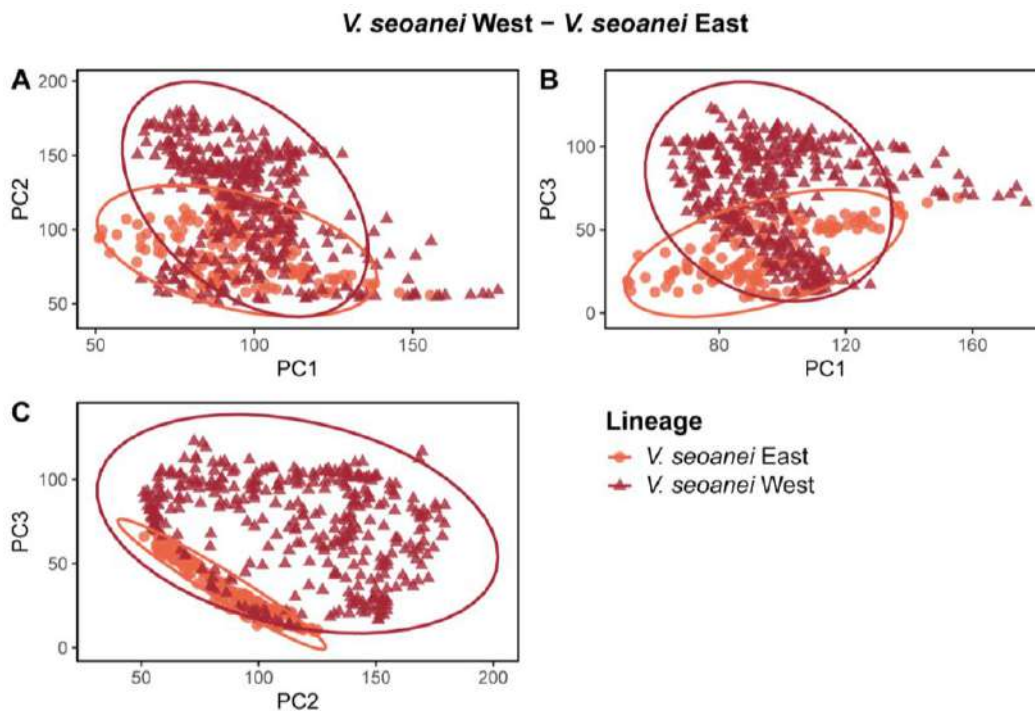


Fig. S34. Environmental PCA for the Western (maroon triangles, n=362) and Eastern (red circles, n=146) lineages of *V. seoanei*. *Vipera seoanei* lineages show great overlap in the three PCs. Dispersion of the occurrences of each species with 95% CI are depicted.

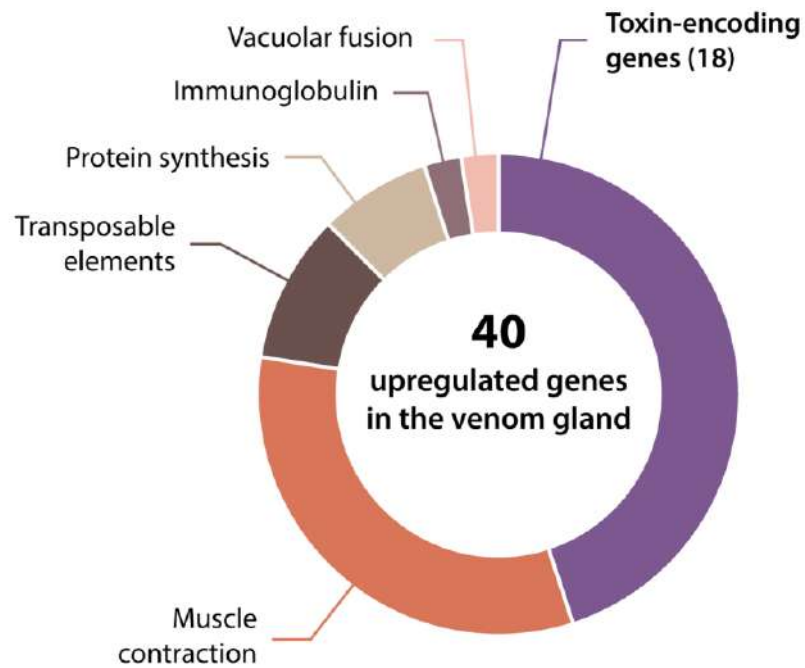


Fig. S35. Pie chart of the differentially upregulated genes in the venom gland of the Iberian vipers. From 40 detected genes, 18 are toxin-encoding genes. The rest are related to muscle contraction (13 genes), transposable elements (4). Protein synthesis (3), vacuolar fusion (1) or immunoglobulins (1).

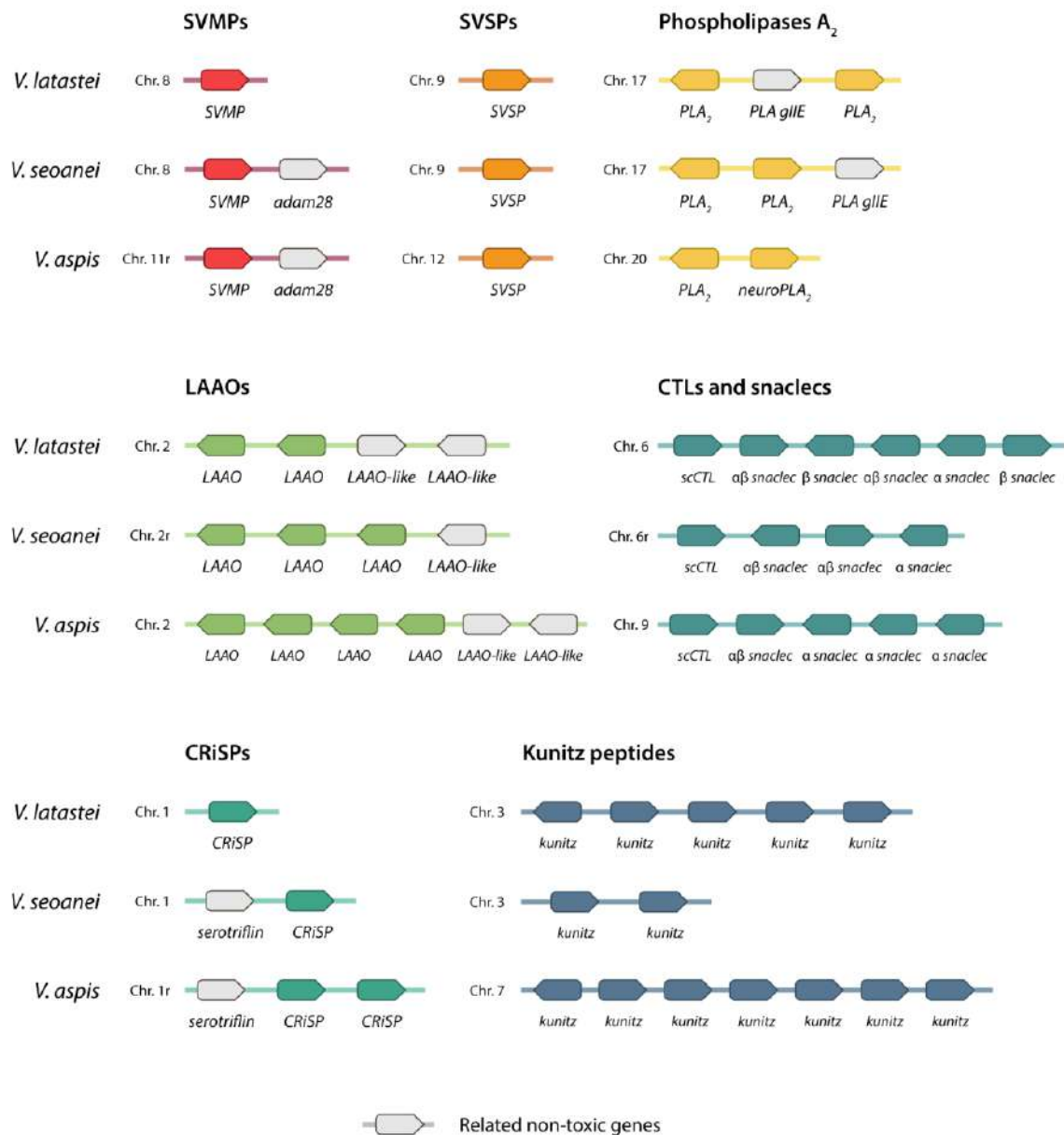


Fig. S36. Microsynteny of the main toxin-encoding gene families for the three Iberian vipers. Genes in grey are non-toxic, but potentially related to the toxin copies. Toxin-encoding genes from other non-main families are not represented.

Table S1. Samples (n=94) used for whole-genome sequencing (WGS). The table shows sample codes, the lineage to which they belong, their approximate origin and their coverage when mapped to rVipLat1. First eight samples were sequenced at high coverage, and first three were used in genome assembly (CNPOR1, CNSOC1, CNMOL1).

Sample	Species	Locality	Country	Coverage
CNPOR1	<i>V. latastei</i>	Els Ports	Spain	59.08
CNSOC1	<i>V. seoanei</i>	Socobio	Spain	50.02
CNMOL1	<i>V. aspis</i>	La Molina	Spain	48.96
MNCN50498	<i>V. monticola</i>	Djurdjurá	Algeria	48.77
CN13667	<i>V. ursinii</i>	Kiskunság, Látó-hegy	Hungary	44.06
SPM002443	<i>V. berus</i>	Brandon	UK	39.61
EMO4112020	<i>V. aspis</i>	Prata	Italy	39.74
CN13858	<i>V. ammodytes</i>	Zagreb	Croatia	41.58
ATP14	<i>V. ammodytes</i>	Ada Bojana	Montenegro	7.05
V6IPEAsp	<i>V. aspis</i>	Grindenwald	Switzerland	7.65
BEV11035	<i>V. aspis</i>	Auzet	France	7.33
BEV9268	<i>V. aspis</i>	Cévennes	France	7.00
CN19720	<i>V. aspis</i>	Girona	Spain	8.14
CN19737	<i>V. aspis</i>	N Burgos	Spain	7.60
SPM003071	<i>V. aspis</i>	Andorra	Andorra	6.06
CN19872	<i>V. aspis</i>	Setcases	Spain	6.12
CN19869	<i>V. aspis</i>	Loza	Spain	6.10
L58	<i>V. aspis</i>	Alto Aneu	Spain	5.09
L59	<i>V. aspis</i>	Montnegre	Spain	5.62
MALZ004	<i>V. aspis</i>	Madonie	Italy	6.49
E091	<i>V. latastei</i>	Sierra Espuña	Spain	8.04
FM14VL028	<i>V. latastei</i>	Ciudad Real	Spain	8.25
FM19VL036	<i>V. latastei</i>	O Pereiro	Spain	7.97
FM22VL206	<i>V. latastei</i>	Santa Cruz	Portugal	7.38
FM22VL233	<i>V. latastei</i>	Évora	Portugal	7.34
FM22VL238	<i>V. latastei</i>	Faro	Portugal	5.04
FM22VL240	<i>V. latastei</i>	Cocharil	Portugal	7.63
FM22VL244	<i>V. latastei</i>	Gerês	Portugal	9.23
CN11278	<i>V. latastei</i>	Doñana	Spain	6.63
CN20035	<i>V. latastei</i>	Hervás	Spain	6.04
CN20036	<i>V. latastei</i>	Navezuelas	Spain	18.93
CN20037	<i>V. latastei</i>	Filiel	Spain	17.86
CN12872	<i>V. latastei</i>	Alcoi	Spain	16.97
CN11554	<i>V. latastei</i>	Cabo de Gata	Spain	17.53
CN12857	<i>V. latastei</i>	Navacerrada	Spain	16.72
SPM003236	<i>V. latastei</i>	Cobujón	Spain	6.76
SPM003248	<i>V. latastei</i>	Sierra de Loja	Spain	6.66
SPM003249	<i>V. latastei</i>	Sierra de Hornachuelos	Spain	7.09
CN19763	<i>V. latastei</i>	N Burgos	Spain	6.58
CN13661R	<i>V. latastei</i>	Alt Penedès	Spain	5.28
CN19857	<i>V. latastei</i>	Soria	Spain	6.39
FA11Vlat	<i>V. latastei</i>	Castellón	Spain	7.02

SPM003079	<i>V. latastei</i>	Peña de Francia	Spain	6.75
SPM003239R	<i>V. latastei</i>	Carcabuey	Spain	5.05
SPM001617	<i>V. latastei</i>	Villaluenga del Rosario	Spain	6.46
FM13VL026	<i>V. latastei</i>	Montemuro	Portugal	5.80
FM13VL024	<i>V. latastei</i>	Peñalén	Spain	6.14
BAR2012	<i>V. latastei</i>	Bárdenas Reales	Spain	6.10
LAT1	<i>V. latastei</i>	Burgos	Spain	6.58
CN10452R	<i>V. latastei</i>	Garraf	Spain	7.79
LAT2	<i>V. latastei</i>	La Hiruela	Spain	7.04
20VL039RR	<i>V. latastei</i>	Vila do Conde	Portugal	6.76
CN19858	<i>V. latastei</i>	Moianés	Spain	7.16
CN20038	<i>V. latastei</i>	Tarifa	Spain	7.01
CN20039	<i>V. latastei</i>	Hurones	Spain	6.94
CN20034	<i>V. latastei</i>	Peguerinos	Spain	6.27
CN19726	<i>V. latastei x aspis</i>	N Burgos	Spain	6.24
SPM002522	<i>V. latastei x aspis</i>	Boumort	Spain	6.62
FM22VM133	<i>V. monticola</i>	Talasantane	Morocco	10.76
FM22VM216	<i>V. monticola</i>	Tamda n'Oughmar	Morocco	9.01
FM19VM061	<i>V. monticola</i>	Plateau du Tichka	Morocco	6.06
FM11VM003	<i>V. monticola</i>	Toubkal	Morocco	5.42
FM19VM052	<i>V. monticola</i>	Oukeimedem	Morocco	6.26
FM16VM003	<i>V. monticola</i>	Jebel Sirwa	Morocco	5.80
FM13VL021	<i>V. monticola</i>	Jebel Azourki	Morocco	6.50
CN10736	<i>V. seoanei</i>	Sopelana	Spain	8.33
CN10760	<i>V. seoanei</i>	Picos de Europa	Spain	12.10
CN13623	<i>V. seoanei</i>	Cistierna	Spain	8.64
CN19710	<i>V. seoanei</i>	Gorbea	Spain	7.21
CN19756	<i>V. seoanei</i>	Gorbea	Spain	9.25
CN19984	<i>V. seoanei</i>	Gorbea	Spain	8.44
CN9321	<i>V. seoanei</i>	Salientes	Spain	7.24
FM11VS005	<i>V. seoanei</i>	Vila Real	Portugal	8.10
FM11VS009	<i>V. seoanei</i>	Sanabria	Spain	7.09
FM12VS017	<i>V. seoanei</i>	Alto de la Garganta	Spain	7.30
FM16VS105	<i>V. seoanei</i>	Boñar	Spain	6.98
FM16VS171	<i>V. seoanei</i>	Cernedo	Spain	7.14
FM18VS027	<i>V. seoanei</i>	Gerês	Portugal	7.91
FM18VS076	<i>V. seoanei</i>	Hoces del Ebro	Spain	6.82
FM18VS077	<i>V. seoanei</i>	Hoces del Ebro	Spain	6.68
FM20VS016	<i>V. seoanei</i>	Feáns	Spain	7.55
FM20VS130	<i>V. seoanei</i>	Hoces del Ebro	Spain	9.06
FM21VS016	<i>V. seoanei</i>	Montederramo	Spain	7.71
CN10747	<i>V. seoanei</i>	Picos de Europa	Spain	6.14
CN19871	<i>V. seoanei</i>	Irati	Spain	5.54
SPM004261R	<i>V. seoanei</i>	Gamboa	Spain	6.62
SPM003076R	<i>V. seoanei</i>	Caurel	Spain	6.33
CN10754	<i>V. seoanei</i>	Picos de Europa	Spain	6.75
CN19795	<i>V. seoanei</i>	Tejedo de Ancares	Spain	5.15
SPM004451R	<i>V. seoanei</i>	Artxanda	Spain	6.15
SPM004546R	<i>V. seoanei</i>	Donostia	Spain	5.91

SPM003072R	<i>V. seoanei</i>	Somiedo	Spain	6.67
SPM004597R	<i>V. seoanei</i>	Covadonga	Spain	9.86
X034	<i>V. seoanei</i>	Muros	Spain	7.42

Table S2. Whole-genome datasets used in this study. The table shows whether the dataset have high- or low-coverage, the number of individuals (n), species (spp.), SNPs (if not all sites) and the analyses that were performed with them.

Coverage	Dataset Name	n	spp.	SNPs	Analyses	Comments
High	<i>high_coverage</i>	8	7	all sites	PSMC, ROHs	
High	<i>high_coverage+out</i>	8	8	all sites	Twisst	With <i>Crotalus</i> as outgroup
Low	<i>all_sites</i>	95	7	all sites	Heterozygosity	
Low	<i>westmed</i>	90	4	1,588,578	PCA, Admixture	0.5-kbp thinning
Low	<i>latastei-monticola</i>	45	2	1,308,394	PCA	0.5-kbp thinning, no hybrids
Low	<i>latastei</i>	37	1	1,309,324	PCA, IBD	0.5-kbp thinning, no hybrids
Low	<i>monticola</i>	8	1	326,645	PCA	0.5-kbp thinning
Low	<i>aspis</i>	13	1	1,152,108	PCA, IBD	0.5-kbp thinning, no hybrids
Low	<i>seoanei</i>	30	1	1,504,794	PCA	0.5-kbp thinning
Low	<i>linked_SNPs</i>	86	8	12,645,411	Dsuite	No LD-pruned, no hybrids, no intraspecific admixture, with <i>Crotalus</i> as outgroup
Low	<i>snapp_species</i>	8	8	1,225,190	SNAPP	0.5-kbp thinning, with <i>Crotalus</i> as outgroup, 0% missingness
Low	<i>snapp_subspecies</i>	15	7	126,083	SNAPP	10-kbp thinning, 0% missingness
Low	<i>treemix_subspecies</i>	15	7	1,095,590	TreeMix	0.5-kbp thinning, 0% missingness

Table S3. Toxin families in *Vipera*, with the number of gene copies found for each one, the relative expression (*sensu* Giribaldi et al., 2020), and the normalized ubiquity of diversifying and purifying selection, corrected by number of copies and length of the alignment. These values have been used in linear regression shown in Fig. 6D.

Family	Copies	Expression (%)	Diversification	Purification
SVMP	3	28.31	0	1
PLA	6	15.45	0.45	0.48
snaclec	12	12.28	1	0.67
CRISP	4	0.79	0.12	0.20
LAAO	9	1.07	0.21	0.46
SVVEGF	3	3.19	0	0.35
SVSP	3	13.31	0	0.44
SVNGF	4	1.89	0	0.07
Kunitz	12	9.87	0.25	0.52
SVMPi	4	13.2	0	0.54
5'-nucleotidase	3	0.25	0	0.65

Table S4. Assembly and annotation quality and completeness parameters of the three genomes assembled in this study (*V. latastei*, *V. seoanei*, and *V. aspis*) and other venomous snake reference genomes. The genomes here presented stand out for their contig N50. BUSCO stats refer to single-copied complete orthologs of vertebrata odb10 (% out of 3,354 genes) for the assembly or to complete orthologs (single-copied or not) of vertebrata for the annotation, respectively. annot. = annotated; Gb = Gigabases; N. = number; Mb = Megabases.

	<i>V. latastei</i>	<i>V. seoanei</i>	<i>V. aspis</i>	<i>V. ursinii</i>	<i>C. tigris</i>	<i>N. naja</i>
Genome size (Gb)	1.63	1.57	1.59	1.62	1.59	1.79
N. of Scaffolds	56	18	26	384	160	1,897
Scaffold N50 (Mb)	222.4	236.1	113.8	212.8	207.7	223.4
Scaffold L50	3	3	6	3	-	-
Contig N50 (Mb)	44.84	75.89	35.32	2.13	2.11	0.3
Assembly BUSCO (%)	97.1	96.2	96.1	-	-	-
GC%	40.96	40.94	40.96	40.93	39.8	40.5
Structurally annot. genes	59,409	61,622	57,979	-	-	-
Annotation BUSCO (%)	98.6	97.8	97.6	-	-	-
Functionally annot. genes	26,312	25,391	26,213	-	-	-

Table S5. Introgression results from ABBA-BABA tests implemented on Dsuite. Results are obtained from a dataset of 86 samples, comprising all studied *Vipera* lineages and excluding individuals from secondary contacts (see Admixture results), which yielded a total of 12,645,411 SNPs. Only trios showing significant results regarding the robust clustering threshold ($p\text{-value} < 0.001$) and Z-score $|Z| > 3$ are shown. Sensitive and robust clustering $p\text{-values}$ are calculated based on Kolmogorov-Smirnov tests to correct for

homoplasies. amm stands for *V. ammodytes*; ber for *V. berus*; urs for *V. ursinii*; aspasp for *V. aspis aspis*; aspzin for *V. aspis zinnikeri*; asphug for *V. aspis hugyi*; aspfra for *V. aspis francisciredi*; latlat for *V. latastei latastei*; latgad for *V. latastei gaditana*; lataru for *V. latastei arundana*; monN for the Northern cluster of *V. monticola*; monS for the Southern cluster of *V. monticola*; seoE for the Eastern cluster of *V. seoanei* and seoW for the Western cluster of the same species (see Admixture results).

P1	P2	P3	Dstatistic	Z-score	p-value	f4-ratio	Sensitive clustering	Robust clustering	BBA	ABBA	BABA
amm	aspasp	ber	0.217	26.854	<0.0001	0.1	<0.0001	0.0004	219428	240074	154569
amm	aspasp	latgad	0.022	3.186	0.0014	0.009	<0.0001	0.0004	299566	167236	159967
amm	aspasp	latlat	0.043	5.248	<0.0001	0.017	<0.0001	0.0001	297957	172900	158719
amm	aspasp	seoE	0.197	19.04	<0.0001	0.08	<0.0001	<0.0001	259147	236950	159034
amm	aspasp	seoW	0.162	20.281	<0.0001	0.074	<0.0001	<0.0001	264977	223004	160763
amm	aspfra	ber	0.132	15.05	<0.0001	0.054	<0.0001	0.0002	219381	192091	147371
amm	aspfra	seoE	0.125	15.147	<0.0001	0.045	<0.0001	<0.0001	256252	192842	149857
amm	aspfra	seoW	0.101	13.694	<0.0001	0.042	<0.0001	<0.0001	260742	184319	150377
amm	aspfra	urs	0.082	7.907	<0.0001	0.045	<0.0001	<0.0001	211949	172856	146792
amm	asphug	ber	0.123	13.389	<0.0001	0.05	<0.0001	<0.0001	217180	186356	145472
amm	asphug	seoE	0.118	14.361	<0.0001	0.042	<0.0001	0.0005	253650	187715	147975
amm	asphug	seoW	0.095	12.714	<0.0001	0.039	<0.0001	0.0004	257921	179610	148376
amm	asphug	urs	0.073	6.777	<0.0001	0.04	<0.0001	<0.0001	209998	168024	145027
amm	aspzin	ber	0.271	23.428	<0.0001	0.136	<0.0001	<0.0001	203384	271894	156007
amm	aspzin	lataru	0.157	8.393	<0.0001	0.062	<0.0001	<0.0001	277905	211905	154294
amm	aspzin	latgad	0.17	9.009	<0.0001	0.079	<0.0001	<0.0001	275760	217314	154219
amm	aspzin	latlat	0.192	9.514	<0.0001	0.086	<0.0001	<0.0001	273771	225299	152596
amm	aspzin	monS	0.142	8.014	<0.0001	0.05	<0.0001	<0.0001	279242	206172	154847
amm	aspzin	monN	0.142	8.021	<0.0001	0.05	<0.0001	<0.0001	276853	205616	154482
amm	aspzin	seoE	0.344	18.115	<0.0001	0.162	<0.0001	<0.0001	233204	309966	151168
amm	aspzin	seoW	0.312	18.243	<0.0001	0.166	<0.0001	<0.0001	239319	292256	153181
amm	aspzin	urs	0.121	14.413	<0.0001	0.076	<0.0001	<0.0001	203507	207862	162889
amm	ber	monS	0.156	23.472	<0.0001	0.057	<0.0001	0.0006	226440	210182	153305
amm	ber	monN	0.156	23.902	<0.0001	0.057	<0.0001	0.0003	224317	209582	153036
amm	seoE	lataru	0.438	73.146	<0.0001	0.247	<0.0001	<0.0001	187523	373556	145931
amm	seoW	lataru	0.467	79.751	<0.0001	0.271	<0.0001	<0.0001	180167	392817	142677
amm	seoE	latgad	0.466	79.509	<0.0001	0.313	<0.0001	<0.0001	183034	394159	143528
amm	seoW	latgad	0.493	86.412	<0.0001	0.341	<0.0001	<0.0001	175786	413332	140384
amm	seoE	latlat	0.447	77.089	<0.0001	0.279	<0.0001	0.0002	185118	381503	145978
amm	seoW	latlat	0.473	82.9	<0.0001	0.304	<0.0001	<0.0001	178057	399779	143021
amm	seoE	monS	0.416	65.352	<0.0001	0.204	<0.0001	<0.0001	190560	358995	148199
amm	seoW	monS	0.445	72.423	<0.0001	0.226	<0.0001	<0.0001	183221	377847	144962
amm	seoE	monN	0.416	65.754	<0.0001	0.207	<0.0001	<0.0001	188345	358276	147812
amm	seoW	monN	0.446	72.621	<0.0001	0.229	<0.0001	<0.0001	181033	377037	144601
seoE	urs	amm	0.123	20.921	<0.0001	0.085	<0.0001	<0.0001	283681	174271	136017
seoW	urs	amm	0.136	22.158	<0.0001	0.093	<0.0001	<0.0001	282854	176143	133974

asphug	aspfra	aspasp	0.08	20.694	<0.0001	0.035	<0.0001	0.0043	302166	90602.1	77169
aspzin	aspasp	aspfra	0.222	15.876	<0.0001	0.162	<0.0001	<0.0001	317264	194230	123727
aspfra	aspasp	ber	0.149	10.812	<0.0001	0.049	<0.0001	0.0006	512130	147663	109444
aspfra	aspasp	seoE	0.127	9.542	<0.0001	0.036	<0.0001	0.0014	561441	144528	111851
aspfra	aspasp	seoW	0.107	8.59	<0.0001	0.034	<0.0001	0.003	573695	137513	111029
aspfra	aspasp	urs	0.058	7.209	<0.0001	0.023	<0.0001	0.0067	515976	118702	105760
aspzin	aspasp	asphug	0.218	15.849	<0.0001	0.142	<0.0001	<0.0001	324578	188829	121220
asphug	aspasp	ber	0.159	10.905	<0.0001	0.053	<0.0001	0.0024	496737	149799	108702
asphug	aspasp	seoE	0.136	9.889	<0.0001	0.039	<0.0001	0.0069	545030	146705	111492
asphug	aspasp	seoW	0.114	9.035	<0.0001	0.037	<0.0001	0.0056	556777	139457	110815
aspasp	aspzin	lataru	0.237	13.379	<0.0001	0.058	<0.0001	<0.0001	760719	140788	86807.1
aspasp	aspzin	latgad	0.243	13.655	<0.0001	0.07	<0.0001	<0.0001	756182	143706	87604.4
aspasp	aspzin	latlat	0.25	13.716	<0.0001	0.07	<0.0001	<0.0001	749671	147201	88393
aspasp	aspzin	monS	0.226	13.237	<0.0001	0.049	<0.0001	<0.0001	764227	137123	86592.2
aspasp	aspzin	monN	0.227	13.356	<0.0001	0.05	<0.0001	<0.0001	760875	136837	86254.4
aspasp	aspzin	seoE	0.282	17.119	<0.0001	0.09	<0.0001	<0.0001	661396	184639	103430
aspasp	aspzin	seoW	0.277	16.863	<0.0001	0.099	<0.0001	<0.0001	678767	178070	100915
aspasp	ber	lataru	0.171	19.351	<0.0001	0.072	<0.0001	0.0013	312272	222328	157479
aspasp	ber	latgad	0.196	22.257	<0.0001	0.099	<0.0001	0.0008	306233	233096	156769
aspasp	ber	latlat	0.171	18.85	<0.0001	0.083	<0.0001	0.0013	304790	229586	162444
aspasp	ber	monS	0.151	16.806	<0.0001	0.056	<0.0001	0.0009	316639	214461	158307
aspasp	ber	monN	0.151	17.229	<0.0001	0.057	<0.0001	0.0017	314296	213871	157818
aspasp	seoE	lataru	0.427	58.956	<0.0001	0.243	<0.0001	<0.0001	266972	375728	150909
aspasp	seoW	lataru	0.452	67.645	<0.0001	0.268	<0.0001	<0.0001	245961	397186	149779
aspasp	urs	lataru	0.09	11.038	<0.0001	0.036	<0.0001	<0.0001	276827	193687	161832
aspasp	seoE	latgad	0.45	61.71	<0.0001	0.306	<0.0001	<0.0001	260056	393983	149445
aspasp	seoW	latgad	0.473	70.421	<0.0001	0.334	<0.0001	<0.0001	239187	415389	148461
aspasp	urs	latgad	0.094	11.615	<0.0001	0.044	<0.0001	<0.0001	273138	197332	163359
aspasp	seoE	latlat	0.414	55.069	<0.0001	0.266	<0.0001	<0.0001	260777	379907	157464
aspasp	seoW	latlat	0.437	60.683	<0.0001	0.292	<0.0001	<0.0001	240259	400579	156833
aspasp	urs	latlat	0.072	8.032	<0.0001	0.033	<0.0001	<0.0001	271764	195229	169070
aspasp	seoE	monS	0.409	54.169	<0.0001	0.204	<0.0001	<0.0001	272056	363127	152373
aspasp	seoW	monS	0.435	62.634	<0.0001	0.225	<0.0001	<0.0001	251017	384135	151219
aspasp	urs	monS	0.084	10.892	<0.0001	0.03	<0.0001	<0.0001	279544	190410	160963
aspasp	seoE	monN	0.41	55.366	<0.0001	0.207	<0.0001	<0.0001	269674	362430	151771
aspasp	seoW	monN	0.436	64.107	<0.0001	0.228	<0.0001	<0.0001	248701	383349	150651
aspasp	urs	monN	0.084	11.159	<0.0001	0.03	<0.0001	0.0002	277206	189860	160391
asphug	aspfra	aspzin	0.073	18.345	<0.0001	0.03	<0.0001	0.0037	362513	83084.7	71735.4
aspfra	aspzin	ber	0.217	13.554	<0.0001	0.086	<0.0001	<0.0001	456421	189871	122092
aspfra	aspzin	lataru	0.208	9.267	<0.0001	0.062	<0.0001	<0.0001	568568	162126	106224
aspfra	aspzin	latgad	0.218	9.615	<0.0001	0.077	<0.0001	<0.0001	565194	166041	106522
aspfra	aspzin	latlat	0.235	9.978	<0.0001	0.08	<0.0001	<0.0001	560835	171382	106135
aspfra	aspzin	monS	0.193	9.002	<0.0001	0.051	<0.0001	<0.0001	570838	157427	106598

aspfra	aspzin	monN	0.193	9.1	<0.0001	0.051	<0.0001	<0.0001	567936	157096	106203
aspfra	aspzin	seoE	0.327	13.553	<0.0001	0.122	<0.0001	<0.0001	493700	225942	114494
aspfra	aspzin	seoW	0.308	13.155	<0.0001	0.13	<0.0001	<0.0001	506119	215162	113832
aspfra	ber	lataru	0.178	26.057	<0.0001	0.076	<0.0001	0.0079	266750	217633	151755
aspfra	ber	latgad	0.207	30.714	<0.0001	0.104	<0.0001	0.0044	261869	229096	150548
aspfra	ber	latlat	0.191	28.606	<0.0001	0.092	<0.0001	0.0083	261255	226497	153790
aspfra	ber	monS	0.154	21.163	<0.0001	0.057	<0.0001	0.0054	269957	208946	153172
aspfra	ber	monN	0.154	21.805	<0.0001	0.058	<0.0001	0.0092	267774	208368	152707
aspfra	seoE	lataru	0.43	61.276	<0.0001	0.246	<0.0001	0.0011	229416	367136	146206
aspfra	seoW	lataru	0.457	70.105	<0.0001	0.27	<0.0001	<0.0001	214371	387048	144240
aspfra	urs	lataru	0.101	18.478	<0.0001	0.039	<0.0001	0.0003	252666	184948	151019
aspfra	seoE	latgad	0.456	65.405	<0.0001	0.311	<0.0001	0.0007	223712	385898	144246
aspfra	seoW	latgad	0.48	74.379	<0.0001	0.339	<0.0001	<0.0001	208823	405779	142438
aspfra	urs	latgad	0.11	20.982	<0.0001	0.05	<0.0001	0.0001	249561	189003	151493
aspfra	seoE	latlat	0.427	61.395	<0.0001	0.274	<0.0001	0.0057	225074	372889	149579
aspfra	seoW	latlat	0.452	68.798	<0.0001	0.299	<0.0001	<0.0001	210466	391986	148053
aspfra	urs	latlat	0.096	17.169	<0.0001	0.042	<0.0001	0.001	248871	187631	154673
aspfra	seoE	monS	0.41	56.24	<0.0001	0.205	<0.0001	0.0008	233230	353759	148187
aspfra	seoW	monS	0.437	64.942	<0.0001	0.226	<0.0001	<0.0001	218217	373304	146254
aspfra	urs	monS	0.09	16.006	<0.0001	0.031	<0.0001	0.0002	254601	181034	151117
aspfra	seoE	monN	0.41	57.418	<0.0001	0.208	<0.0001	0.0003	231007	353073	147578
aspfra	seoW	monN	0.438	66.161	<0.0001	0.229	<0.0001	<0.0001	216046	372524	145679
aspfra	urs	monN	0.09	16.789	<0.0001	0.032	<0.0001	0.0006	252403	180551	150591
asphug	aspzin	ber	0.226	13.879	<0.0001	0.09	<0.0001	<0.0001	443276	191001	120694
asphug	aspzin	lataru	0.206	9.392	<0.0001	0.062	<0.0001	<0.0001	551842	161725	106560
asphug	aspzin	latgad	0.216	9.713	<0.0001	0.076	<0.0001	<0.0001	548561	165635	106856
asphug	aspzin	latlat	0.232	9.993	<0.0001	0.08	<0.0001	<0.0001	544299	171004	106581
asphug	aspzin	monS	0.189	9.036	<0.0001	0.05	<0.0001	<0.0001	553884	156895	106998
asphug	aspzin	monN	0.189	9.079	<0.0001	0.051	<0.0001	<0.0001	550999	156487	106720
asphug	aspzin	seoE	0.332	13.72	<0.0001	0.125	<0.0001	<0.0001	479814	226759	113773
asphug	aspzin	seoW	0.312	13.401	<0.0001	0.132	<0.0001	<0.0001	491744	215796	113275
asphug	ber	latgad	0.207	31.052	<0.0001	0.104	<0.0001	0.0067	255618	226631	149035
asphug	ber	monS	0.153	21.146	<0.0001	0.057	<0.0001	0.0062	263370	206495	151717
asphug	ber	monN	0.153	21.731	<0.0001	0.057	<0.0001	0.0082	261160	205858	151363
asphug	seoE	lataru	0.429	61.358	<0.0001	0.246	<0.0001	0.0002	224174	362952	144858
asphug	seoW	lataru	0.456	70.749	<0.0001	0.27	<0.0001	<0.0001	209738	382645	142850
asphug	urs	lataru	0.101	19.066	<0.0001	0.039	<0.0001	0.0012	246698	182799	149366
asphug	seoE	latgad	0.455	65.746	<0.0001	0.31	<0.0001	0.0001	218625	381552	142961
asphug	seoW	latgad	0.48	75.14	<0.0001	0.338	<0.0001	<0.0001	204332	401223	141096
asphug	urs	latgad	0.11	21.744	<0.0001	0.05	<0.0001	0.0009	243740	186848	149864
asphug	seoE	latlat	0.426	61.296	<0.0001	0.273	<0.0001	0.0009	219911	368659	148235
asphug	seoW	latlat	0.451	69.18	<0.0001	0.298	<0.0001	<0.0001	205884	387558	146635
asphug	urs	latlat	0.096	17.94	<0.0001	0.042	<0.0001	0.0051	243026	185481	153044

asphug	seoE	monS	0.408	56.031	<0.0001	0.204	<0.0001	0.0001	227822	349666	146967
asphug	seoW	monS	0.436	65.016	<0.0001	0.225	<0.0001	<0.0001	213404	368933	144976
asphug	urs	monS	0.089	16.364	<0.0001	0.031	<0.0001	0.0011	248527	178837	149649
asphug	seoE	monN	0.409	57.105	<0.0001	0.207	<0.0001	<0.0001	225575	348892	146490
asphug	seoW	monN	0.436	66.213	<0.0001	0.229	<0.0001	<0.0001	211198	368061	144517
asphug	urs	monN	0.089	16.835	<0.0001	0.031	<0.0001	0.0016	246289	178259	149225
seoW	ber	aspzin	0.046	4.775	<0.0001	0.025	<0.0001	<0.0001	420215	188751	171994
urs	ber	aspzin	0.179	13.989	<0.0001	0.089	<0.0001	<0.0001	274862	203821	141893
aspzin	seoE	lataru	0.341	24.868	<0.0001	0.196	<0.0001	0.0002	308281	335871	164929
aspzin	seoW	lataru	0.371	28.897	<0.0001	0.222	<0.0001	<0.0001	283212	357342	163805
aspzin	seoE	latgad	0.365	26.781	<0.0001	0.254	<0.0001	0.0002	299738	352496	163949
aspzin	seoW	latgad	0.393	30.733	<0.0001	0.284	<0.0001	0.0002	274857	373963	163020
aspzin	seoE	latlat	0.32	20.642	<0.0001	0.211	<0.0001	<0.0001	299998	337959	174222
aspzin	seoW	latlat	0.348	23.301	<0.0001	0.238	<0.0001	<0.0001	275518	358740	173694
urs	aspzin	latlat	0.077	3.562	0.0004	0.039	<0.0001	<0.0001	262625	211774	181420
aspzin	seoE	monS	0.327	24.752	<0.0001	0.163	<0.0001	<0.0001	315462	325350	165028
aspzin	seoW	monS	0.358	28.945	<0.0001	0.185	<0.0001	<0.0001	290397	346402	163913
aspzin	seoE	monN	0.327	25.059	<0.0001	0.165	<0.0001	<0.0001	312909	324687	164512
aspzin	seoW	monN	0.358	29.404	<0.0001	0.188	<0.0001	<0.0001	287914	345641	163423
urs	seoE	aspzin	0.16	7.697	<0.0001	0.097	<0.0001	<0.0001	248267	253347	183463
urs	seoW	aspzin	0.11	6.254	<0.0001	0.065	<0.0001	<0.0001	251153	235940	189065
ber	seoE	lataru	0.404	79.096	<0.0001	0.184	<0.0001	0.0005	468788	265888	112856
ber	seoW	lataru	0.446	87.188	<0.0001	0.21	<0.0001	<0.0001	460410	283522	108548
urs	ber	lataru	0.112	28.187	<0.0001	0.038	<0.0001	0.0081	342919	156203	124748
monS	latgad	ber	0.198	41.672	<0.0001	0.033	<0.0001	<0.0001	755316	87851.4	58792.9
monN	latgad	ber	0.191	39.968	<0.0001	0.032	<0.0001	0.0009	754204	85822	58273.4
ber	seoE	latgad	0.411	85.718	<0.0001	0.23	<0.0001	<0.0001	454526	276102	115361
ber	seoW	latgad	0.451	93.985	<0.0001	0.261	<0.0001	<0.0001	446236	293634	111143
ber	seoE	latlat	0.387	78.584	<0.0001	0.2	<0.0001	0.004	458924	266179	117697
ber	seoW	latlat	0.427	84.381	<0.0001	0.228	<0.0001	<0.0001	450893	282908	113737
ber	seoE	monS	0.4	76.057	<0.0001	0.156	<0.0001	0.0012	479224	259054	111009
ber	seoW	monS	0.443	85.564	<0.0001	0.179	<0.0001	<0.0001	470845	276288	106700
urs	ber	monS	0.093	22.362	<0.0001	0.027	<0.0001	0.0044	348502	149941	124505
ber	seoE	monN	0.401	77.246	<0.0001	0.159	<0.0001	0.0011	476431	258584	110509
ber	seoW	monN	0.444	86.228	<0.0001	0.181	<0.0001	<0.0001	468094	275741	106245
urs	ber	monN	0.093	22.399	<0.0001	0.027	<0.0001	0.0028	346143	149462	124134
seoW	ber	urs	0.23	47.053	<0.0001	0.152	<0.0001	0.0009	345398	193518	121182
seoE	seoW	lataru	0.08	17.316	<0.0001	0.032	<0.0001	<0.0001	751170	151809	129214
urs	seoE	lataru	0.42	74.83	<0.0001	0.215	<0.0001	<0.0001	279812	311067	127195
urs	seoW	lataru	0.455	83.03	<0.0001	0.24	<0.0001	<0.0001	273059	328718	123155
latlat	latgad	seoW	0.076	17.983	<0.0001	0.026	<0.0001	0.0011	618020	111763	95878.6
monS	latgad	urs	0.109	38.052	<0.0001	0.021	<0.0001	0.0014	769813	67797.3	54459.8
seoE	seoW	latgad	0.077	17.721	<0.0001	0.04	<0.0001	<0.0001	730254	155731	133335

urs	seoE	latgad	0.442	82.269	<0.0001	0.274	<0.0001	<0.0001	271949	327377	126664
urs	seoW	latgad	0.475	90.582	<0.0001	0.303	<0.0001	<0.0001	265292	344937	122723
monS	latlat	seoE	0.145	25.832	<0.0001	0.034	<0.0001	0.0075	628460	111099	83029.5
monN	latlat	seoE	0.136	24.965	<0.0001	0.032	<0.0001	0.003	627109	108800	82673.3
seoE	seoW	latlat	0.075	16.509	<0.0001	0.035	<0.0001	<0.0001	742208	151801	130491
urs	seoE	latlat	0.421	78.826	<0.0001	0.242	<0.0001	0.0006	274792	316048	128783
urs	seoW	latlat	0.454	86.331	<0.0001	0.268	<0.0001	<0.0001	268349	332769	125055
monS	monN	seoW	0.028	7.311	<0.0001	0.003	<0.0001	0.0062	769088	32012.8	30245.8
seoE	seoW	monS	0.081	17.23	<0.0001	0.027	<0.0001	<0.0001	766143	148539	126372
urs	seoE	monS	0.404	66.375	<0.0001	0.179	<0.0001	0.0001	286092	300293	127474
urs	seoW	monS	0.44	77.228	<0.0001	0.201	<0.0001	<0.0001	279361	317539	123460
seoE	seoW	monN	0.08	17.264	<0.0001	0.027	<0.0001	<0.0001	762522	148121	126077
urs	seoE	monN	0.405	67.442	<0.0001	0.182	<0.0001	<0.0001	283696	299641	126964
urs	seoW	monN	0.441	77.969	<0.0001	0.204	<0.0001	<0.0001	276990	316808	122972

Table S6. *f*-branch statistics for the introgression analysis with 86 *Vipera* samples and sensitive clustering threshold. *f*-branch values for trios whose Dstatistic *p*-value > 0.001, sensitive clustering threshold *p*-value > 0.001 and/or Z-score |Z| < 3 have been considered equal to 0. Lineage codes are the same as in Table S3. Dashes indicate comparisons that cannot be made. Fig. 4A shows the placement of the branches (bX) in the species tree.

branch	branch_descendants	Outgroup	latlat	latgad	lataru	monN	monS	amm	aspzin	aspasp	aspfra	asphug	urs	berus	seoW	seoE
b2	amm, asp, urs, berus, seo	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
b3	mon, lat	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
b4	amm, asp	-	0	0	0	0	0	-	-	-	-	-	-	-	-	-
b5	urs, berus, seo	-	0.042	0.05	0.039	0.031	0.03	-	-	-	-	-	-	-	-	-
b6	lat	-	-	-	-	-	-	0	0.008	0.003	0.001	0.001	0.008	0.013	0.027	0.021
b7	mon	-	-	-	-	-	-	0	0	0	0	0	0	0	0	0
b8	amm	-	0	0	0	0	0	-	-	-	-	-	0	0	0	0
b9	asp	-	0.007	0	0	0	0	-	-	-	-	-	0.04	0.05	0.039	0.042
b10	urs	-	0	0	0	0	0	0.085	0	0	0.028	0.026	-	-	-	-
b11	berus, seo	-	0.052	0.057	0.038	0.027	0.027	0	0.065	0	0	0	-	-	-	-
b12	latgad, lataru	-	-	-	-	0.075	0.07	0	0	0	0	0	0	0	0	0
b13	latlat	-	-	-	-	0	0	0.002	0.017	0.011	0.007	0.006	0.005	0.009	0.007	0.006
b14	monS	-	0	0	0	-	-	0	0	0	0	0	0	0	0	0
b15	monN	-	0.02	0.073	0.023	-	-	0.003	0.002	0	0	0.002	0	0	0.003	0.002
b16	aspasp, aspzin	-	0.01	0	0	0	0	0	-	-	-	-	0.026	0.051	0.035	0.037
b17	aspfra, asphug	-	0	0	0	0	0	0.049	-	-	-	-	0	0	0	0
b18	seo	-	0.2	0.23	0.184	0.159	0.156	0	0	0	0	0	0	-	-	-
b19	berus	-	0	0	0	0	0	0.075	0.012	0.063	0.051	0.047	0.15	-	-	-
b20	latgad	-	0	-	-	0	0	0	0	0	0.004	0.004	0.012	0.02	0.04	0
b21	lataru	-	0.059	-	-	0.148	0.138	0	0	0	0	0	0	0	0	0
b22	aspasp	-	0	0	0	0	0	0.047	-	-	0.162	0.142	0	0	0	0
b23	aspzin	-	0.07	0.07	0.058	0.05	0.049	0	-	-	0	0	0.008	0.04	0.099	0.09
b24	aspfra	-	0	0	0	0	0	0	0.03	0.035	-	-	0.005	0.005	0.003	0.003
b25	asphug	-	0	0	0	0	0	0	0	0	-	-	0	0	0	0
b26	seoW	-	0.035	0.04	0.032	0.027	0.027	0	0	0	0	0	0	0	-	-
b27	seoE	-	0	0	0	0	0	0.009	0.034	0.023	0.016	0.015	0.006	0.008	-	-

Table S7. *f*-branch statistics for the introgression analysis with 86 *Vipera* samples and robust clustering threshold. *f*-branch values are computed only with trios whose Dstatistic *p*-value < 0.001, robust clustering threshold *p*-value < 0.001 and Z-score |Z| > 3. Lineage codes are the same as in Table S3. Dashes indicate comparisons that cannot be made. Fig. 4A shows the placement of the branches (bX) in the species tree.

[illegible]

Table S8. Niche overlap analyses between *V. latastei* and *V. seoanei*, including their Western and Eastern genomic lineages. Niche overlap is measured by the Sørensen index (K) and the Overlap Index (OI). As well, the niche hypervolume of each species' niche of the comparison is shown, the intersection and union among them, and the unique components of each compared species (Uniq. spX relative to spY). Vlat stands for *V. latastei*, Vseo for *V. seoanei*, W or E for Western or Eastern lineage of both species. Some comparisons (pure) excluded admixed populations from other species.

sp1	sp2	Vol. sp1	Vol. sp2	Intersection	Union	Uniq. sp1 rel. to sp2	Uniq. sp2 rel. to sp1	K	<i>p</i>	K _{max}	OI	<i>p</i>
Vlat (all)	Vseo (all)	2435322.1	1336650.6	410952.5	3361020.2	2024369.5	925698.1	0.215	0.01	0.696	0.309	0.01
Vlat (pure)	Vseo (pure)	2080461.9	1267075.6	235355.1	3112182.4	1845106.8	1031720.5	0.146	0.01	0.747	0.196	0.01
VlatE	VlatW	656106.52	1868500.26	78470.62	2446136.16	577635.89	1790029.64	0.062	0.01	0.521	0.119	0.01
VseoW	VseoE	1225437	322170.2	204707.1	1342900	1020729.9	117463.1	0.255	0.01	0.413	0.617	0.01
VlatE	VseoE	657782.28	321384.46	34745.53	944421.21	623036.76	286638.93	0.072	0.01	0.661	0.109	0.01
VlatW (all)	VseoW (all)	1893085.7	1233109.9	388438.7	2737757	1504647	844671.2	0.252	0.01	0.791	0.318	0.01
VlatW (pure)	Vseo W (pure)	1468990.1	1178598.9	107860.7	2539728.3	1361129.4	1070738.2	0.083	0.01	0.89	0.093	0.01

Table S9. Factor loadings of ecological PCA with 19 WorldClim v2.1 (Fick & R.J. Hijmans, 2017) variables for the three main Principal Components (88.04% of explained variance). Main factor loadings in bold.

Variable	Meaning	PC1	PC2	PC3
bio 1	Annual Mean Temperature	0.697	0.595	-0.305
bio 2	Mean Diurnal Range	0.723	-0.443	0.264
bio 3	Isothermality (BIO2/BIO7)	0.171	0.524	-0.255
bio 4	Temperature Seasonality	0.535	-0.730	0.355
bio 5	Max Temperature of Warmest Month	0.946	-0.034	0.094
bio 6	Min Temperature of Coldest Month	0.243	0.872	-0.394
bio 7	Temperature Annual Range (BIO5-BIO6)	0.650	-0.630	0.352
bio 8	Mean Temperature of Wettest Quarter	0.442	0.068	-0.736
bio 9	Mean Temperature of Driest Quarter	0.536	0.434	0.207
bio 10	Mean Temperature of Warmest Quarter	0.901	0.242	-0.126
bio 11	Mean Temperature of Coldest Quarter	0.438	0.791	-0.392
bio 12	Annual Precipitation	-0.853	0.417	0.247
bio 13	Precipitation of Wettest Month	-0.681	0.627	0.296
bio 14	Precipitation of Driest Month	-0.896	-0.325	-0.147
bio 15	Precipitation Seasonality	0.462	0.787	0.303
bio 16	Precipitation of Wettest Quarter	-0.683	0.627	0.317
bio 17	Precipitation of Driest Quarter	-0.924	-0.258	-0.108
bio 18	Precipitation of Warmest Quarter	-0.919	-0.235	-0.163
bio 19	Precipitation of Coldest Quarter	-0.632	0.659	0.376

Table S10. Toxin-encoding genes identified in *V. latastei* genome rVipLat1. Genes are ordered by chromosome and sorted out in toxin families. The transcript name is the specific isoform whose CDS has been used in selection analyses. Peptide hits are the number of unique peptide sequences from the *V. latastei* proteome which have been found in that isoform. Upreg. Column shows whether that gene has been found to be significantly upregulated in the venom gland. Microsynteny of the primary and secondary toxin families is displayed in Fig. S33. Three non-venomous genes related to toxin-encoding counterparts are shown at the bottom.

Family	Gene	Transcript	Peptide hits	Upreg.	Chrom	Strand	Description
CRISP	g_96919	long_reads1.PB.4209.307	10	YES	1	+	Cysteine Rich Venom Protein-1
5'nucleotidase	g_97788	long_reads1.PB.5295.83	1		1	+	5'-nucleotidase
PDE	g_98211	long_reads1.PB.5717.6	3		1	+	Venom phosphodiesterase
LAAO	g_48392	long_reads1.PB.13771.87	20		2	-	L-amino acid oxidase LAAO-II
LAAO	g_48393	anno4.cro_tig_rna-XM_039361015.1_R1	3		2	-	L-amino acid oxidase
Kunitz	g_162	long_reads1.PB.14656.7	0	YES	3	-	Serine protease inhibitor 3
Kunitz	g_5167	anno4.cro_tig_rna-XM_039342386.1_R0	1		3	+	Kunitz peptide
Kunitz	g_14506	anno4.pan_gut_rna-XM_034405137.1_R7	7		3	+	Kunitz-type serine protease inhibitor
Kunitz	g_14507	anno4.pan_gut_rna-XM_034405137.1_R5	6		3	+	Kunitz-type serine protease inhibitor
Kunitz	g_14509	anno4.pan_gut_rna-XM_034405137.1_R4	7	YES	3	+	Kunitz-type serine protease inhibitor
vNGF	g_1242	long_reads1.PB.15420.184	1	YES	3	+	Venom nerve growth factor
vNGF	g_14670	anno4.cro_tig_rna-XM_039356926.1_R1	1		3	+	Venom nerve growth factor
svMPi	g_82078	long_reads1.PB.21558.84	1	YES	5	-	MPi-3
svMPi	g_82081	long_reads1.PB.21557.998	1	YES	5	+	endogenous tripeptide metalloproteinase inhibitor precursor
CTL	g_71781	long_reads1.PB.23578.6	0	YES	6	+	Ancestral single-chain lectin toxin

CTL	g_71812	long_reads1.PB.23585.6	5	YES	6	+	Snaclec subunit A and B. Transcript example of subunit B: long_reads1.PB.23585.231
CTL	g_77666	long_reads1.PB.23587.74	1	YES	6	-	Snaclec subunit B
CTL	g_71813	long_reads1.PB.23587.53	6	YES	6	-	Snaclec subunit A and B. Transcript example of subunit B: long_reads1.PB.23587.253
CTL	g_71814	long_reads1.PB.23587.423	3	YES	6	-	Snaclec subunit A
CTL	g_71817	long_reads1.PB.23585.597	0	YES	6	+	Snaclec subunit B
PLB	g_72774	long_reads1.PB.24340.6	1		6	+	Phospholipase B
svMP	g_134876	long_reads1.PB.27234.89	13	YES	8	+	Metalloproteinase MPIII
svSP	g_130459	long_reads1.PB.28174.636	11	YES	9	-	Serine protease
svVEGF	g_131055	anno1.g48961.t1	7	YES	9	+	vammin-1'
XaaPro	g_38002	long_reads1.PB.30060.88	6		12	+	xaa-Pro aminopeptidase 2
CYS	g_79330	anno4.pan_gut_rna-XM_034427211.1_R7	1		16	+	Cystatin 1
PLA	g_133624	long_reads1.PB.32683.3	13	YES	17	-	Ammodytin I2(A)
PLA	g_133627	long_reads1.PB.32686.117	9	YES	17	+	Ammodytin I1(D)
LAAO-like	g_48408	anno4.cro_tig_rna-XM_039361019.1_R0	2			-	L-amino-acid oxidase-like. Non venomous
LAAO-like	g_50929	anno2.g17375.t1		2		+	L-amino acid oxidase-like. Non venomous
PLA_gIIE	g_134145	anno4.pan_gut_rna-XM_034424295.1_R9	17			+	Ancestral non-venomous PLA ₂ . Non venomous

Table S11. Toxin-encoding genes identified in *V. aspis* genome rVipAsp1. Genes are ordered by chromosome and sorted out in toxin families. The transcript name is the specific isoform whose CDS has been used in selection analyses. Peptide hits are the number of unique peptide sequences from the *V. aspis* proteome which have been found in that isoform. Microsynteny of the primary and secondary toxin families is displayed in Fig. S33. Four non-venomous genes related to toxin-encoding counterparts are shown at the bottom.

Family	Gene	Transcript	Peptide hits	Chrom.	Strand	Description
CRISP	g_33918	long_reads1.PB.695.69	12	1	-	Cysteine-rich venom protein
CRISP	g_28182	long_reads1.PB.695.4	14	1	-	Cysteine-rich venom protein
5'nucleotidase	g_28618	long_reads1.PB.1209.71	22	1	+	5'-nucleotidase
5'nucleotidase	g_28619	long_reads1.PB.1209.244	14	1	+	5'-nucleotidase
PDE	g_29051	long_reads1.PB.1629.1	15	1	+	Phosphodiesterase
vQC	g_29579	long_reads1.PB.2193.13	11	1	-	Glutaminyl-peptide cyclotransferase
svVEGF	g_29661	long_reads1.PB.2291.57	4	1	+	Vascular endothelial growth factor A
LAAO	g_6353	long_reads1.PB.7791.79	33	2	-	L-amino acid oxidase
LAAO	g_3579	anno4.cro_tig_rna-XM_039361015.1_R2	2	2	-	L-amino acid oxidase
LAAO	g_6356	anno4.cro_tig_rna-XM_039361013.1_R4	3	2	-	L-amino acid oxidase
LAAO	g_3580	anno4.cro_tig_rna-XM_039361019.1_R3	5	2	-	L-amino acid oxidase
PPi	g_13950	long_reads1.PB.15243.2	1	5	-	Peptidyl-prolyl cis-trans isomerase C
Kunitz	g_70866	anno4.cro_tig_rna-XM_039342380.1_R0	0	7	-	Serine protease inhibitor 3
Kunitz	g_74066	anno4.cro_tig_rna-XM_039342386.1_R1	0	7	+	Kunitz/BPTI-like toxin
Kunitz	g_74067	anno4.pan_gut_rna-XM_034405137.1_R4	4	7	+	Kunitz-type serine protease inhibitor
Kunitz	g_74068	anno4.pan_gut_rna-XM_034405137.1_R5	5	7	+	Kunitz-type serine protease inhibitor

Kunitz	g_74069	anno4.pan_gut_rna-XM_034405137.1_R6	5	7	+	Kunitz-type serine protease inhibitor
Kunitz	g_74071	long_reads1.PB.18914.159	3	7	+	Kunitz-type serine protease inhibitor
Kunitz	g_74072	anno4.pan_gut_rna-XM_034405137.1_R7	1	7	+	Kunitz-type serine protease inhibitor
vNGF	g_71982	long_reads1.PB.19627.61	13	7	+	Venom nerve growth factor
RLAP	g_72151	long_reads1.PB.19854.45	4	7	+	Renin-like aspartic protease
Prosaposin	g_56972	long_reads1.PB.21396.97	4	8	-	Prosaposin isoform X2
svMPi	g_58598	long_reads1.PB.22989.178	22	8	-	Endogenous tripeptide metalloproteinase inhibitor precursor
svMPi	g_59515	long_reads1.PB.22987.159	27	8	+	Endogenous tripeptide metalloproteinase inhibitor precursor
CTL	g_78881	long_reads1.PB.23130.9	7	9	+	Ancestral single-chain lectin toxin
CTL	g_78909	long_reads1.PB.23137.139	22	9	+	Snaclec subunit A and B. transcript example of subunit B: long_reads1.PB.23137.121
CTL	g_78912	long_reads1.PB.23138.95	17	9	-	Snaclec subunit A
CTL	g_78913	long_reads1.PB.23138.61	2	9	-	Snaclec subunit A
CTL	g_78914	long_reads1.PB.23138.490	6	9	-	Snaclec subunit A
PLB	g_79927	long_reads1.PB.23926.4	9	9	+	Phospholipase B
hyaluronidase	g_80652	long_reads1.PB.24587.28	3	9	+	Hyaluronidase
aminopeptidase	g_40343	long_reads1.PB.25597.11	18	10	+	Glutamyl aminopeptidase
svMP	g_9891	long_reads1.PB.27853.3060	41	11	-	Metalloproteinase of class P-III
CP	g_67109	long_reads1.PB.28470.8	2	12	+	Cysteine proteinase 2-like
svSP	g_67321	long_reads1.PB.28631.537	21	12	-	Serine proteinase SP-8
svVEGF	g_67922	long_reads1.PB.28961.109	15	12	+	vammin-1'
DPP	g_84414	long_reads1.PB.29453.2	3	13	-	dipeptidase 2-like
XaaPro	g_75857	long_reads1.PB.29901.62	26	14	-	xaa-Pro aminopeptidase 2
PPi	g_69564	long_reads1.PB.30554.65	3	15	+	Peptidyl-prolyl cis-trans isomerase B

PLA	g_11106	long_reads1.PB.32751.122	28	20	-	ammodytin I1
neuroPLA	g_11113	anno4.cro_tig_rna-XM_039367476.1_R2	4	20	+	Vaspin basic subunit variant. Transcript example of acidic subunit: anno4.cro_tig_rna-XM_039367475.1_R0
CRISP-like	g_33919	anno4.cro_tig_rna-XM_039329609.1_R0	0	1	-	serotriflin. Non venomous.
LAAO-like	g_6374	anno4.naj_naj_rna-gnl_WGS:SOZL_mRNA16818_R5	0	2	-	L-amino acid oxidase-like. Non venomous.
LAAO-like	g_5192	anno2.g34869.t1	0	2	+	L-amino acid oxidase-like. Non venomous.
svMP-ancestral	g_9895	long_reads1.PB.27853.3260	1	11	-	ADAM28. Non venomous.

Table S12. Toxin-encoding genes identified in *V. seoanei* genome rVipSeo1. Genes are ordered by chromosome and sorted out in toxin families. The transcript name is the specific isoform whose CDS has been used in selection analyses. Peptide hits are the number of unique peptide sequences from the *V. seoanei* proteome which have been found in that isoform. Microsynteny of the primary and secondary toxin families is displayed in Fig. S33. Four non-venomous genes related to toxin-encoding counterparts are shown at the bottom.

Family	Gene	Transcript	Peptide hits	Chrom.	Strand	Description
CRISP	g_41326	long_reads1.PB.4352.28	17	1	+	Cysteine-rich venom protein
LAAO	g_877	long_reads1.PB.8947.374	30	2	+	L-amino acid oxidase LAAO-II
LAAO	g_875	anno1.g991.t1	4	2	+	L-amino acid oxidase
LAAO	g_7610	anno4.cro_tig_rna-XM_039361015.1_R1	6	2	+	L-amino acid oxidase
Kunitz	g_95072	anno4.pan_gut_rna-XM_034424188.1_R7	1	3	+	Kunitz-type serine protease inhibitor
Kunitz	g_114630	long_reads1.PB.14544.37	3	3	+	Kunitz-type serine protease inhibitor

vNGF	g_96318	long_reads1.PB.15297.63	8	3	+	Venom nerve growth factor 1
CTL	g_136710	long_reads1.PB.26260.83	3	6	+	Snaclec subunit A
CTL	g_136709	long_reads1.PB.26261.190	13	6	-	Snaclec subunit A and B. Transcript example of subunit B: long_reads1.PB.26261.539
CTL	g_136713	long_reads1.PB.26260.403	9	6	+	Snaclec subunit A and B. Transcript example of subunit B: long_reads1.PB.26260.348
CTL	g_136734	long_reads1.PB.26269.19	0	6	-	Ancestral single-chain lectin toxin
svMP	g_143356	long_reads1.PB.28146.617	22	8	+	Zinc Metalloproteinase of class P-III
svSP	g_26734	long_reads1.PB.29112.1266	23	9	-	Snake Venom Serine Protease
svVEGF	g_27370	long_reads1.PB.29435.74	24	9	+	Vammin-1
PLA	g_117442	long_reads1.PB.33347.218	24	17	-	Ammodytin I1
PLA	g_117449	long_reads1.PB.33345.60	21	17	+	Ammodytin I2
CRISP-like	g_66533	anno4.cro_tig_rna-XM_039329609.1_R0	0	1	+	Serotriflin. Non venomous.
LAAO-like	g_10173	anno4.naj_naj_rna-gnl_WGS:SOZL_mRNA16818_R4	0	2	+	L-amino-acid oxidase-like. Non venomous.
svMP-ancestral	g_145534	anno4.cro_tig_rna-XM_039354198.1_R0	1	8	+	ADAM28. Non venomous.
PLA_gIIE	g_117452	long_reads1.PB.33346.10	0	17	+	Phospholipase A2-like group IIE. Non venomous.

Table S13. Selection tests on toxin-encoding genes per family. d_N/d_S tests were performed with BUSTED and FEL implemented on Datamonkey. BUSTED looks for codons with episodic diversifying selection (significant values in bold), whereas FEL tests for codons with either pervasive diversifying or purifying selection, with a fixed p -value in 0.1. Only toxin families with at least 3 copies in at least 2 species (spp.) were tested.

Family	spp.	copies	codons	BUSTED		FEL		
				p -value	codons	p -value	codons diversifying sel.	codons purifying sel.
SVMP	3	3	614	0.018	4	0.1	0	27
PLA ₂	3	6	143	<0.001	20	0.1	3	6
snaclec	3	12	162	<0.001	19	0.1	15	19
CRISP	3	4	260	<0.001	51	0.1	1	3
LAAO	3	9	549	<0.001	56	0.1	8	33
SVVEGF	3	3	131	0.5	0	0.1	0	2
SVSP	3	3	259	0.0042	10	0.1	0	5
SVNGF	3	4	260	0.38	3	0.1	0	1
Kunitz	3	12	87	<0.001	11	0.1	2	8
SVMPi	2	4	253	<0.001	7	0.1	0	8
5'-nucleotidase	2	3	592	0.5	0	0.1	0	17

Table S14. Samples used for phylogenetic inference and the lineage to which they belong. All these samples were used to build the *snapp_subspecies* and *treemix_subspecies* datasets, as well as for extracting their mitogenomes to infer the mitochondrial phylogeny, but only the first seven samples were included in *snapp_species* dataset (see Table S11).

Sample	Species	Subspecies/lineage	<i>snapp_species</i>
ATP14	<i>V. ammodytes</i>	Northwestern	Y
CNMOL1	<i>V. aspis</i>	<i>zinnikeri</i>	Y
SPM002443	<i>V. berus</i>	<i>berus</i>	Y
CNPOR1	<i>V. latastei</i>	<i>latastei</i>	Y
FM19VM052	<i>V. monticola</i>	<i>monticola</i>	Y
CNSOC1	<i>V. seoanei</i>	East	Y
CN13667	<i>V. ursinii</i>	<i>rakosiensis</i>	Y
MALZ004	<i>V. aspis</i>	<i>hugyi</i>	N
FM22VL233	<i>V. latastei</i>	<i>gaditana</i>	N
EMO4112020	<i>V. aspis</i>	<i>francisciredi</i>	N
BEV11035	<i>V. aspis</i>	<i>aspis</i>	N
CN20038	<i>V. latastei</i>	<i>arundana</i>	N
MNCN50498	<i>V. monticola</i>	<i>saintgironsi</i>	N
FM19VM061	<i>V. monticola</i>	<i>atlantica</i>	N
SPM003076R	<i>V. seoanei</i>	West	N

Table S15. Partitions for the mitochondrial protein-coding gene alignment of 10,695 bp arranged by PartitionFinder. This partition scheme and its respective nucleotide alignment were used to infer the mitochondrial phylogeny of the studied *Vipera* spp., depicted in Figs. 2, S7.

Partition	Best Model	Sites	Genes and positions
1	TVM+I	676	ATP6_1 st , ND4L_1 st , ND4_1, ND3_1 st
2	GTR+I+G	1617	ND2_2 nd , ND4L_2 nd , ND4_2 nd , ND5_2 nd , ND3_2 nd , ATP6_2 nd
3	GTR+I+G	465	ATP6_3 rd , ND4_3 rd
4	TRN+I	110	ATP8_1 st , ATP8_2 nd
5	TIM+I+G	286	ATP8_3 rd , COX2_3 rd
6	TRNEF+G	534	COX1_1 st
7	K81UF+I	534	COX1_2 nd
8	GTR+I+G	909	ND3_3 rd , COX1_3 rd , COX3_3 rd
9	TRN+I	231	COX2_1 st
10	TIM+I	552	COX2_2 nd , ND1_2 nd
11	GTR+I+G	954	COX3_1 st , CTYB_1 st , ND1_1 st
12	TIM+I	633	CTYB_2 nd , COX3_2 nd
13	TIM+I+G	969	CTYB_3 rd , ND5_3 rd
14	TIM+I+G	762	ND1_3 rd , ND4L_3 rd , ND2_3 rd
15	GTR+G	941	ND2_1 st , ND5_1 st
16	TRN+I+G	174	ND6_1 st
17	HKY+G	174	ND6_2 nd
18	TIM+G	174	ND6_3 rd

Table S16. Cross Validation (CV) average and standard deviation (SD) values for Admixture analysis. A likelihood plateau is reached at K=3, although there are four species in the dataset. The highest likelihood is reached at K=5, although K=7 has the highest biological significance.

K	CV Average	CV SD
1	0.4371	0.0002
2	0.3231	0.0211
3	0.2484	0.0002
4	0.2453	0.0063
5	0.2439	0.0011
6	0.2463	0.0019
7	0.2599	0.0092
8	0.2623	0.0047
9	0.2709	0.0061
10	0.2813	0.0048
11	0.2948	0.0109
12	0.3038	0.0072

Supplementary Information. Material and Methods

Sampling

We extracted blood samples from four different individuals to produce the sequencing data for three reference genomes: a male (CNPOR1) *V. latastei* from Els Ports (Catalonia, Spain), a female and a male (CNMOL1) *V. aspis* specimens from La Molina (Catalonia, Spain) and a male (CNSOC1) *V. seoanei* from Socobio (Cantabria, Spain). Another four specimens were euthanized (a female *V. aspis* from La Molina, VA; a male *V. seoanei* from Burgos, VS; and two male juvenile and adult *V. latastei* from Burgos as well, VL and VL2, respectively) to collect their tissues, which were flash-frozen at -80°C until RNA extraction. Regarding WGS, samples from 94 individuals of different European and North African viper species of the genus *Vipera*, including *V. latastei*, *V. seoanei*, *V. monticola*, *V. aspis*, *V. berus*, *V. ursinii* and *V. ammodytes* were sequenced in the present study (See Table S1). We focused on the Western Mediterranean diversity, including representatives from all described lineages of *V. latastei*, *V. seoanei*, *V. aspis* and *V. monticola*, specially aiming to cover most of the Iberian Peninsula and the contact zones among species. Samples varied from tail tips, tissue from road killed animals, museum specimens or scale clips, and were stored in 99% ethanol at -20°C. Finally, venom samples were collected in 2018 from three *Vipera latastei* specimens in Segovia (adult female, VI2), Hervás, Cáceres (adult male, VI14), and Navezuelas, Cáceres (adult male, VI5).

DNA extraction, library preparation and sequencing

We extracted high molecular weight genomic DNA from diverse tissues (tail tips, ventral scale clips or blood, mainly) from wild and museum specimens of 94 individuals of the genus *Vipera*. We followed manufacturer's protocol of MagAttract HMW Kit (Qiagen), sonicated samples with QSonica Q800R3, built whole genome libraries following the BEST method¹ with minor adjustments, and quality-checked with Bioanalyzer (Agilent). We sequenced 150bp pair-end reads aiming for 10x coverage for 86 samples in Illumina NovaSeq6000 lanes from Macrogen and Novogene Co. Five samples were sequenced at high coverage (45x) as representatives of their species, and another three samples were sequenced at very high coverage (60-75x) for reference genome assembly purposes (see Table S1). Secondly, long-read Oxford Nanopore Technologies (ONT) data was sequenced from blood DNA extractions of male *V. latastei*, *V. seoanei* and *V. aspis* at CNAG, aiming for 60x coverage for each species. Additionally, 60x long-read data from PacBio HiFi was sequenced for *V. latastei*. Finally, we combined Bionano optical mapping (248 cmaps) and Hi-C data (100x) to scaffold *V. latastei*'s draft genome into pseudochromosomes. On the other hand, for *V. aspis* we built Omni-C libraries from another female conspecific's blood using Omni-C kit by Dovetail Genomics and sequenced 150bp PE reads with Illumina NovaSeq6000 (CNAG) aiming for 120x coverage.

To annotate the reference genomes and to explore differential gene expression among organs, four individuals (young and adult *V. latastei*, *V. aspis* and *V. seoanei*) were sacrificed and their individualized tissues (heart, lungs, brain, venom gland, salivary gland, tongue, liver, eye, gall bladder, kidney, stomach, pancreas, and gonads) were flash frozen with liquid nitrogen. RNA was extracted with HigherPurityTM Tissue Total RNA Purification Kit (Canvax) and quality-checked with Bioanalyzer. The 19 best-preserved RNA samples (including 4 venom glands) were used for building stranded RNA libraries with poly-A enrichment. 12 Gb of data per sample were sequenced with Illumina short reads in NovaSeq PE150 (Novogene Co.). Furthermore, for each species, one pooled non-venomous RNA sample and one venom gland sample (pooled for adult and young *V. latastei*) were sequenced with PacBio long reads (Iso-Seq) to obtain HiFi full-length transcriptomes for a better isoform annotation. In total, six samples of this project and two external ones were sequenced on a PacBio HiFi SMRT Cell 8M using the Sequel II sequencing kit (Novogene Co.).

Genome assemblies

Vipera latastei's PacBio HiFi reads were trimmed from PacBio adapters using Cutadapt v3.4². ONT long reads for the three species were quality checked with nanoQC³. To obtain the draft genome assemblies, *V. latastei* PacBio reads were assembled using Hifiasm v0.13⁴ with default settings, and *V. seoanei* and *V. aspis* long read data above 8 kbp was assembled using NextDenovo v2.5.2⁵. Illumina short read data was quality checked with FastQC v0.11.9⁶ and multiQC⁷, adapters were trimmed with fastp v0.20⁸ and mapped to the respective draft assembly with BWA v0.7.17-r1188 mem algorithm⁹, and used to polish *V. seoanei* and *V. aspis* drafts with two rounds of Pilon v1.24¹⁰. Afterwards, haplotigs were purged from the three draft assemblies with *purge_dups* v1.2.5¹¹. A first round of scaffolding in *V. latastei* was performed using Bionano data with Bionano Solve v3.6.1 (Bionano Genomics), and a second round using SALSA2 v2.2¹² with parameters “-e GATC,GATC -m yes”. A final gap filling step was performed using RagTag v2.0.1¹³ and another draft assembly of *V. latastei* from ONT data. This ONT raw data was filtered with Q>7 and minimum length of one kbp, assembled with Flye v2.8.3¹⁴ with two polishing iterations, polished again with HyPo v2.4.1¹⁵ and scaffolded as the PacBio draft assembly, first with Bionano Solve and then with SALSA2. *Vipera aspis* Omni-C short read data was again quality-checked, and adapters were removed as described above. Three iterations of SALSA2 were implemented to scaffold and correct previous misassemblies. Finally, manual curation was performed on PretextView¹⁶, with mapping threshold Q=10, to visualize the contact map, aided by coverage, gap, telomere, and repetitive regions graphs, following the rapid curation manual (<https://gitlab.com/wtsi-grit/rapid-curation>). The gap bed was obtained through the fasta-stats.py script (available in https://github.com/nextgenusfs/genome_scripts/tree/master). Telomere regions were identified using the conserved vertebrate telomeric sequence¹⁷. We soft-masked repetitive areas in the genomes with RED¹⁸ using the redmask.py wrapper (<https://github.com/nextgenusfs/redmask/blob/master/redmask.py>). On the other hand, for scaffolding *V. seoanei*'s genome, we used the assembly of the closely related *V. ursinii* (rVipUrs1.1, PRJEB57183), with the same chromosome number¹⁹, as reference input for RagTag. Finally, potential contamination in the assemblies was assessed with BlobToolsKit²⁰. Two small scaffold/contigs were removed in the case of *V. seoanei* for extremely high coverage and the detection of mitochondrial genes, resulting in all genomic information contained in 18 scaffolds, matching the expected number of chromosomes²¹. Genome assemblies were depicted as snail plots with BlobToolsKit (Fig. S1-S3), and their contiguity and completeness were assessed with gfastats²² and BUSCO (Vertebrate databases²³), respectively.

Genome annotation

We annotated our three reference genomes by combining four annotation pipelines: 1) First we used GeMoMa v1.9²⁴, which is based on the conservation of intron positions in related organisms. As input for GeMoMa, we downloaded the annotated reference genomes of *Crotalus tigris* (ASM1654583v1; NCBI BioSample ID: SAMN12497667; Margres et al., 2021), *C. adamanteus* (Cadam_11369; NCBI BioSample ID: SAMN16378231; Hogan et al., 2021), *Naja naja* (Nana_v5; NCBI BioSample ID: SAMN11155092), *Ophiophagus hannah* (OphHan1.0; NCBI BioSample ID: SAMN02439592; Vonk et al., 2013) and *Pantherophis guttatus* (UNIGE_PanGut_3.0; NCBI BioSample ID: SAMN12106767; Ullate-Agote et al., 2014). 2) Based on *de novo* sequenced full-length transcriptomes of different tissues. Subreads from Iso-Seq PacBio were first converted into circular consensus sequences (CCS) with the ccs function from pbccs v6.4 (Pacific Biosciences of California, Inc), followed by demultiplexing and primer removal with lima v2.7.1 (Pacific Biosciences of California, Inc). Non-chimeric full-length (NCFL) transcripts were then obtained by removing concatemers with isoseq3 v3.8.3 (Pacific Biosciences of California, Inc) and trimming poly-A tails with the python script *tama_flnc_polya_cleanup.py* (<https://github.com/GenomeRIK/tama/wiki>) mapped to each reference genome assembly with the splice-aware mapper pbmm2 v1.10 (Pacific Biosciences of California, Inc). The resulting mapped reads were collapsed into unique isoforms with isoseq3 v3.8.3 with the --do-not-collapse-extra-5exons flag. Collapsed unique isoforms were then used to

identify protein-coding regions with GeneMarkS-T²⁹. We employed transcripts from all the six libraries to annotated each of the three genomes to account for the maximum number of putative genes, even if they are not expressed in a particular species. 3) Based on RNA-Seq evidence from different tissues within each species, following BRAKER1 pipeline³⁰. We quality-checked RNA-Seq Illumina reads with fastQC, and trimmed poly-A and poly-G tails, remove adapters and low-quality bases with fastp. Filtered reads were mapped to the respective reference using the splice-aware mapper STAR v2.7.10b³¹ and then used as input for BRAKER1, which creates gene sets with GeneMark-ET to train Augustus. 4) Based on external protein databases following BRAKER2 pipeline³². We ran BRAKER2 twice per species, using the ortholog database Sauropsida OrthoDB v10³³ and all Sauropsida toxin isoforms from the Animal Toxin Annotation Project (<https://www.uniprot.org/help/Toxins>) as input to creates gene sets with GeneMark-EP+ to train Augustus. Finally, we integrated the different annotation files with TSEBRA³⁴, keeping all predictions from GeMoMa and GeneMarkS-T. BUSCO scores and general metrics from the resulting annotation files for each of the three species were calculated (Table S4).

Macrosynteny analysis

Chromosomal rearrangements were explored with MCscan³⁵ in between the three *Vipera* assemblies: *V. latastei*, *V. aspis* and *V. seoanei*, the latter representing the general chromosomal organization of the *berus* group, as its scaffolding was done using the closely related *V. ursinii* as reference. As outgroup we used the reference genome of *Crotalus viridis* (UTA_CroVir_3.0; NCBI BioSample ID: SAMN07738522; Pasquesi et al., 2018) to represent the sister subfamily Crotalinae. Protein sequences from each viperid species were extracted using AGAT v0.8.0³⁷ and pairwise aligned using LAST³⁸ with the JCVI python module³⁹. Initial alignments between species were used to identify the Z chromosomes as well as chromosomes assembled in the reverse complement, which were corrected in some of the assemblies by reverse complementing the fasta sequence with SAMtools faidx v0.7.2.2⁴⁰. Afterwards, GeMoMa was used to generate a draft annotation of the new genomes with reversed sequences, using as reference the original reference genome of each respective species. Then, we re-ran the MCscan pipeline to generate final re-oriented synteny plots.

Processing of whole genome sequencing

Raw Illumina data of all WGS samples (n=94) was quality-checked with fastQC and processed with fastp, filtering out poly-X and poly-A tails and bases with Q<30. Filtered reads were mapped to the reference genome of *V. latastei* (rVipLat1) with BWA, and then filtered by mapping quality with SAMtools (-q 30 -F 260). PCR duplicates were marked and removed with PicardTools v3.0.0 (<https://broadinstitute.github.io/picard/>). The individual read mean coverage was determined with SAMtools depth function. At this point, data processing was split into a low- and a high-coverage dataset. On the one hand, for the former one, samples with coverages higher than 10x were downsampled with SAMtools to 7x, which was the mean coverage among all the low-coverage samples. On the other one, high-coverage samples were kept at their original depth or downsampled to 45x in the case of the Illumina resequencing data from the three reference genomes to be part of the high-coverage dataset. Afterwards, we used GATK v4.1.3⁴¹ to proceed with the SNP-calling in both datasets, through the functions HaplotypeCaller, CombineGVCFs and GenotypeGVCFs. The resulting datasets of genotype callings underwent then a stringent filtering pipeline. Density plots of six highly informative annotations from GATK tools were visualized to set optimized thresholds for each dataset: i.e. Quality by Depth (QD), Fisher Strand bias (FS), Strand Odds Ratio (SOR), the root mean square Mapping Quality over all the reads at the site (MQ), the u-based z-approximation from the Rank Sum test for mapping qualities (MQRankSum) and the u-based z-approximation from the Rank Sum test for site position within reads (ReadPosRankSum). For the high-coverage dataset, calls were kept only if QD > 7, FS < 70, SOR < 4, MQ > 55, -1.5 < MQRankSum < 1.5, -5 < ReadPosRankSum < 5, whereas for the low-coverage dataset: QD > 5, FS < 50, SOR < 4, MQ > 55, -1.5 < MQRankSum < 1.5, -3 < ReadPosRankSum < 3. Indels were normalized and SNPs within 10 bp to any indel were as well

discarded. Furthermore, we excluded SNPs within repetitive regions previously inferred with RED, and sites which mapping was prone to be ambiguous following the SNPable pipeline (<http://lh3lh3.users.sourceforge.net/snpable.shtml>). These regions were inferred by splitting the reference genome of *V. latastei* into overlapping fragments of 100 bp (overlapping by 99 bp), mapping those regions back to the reference, and estimating the number of fragments that correctly mapped at each site. For those sites with a mapping rate below 95%, calls were discarded. Several datasets were split from both the high- and low-coverage branches of the data processing into more specific datasets to accommodate different purposes before further filtering (Table S2).

Genomic and mitochondrial phylogenies

Genomic species tree inferences of the studied European *Vipera* were performed using SNAPP⁴² implemented on BEAST v2.6.4⁴³. We produced two different datasets of low-coverage autosomal bi-allelic unlinked SNPs and 0% missing data with VCFtools⁴⁴. The first one, a species dataset (i.e., *snapp_species* dataset, Table S2, samples in Table S14) comprising one representative of *V. latastei*, *V. monticola*, *V. aspis*, *V. ammodytes*, *V. berus*, *V. seoanei*, *V. ursinii* and *Crotalus viridis* as outgroup (Illumina raw data GenBank accession number SRR10593867; Schield et al., 2019). For the second dataset, we included representatives from all putative subspecies (i.e., *snapp_subspecies* dataset, see Tables S3,S13) from the above-mentioned species plus another one from the Western lineage of *V. seoanei* (see Admixture results). We built two SNAPP time-calibrated species trees following Stange et al. ⁴⁶. The species dataset was LD-pruned by 0.5-kbp windows, yielding 1,225,190 uSNPs, whereas the subspecies dataset by 10-kbp windows to reduced computational demands, yielding 126,083 uSNPs (see “Population genomics” section below for LD-pruning thresholds). The vcf files of both datasets were converted to xml format by using the script *snapp_prep.rb* (available at https://github.com/mmatschiner/snapp_prep). First, we time-calibrated the species phylogeny with the divergence between Viperinae and Crotalinae, set with a lognormal distribution in 38.61 ± 4.85 mya following a fossil-calibrated phylogeny⁴⁷. For the subspecies phylogeny without *Crotalus*, we opted for dating the diversification of the crown *berus* group, which appears to be the only one whose topology is consistent across the literature and with our own results at species level (see SNAPP results, Fig. S6), with a lognormal distribution in 5.97 ± 0.75 mya⁴⁷. We carried out three different runs of each of the two SNAPP analyses for 20 million iterations each, sampling every 5,000 generations. Stationarity (ESS>200) and convergence between runs from the same dataset was checked with Tracer v1.7.2⁴⁸. Converging log and tree files were combined with logCombiner v1.10 discarding 10% as burn-in. The cloudograms were visualized with DensiTree v2.6.4⁴⁹, and TreeAnnotator v2.6.3 was used to construct the maximum clade credibility tree and annotate nodes with their corresponding posterior probabilities.

On the other hand, we produced a mitochondrial phylogeny to account for potential mito-nuclear discordance, using the whole mitogenomes of the same samples as the *snapp_subspecies* dataset (Table S14). From the adapter-trimmed Illumina reads of those samples, we extracted and assembled 15 mitogenomes with GetOrganelle v1.7.7.0⁵⁰, using the published mitochondrial assembly of *V. berus* (GenBank accession number: MF945570.1⁵¹) as reference and k-values: 21, 55, 85, 115. Additionally, we downloaded the mitogenome of *Crotalus horridus* (GenBank accession number: HM641837.1) as outgroup. The 16 mitogenome assemblies were annotated with MitoFinder⁵² using *C. adamanteus* as reference (GenBank accession number: NC_041524.1). We aligned the thirteen protein-coding genes present in the mitogenomes with MEGA11⁵³, using the MUSCLE algorithm⁵⁴ and manually inspected for stop codons. We used only a 714-bp fragment of the gen ND4 to avoid stop codons. Single-gen alignments were concatenated with Geneious Prime v2023.2.1, yielding a 10,695-bp alignment. The best partition scheme was determined with PartitionFinder v2.1.1⁵⁵ using the *greedy* algorithm, constraining results to “beast” models, specifying AICc for model selection, linked branch lengths and potential partitions for each gen and each codon position within a gen in the alignment. It arranged the data into 18 partitions (Table S15). The model for partition 7, K81UF+I, was replaced by

GTR+I instead in BEAUTi. We linked a relaxed clock model with frequencies sampled from a lognormal distribution to all partitions, but a relative substitution rate was estimated for each partition and trees among partitions were linked. Similar to the genomic phylogeny built on SNAPP, we calibrated the mitochondrial tree both with the Viperinae-Crotalinae split and the diversification of the *berus* group, with the same dates stated above. Applying such partition scheme and priors, we ran BEAST v2.7.4 three independent times for 50 million iterations each, sampling every 1,000 generations. We checked for the stationarity and convergence of the three runs in Tracer and used TreeAnnotator to construct the maximum clade credibility tree. The resulting mitochondrial phylogeny was overlapped to visualize mito-nuclear discordance with the subspecies cloudogram of SNAPP.

Introgression analyses

On the grounds of exploring introgression and geneflow events among the *Vipera* lineages, while considering ILS, we calculated ABBA-BABA tests, and their related estimate of admixture fraction, f_4 -ratio, with the function Dtrios implemented in Dsuite v0.5 r53⁵⁶. For that purpose, we filtered out from the low-coverage dataset individual genotypes with a read depth below four or above 150, or a quality score below 30, non-biallelic alleles, SNPs with more than 10% missingness, singletons (adjusting a different minimum allele frequency threshold per dataset) and invariant sites and only kept SNPs from the autosomes. As an outgroup was needed to test the maximum trio combinations, we included *Crotalus viridis* in the analysis. On the other hand, from this dataset we excluded *V. latastei* \times *aspis* putative hybrids, as well as individuals resulting from secondary contacts between intraspecific lineages, as depicted by Admixture (see Admixture results), yielding finally 86 individuals and 12,645,411 SNPs (*linked SNPs* dataset, Table S2). We assumed the genomic species tree inferred by SNAPP to be the true phylogeny, and used all the taxa and lineages supported by that tree but *V. monticola* spp., which were split in North and South lineages instead due to incongruent relationships among subspecies (see Results). Dtrios was ran with the “--ABBAclustering” option to avoid false signals of introgression due to substitution rate variation among branches, which can lead to “ABBA” homoplasies⁵⁷. Then, for a better interpretation of the resulting f_4 -ratio values, and to correctly take correlated excess allele-sharing into account, we calculated f -branch statistics⁵⁸ implemented as well in Dsuite. Per default, all f_4 -ratio values are set to zero when the p -value of the associated D statistic for that trio is > 0.01 ⁵⁶, but we chose to equally purge out Z scores $|Z| < 3$ and, moreover, the ABBA-clustering tests’ p -values > 0.01 (Table S5). These tests calculate two sets of p -values for each trio with different sensitivity and robustness (Tables S6-S7). We plotted the results following both types of thresholds (sensitive and robust) with the Dsuite script dtools.py. Finally, to investigate the direction of the detected introgression events, we ran TreeMix v1.13, which uses SNP frequency data to evaluate population splitting, drift and directional introgression⁵⁹. We used a dataset of 15 individuals representing all subspecies or lineages in the studied *Vipera* spp. with 0.5-kb thinning and 0% missingness, yielding a total of 1,095,590 bi-allelic autosomal uSNPs (*treemix_subspecies* dataset, Tables S2-S3, S13). We analyzed the data in blocks of 20, 100 and 500 SNPs to account for non-independence of the data, although we already used LD-pruned SNPs. Migration edges were calculated for values of m ranging from 1-10 and the options “-noss” and “-global” activated. The *V. latastei*-*monticola* complex was used to root the resulting trees. Finally, the optimum number of migration edges was calculated with the R package OptM v0.1.6⁶⁰ following the Evanno method⁶¹. The highest Δm is supposed to be reached with the optimum number of edges that can best explain the genetic variance from complex histories.

To explore the introgression landscape of the strongest identified introgression event (i.e., *V. latastei* to *V. seoanei*, see Results), we investigated the local window topologies across the genome. We first phased haplotypes of a high coverage dataset of all studied species (*high_coverage+out* dataset, see Table S2) with Beagle v5.2⁶². Then we inferred Neighbour-Joining local trees in non-overlapping 25-kbp windows with PhyML v3.3⁶³ and the script *phyml_sliding_windows.py* (available in https://github.com/simonhmartin/genomics_general), with the requirement of bearing at least 50 SNPs per window. For the resulting 27,217 local trees,

complete topology weightings were computed by TWISST⁶⁴, using two haploid sequences per species at 45x coverage. Topologies in which *V. latastei-monticola* clustered with *V. seoanei* were grouped and highlighted across the genome.

Demographic inference

We used Pairwise Sequentially Markovian Coalescent (PSMC) models⁶⁵ to infer the demographic histories of the eight main studied lineages, namely *V. latastei*, *V. monticola*, Western and Eastern *V. aspis*, *V. ammodytes*, *V. ursinii*, *V. seoanei* and *V. berus*, at high coverage. The three Iberian samples come from the resequencing data used for each of the three reference genomes, which was downsampled to 45x to equal the coverage of the other samples and filtered for low mapping (<30) and base quality (<30). Diploid consensus sequences for autosomal data of the eight high-coverage genomes were obtained by the pileup command of SAMtools. Minimum and maximum depths were set at a third and double the average coverage of each sample, respectively. A rate of 2.4e-9 substitutions/site/year, reported in⁶⁶ from⁶⁷, and a 6-year generation period, as estimated for *V. aspis*⁶⁸, were used. Other parameters were set following previous studies on Viperidae⁶⁹ and ten bootstraps were calculated for each genome. Results were plotted in base R, dividing Eurosiberian (*V. seoanei*, *V. berus* and *V. ursinii*) and Mediterranean species (*V. latastei*, *V. monticola*, Western and Eastern *V. aspis* and *V. ammodytes*) in two different plots to analyze common climatic oscillations due to similar affinities.

Population genomics

In order to explore population genomics of the genus *Vipera* in the Western Mediterranean, we created a dataset including only samples from *V. seoanei*, *V. aspis*, *V. latastei* and *V. monticola* (*westmed* dataset, see Table S2); apart from four species-specific datasets for the above mentioned species and another one for the joint *V. latastei-monticola* complex. Two putatively *V. latastei x aspis* hybrids were included only in the *westmed* dataset. For all these resulting low-coverage datasets, we filtered out individual genotypes with a read depth below four or above 150, or a quality score below 30, non-biallelic alleles, SNPs with more than 10% missingness, singletons (adjusting a different minimum allele frequency threshold per dataset) and invariant sites, and only kept SNPs from the autosomes with VCFtools. Then, to work with unlinked SNPs (uSNPs), we calculated Linkage Disequilibrium (LD) decay with PopLDdecay⁷⁰ in *V. seoanei*. As LD falls below 20% of correlation in between sites 0.5 kbp apart (Fig. S4), we pruned the datasets to kept one SNP per 0.5 kbp with VCFtools (--thin).

We first explore population structure of the four above-mentioned species and phylogenomic relationships between them by performing Principal Component Analyses (PCA) on the following datasets: *westmed*, *latastei-monticola*, *latastei*, *monticola*, *seoanei* and *aspis* (Table S2). The PCAs were implemented in Plink v1.90b6.21⁷¹. Secondly, to delve into population structure and secondary contacts in between lineages, we ran Admixture v1.3.0⁷² on the *westmed* dataset, calculating 10 cross-validation errors with 10 replicates. K values were set to 12, coinciding with the number of putative subspecies/lineages comprised within the four species. Optimal K values were chosen based on CV scores (see Table S16), the plateau profile of those values and the likelihood of the results in biological terms. Admixture results were visualized and plotted R v4.3.1⁷³, and the most likely K was depicted with georeferenced pie charts using QGIS⁷⁴.

Finally, we explored genomic Isolation-by-distance (IBD) patterns within *V. latastei* and *V. seoanei* following Hausdorf and Hennig.⁷⁵ From the *latastei* and *seoanei* datasets, we excluded all admixed individuals (in regard to the chosen Admixture results for the *westmed* dataset) to calculate genomic chord distances matrices among all individuals of each species with genodive v3.06⁷⁶. Two analogous matrices of geographic distances among samples were calculated as well with genodive. Geographic distances were transformed as $f_{ij} = \log_{10}(d_{ij} + c)$, being d_{ij} the

geographical distances in meters and c the 0.25-quantile of such raw distances⁷⁵. For these two species, we used the R package *prabclus*⁷⁷ to perform a Jackknife-based test for the equality of two independent within-lineage regressions (i.e. within the Western and Eastern lineages of *V. latastei* and *V. seoanei*, independently) to confirm that IBD develops alike within each lineage. Under those circumstances, within-lineage regressions of the same species can be merged into a single joint within-lineage regression. Secondly, either this joint regression or the independent specific within-lineage regressions previously calculated were compared to another regression over all comparisons within a species. The non-equality of them imply that the observed genomic differences cannot be explained by IBD alone. IBD regressions were plotted with *ggplot2*⁷⁸ in R.

Heterozygosity and Runs of Homozygosity

We calculated autosomal heterozygosity per individual from the unfiltered low-coverage dataset, following Mochales-Riaño et al.⁷⁹. We generated 100-kbp non-overlapping sliding windows along the autosomes and took variant and invariant sites with high base quality ($Q > 30$). Per window heterozygosity values, relative to the number of called sites, were calculated and then individual genome-wide heterozygosity means were obtained. For each species and the *V. latastei* \times *aspis* putative hybrids, the individual mean or a species mean among all samples were plotted as a barplot with *ggplot2*. To geographically depict gradients in genomic diversity and identify possible barriers to gene flow, we created three raster surfaces in QGIS through inverse distance weighted interpolation of the mean individual heterozygosity values for *V. latastei-monticola*, *V. seoanei* and *V. aspis*. *Vipera latastei* \times *aspis* hybrids were excluded from these interpolations.

Furthermore, we explored how different *Vipera* species are affected by Runs of Homozygosity (ROHs). First, we filtered the high-coverage dataset with different *Vipera* spp. from individual genotypes with a read depth below four or above 150, or a quality score below 30, non-biallelic alleles, and invariant sites and only kept SNPs from the autosomes. This dataset, with 10,754,630 SNPs was divided into several single-sample vcf files and purged out invariant sites again. ROHs were calculated then based on the density of those remaining heterozygous sites by the implemented hidden Markov model (HMM) in *bcftools roh* function⁸⁰, setting the allele frequency as 0.4 per default, being thus conservative when identifying ROHs⁸¹. We kept ROHs with a PhredScore above 70 and a minimum length of 100 kbps⁷⁹. Finally, we divided the amount of ROHs in each high-coverage genome by size, as different-sized ROHs arise from different causes⁸², in the following categories: short (0.1-0.5 Mbp), medium (0.5-1 Mbp) and long (>1 Mbp); and plotted the result with *ggplot2*, showing the amount of ROHs as a percentage of the total autosomal genome (*V. latastei* as reference).

Climatic niche overlap

With the purpose of investigating the role of adaptative introgression in between the ecologically divergent *V. latastei* and *V. seoanei*, we performed pairwise niche overlap tests between the major lineages identified in both species (see Admixture results). We included 1,105 and 508 occurrence records for *V. latastei* and *V. seoanei*, respectively, with a spatial resolution of 5 arc minutes, throughout the entire distribution of the two species, extracted from⁸³⁻⁸⁵. For these locations, we obtained a set of 19 environmental variables from WorldClim v2.1⁸⁶. To correctly assign the records into the genomically-identified lineages (i.e. Eastern or Western *V. seoanei* and Eastern or Western *V. latastei*), we interpolated the genomic affinities inferred from Admixture results using ordinary kriging as interpolator method in ArcGis v.10.5⁸⁷. We separated intraspecific lineages in both species, calculating the contact zones between them and delimitating the records for each cluster with a genomic affinity higher than 0.73 for each lineage. Afterwards, based again on Admixture results ($K=7$ for *westmed* dataset), we identified isolated populations of the Western *V. seoanei* vs Western *V. latastei* contact zone as putatively admixed: the nuclei of Montalegre and Eastern Galicia, south to Sil River for *V. seoanei* and the nuclei of Gerês, Montesinho-Sanabria and northern coastal Porto for *V. latastei* (admixed individuals for those five nuclei shown in Fig. A2). Later, we subset six different record datasets with their respective

environmental variables: “Western *V. seoanei*” (n=362 records; including all populations with or without admixture from *V. latastei*), “Western *V. seoanei* no-admix” (n=326, excluding records from the cited two admixed nuclei), “Eastern *V. seoanei*” (n=146), “Western *V. latastei*” (n=488, with or without *V. seoanei* admixture), “Western *V. latastei* no-admix” (n=412, excluding records from the three cited nuclei) and “Eastern *V. latastei*” (n=617). Climatic data was summarized by a PCA into three Principal Components (factor loadings in Table S9) and, following a three-dimensional hypervolume approach⁸⁸, we used those PCs to describe the environmental niche of the subgroups. Analyses were performed in R with the hypervolume package⁸⁹, using a Silverman bandwidth estimator and a set of 999 random points to sample the kernel density to build the hypervolumes. To quantify the climatic niche overlap the Sørensen index (K) and the Overlap Index (OI) were used^{88–90}. These statistics quantify predicted niche similarity, and they both range from no overlap (0) to identical niche (1). The OI index relates the observed and maximum values of K and is preferred when niches present different sizes^{90,91}. We statistically tested the significance of the overlap comparisons with a randomization procedure⁹⁰. By comparing the produced K and OI indices, we tested if observed overlaps statistically differed from overlaps obtained from random lineage-blind occurrences. This procedure was performed 999 times to generate a null distribution and to evaluate the significance of each overlap. Depending on how much the niche overlap in between Western *V. seoanei* and Western *V. latastei* increases when adding admixed populations, this could be a line of evidence supporting the adaptive introgression, i.e. both species benefit from the introgression with a distant and ecologically differing species, as these events allow the resulting admixed individuals to increase their range beyond the potential range of the species. Once lineage-pairwise niche overlap was tested through hypervolumes, we plotted as well 2D PCAs in between sympatric species/lineages, highlighting putatively admixed individuals to better visualized if they inhabit areas beyond the common niche of the species in absence of recent introgression.

Differential Gene Expression analysis

RNA-Seq data, already filtered and adapter-trimmed (see “Genome annotation” section), was again mapped to the annotated reference genome of *V. latastei* with STAR v2.7.10b. Mapping was quality-checked with PicardTools CollectRnaSeqMetrics and then used the sorted bam files to count gene hits with the count function of HTSeq v2.0.4⁹², in which the number of reads are taken as an indicator of gene expression. We used the mode “-m intersection-nonempty” to be as stringent as possible, finally generating a gene expression matrix of raw read counts, which was then normalized by both library size and composition using the median-of-ratios algorithm of DESeq2⁹³ in R. After normalization, genes with lower than ten counts in total were filtered out, keeping 17,811 genes, and data was transformed with the regularized logarithm of DESeq2 (rlog function). We calculated the variance of each gene among the 19 samples and selected the top 1,000 genes with the highest variance to perform an unsupervised hierarchical clustering analysis, plotted with the R package ComplexHeatmap⁹⁴. In this plot, we checked whether venom gland samples and other duplicated organs (i.e., kidneys, liver, pancreas) clustered together. Gene expression was standardized with Z-score scaling. Finally, we tested for significantly upregulated genes (log2 fold-change > 2) in the venom glands vs all other tissues with DESeq2, correcting for multiple tests with Benjamini and Hochberg adjusted *p-values* (FDR: 5%). The function of the putatively venom gland-upregulated genes was identified through protein BLAST⁹⁵.

Venom decomplexation and initial characterization of *V. latastei* proteomes

Crude lyophilized venom was dissolved in 0.05% trifluoroacetic acid (TFA) and 5% acetonitrile (ACN) to a final concentration of 10 mg/mL. Insoluble material was removed by centrifugation in an Eppendorf centrifuge at 13,000xg for 10 min at room temperature, and 600 µg were fractionated by RP-HPLC using an Agilent LC 1200 Infinity II High Pressure Gradient System equipped with a Teknokroma Europa C18 (25 cm x 5 mm, 5 µm particle size, 300 Å pore size) column and a DAD detector. The column was developed at a flow rate of 1.0 mL/min with a linear gradient of 0.1% TFA in MilliQ® water (solution A) and in acetonitrile (ACN, solution

B), isocratic (5% B) for 5 min, followed by 5-25% B for 10 min, 25-45% B for 60 min, and 45-70% B for 10 min. The elution was monitored at 215 nm with a reference wavelength of 400 nm. Fractions were collected manually and dried in a vacuum centrifuge (Savant™, ThermoFisher Scientific). Molecular masses of the purified proteins were estimated from Coomassie Brilliant Blue G-250 SDS-10% and 15% polyacrilamide gels, run under non-reduced and reduced conditions, respectively. Molecular masses were also measured by electrospray ionization (ESI) mass spectrometry (MS). To this end, the proteins eluted in the different RP-HPLC fractions were separated by nano-Acquity UltraPerformance LC® (UPLC®) using BEH130 C18 (100µm x 100mm, 1.7 µm particle size) column in-line with a Waters SYNAPT G2 High Definition Mass Spectrometry System. The flow rate was set to 0.6 µL/min and the column was developed with a linear gradient of 0.1% formic acid in MilliQ® (solution A) and ACN (solution B), isocratically 1% B for 1 min, followed by 1-12% B for 1min, 12-40% B for 15min, 40-85% B for 2min. Monoisotopic and isotope-averaged molecular masses were calculated by manually deconvolution of the isotope-resolved multiply-charged MS1 mass spectra.

Identification of the venom proteins eluted in the RP-HPLC fractions of the *V. latastei* venoms

Protein bands found in each RP-HPL fraction were excised from Coomassie Brilliant Blue-stained SDS-PAGE and the pieces were subject to automated in-gel disulphide bond reduction (10 mM dithiothreitol, 30 min at 65 °C), cysteine alkylation (50 mM iodoacetamide, 2h in the dark at room temperature), and overnight digestion with sequencing-grade trypsin (66 ng/µL in 25 mM ammonium bicarbonate, 10% ACN; 0.25 µg/sample), using a Genomics Solution ProGest™ Protein Digestion Workstation. Tryptic digests were dried in a vacuum centrifuge (SPD SpeedVac®, ThermoSavant), redissolved in 14 µL of 5% ACN containing 0.1% formic acid, and 7 µL submitted to LC-MS/MS. Tryptic peptides were separated by nano-Acquity UltraPerformance LC® (UPLC®) in-line with a Waters SYNAPT G2 High Definition Mass Spectrometry System as above. For peptide ion fragmentation by collision-induced dissociation tandem mass spectrometry (CID-MS/MS), the electrospray ionization (ESI) source was operated in positive ion mode, and both singly and multiply charged ions were selected for CID-MS/MS at sample cone voltage of 28 V and source temperature of 100 °C. The UPLC eluate was continuously scanned from 300 to 1990 *m/z* in 1 s and peptide ion MS/MS analysis was performed over the range *m/z* 50–2000 with scan time of 0.6 s. Search parameters were: taxonomy: bony vertebrates; enzyme: trypsin (two-missed cleavage allowed); MS/MS mass tolerance was set to ±0.6 Da; carbamidomethyl cysteine and oxidation of methionine were selected as fixed and variable modifications, respectively.

All matched MS/MS data were manually checked. For missing/incomplete identifications, the MS/MS fragmentation spectra were interpreted manually (*de novo* sequencing), and amino acid sequence similarity searches were performed at <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. against the non-redundant NCBI protein sequence database, using the default parameters of the BLASTP. Fragmentation spectra of peptides that yielded daughter ions diagnostic of the endogenous peptides BPP (*m/z* 116.1, *y*₁ = P, and *m/z* 213.1, *y*₂ = PP) and tripeptide inhibitors of metalloproteinases (SVMPi) at *m/z* 205.1 (*y*₁ = W), *m/z* 112.1 (*b*₁ = pyroglutamate, Z) and *m/z* 240.1 (*b*₂ = Z(K/Q) or *m/z* 226.1 (*b*₂ = ZN)^{96,97} were also sequenced manually. The final proteomics-gathered proteomes were manually assigned to the subset of venom toxins annotated in the reference genome of *V. latastei*.

Quantification of the venom proteomes

The relative abundances of the chromatographic peaks obtained by reverse-phase HPLC fractionation of the crude venom were calculated as “% of total peptide bond concentration in the peak” by dividing the peak area by the total area of the chromatogram^{98,99}. For chromatographic peaks containing single components (as judged by SDS-PAGE and/or MS1), this figure is a good estimate of the % by weight (g/100 g) of the pure venom component¹⁰⁰. When more than one

venom protein was present in a reverse-phase fraction, their proportions (% of total protein band area) were estimated by densitometry of Coomassie-stained SDS-polyacrylamide gels using MetaMorph® Image Analysis Software (Molecular Devices). Conversely, the relative abundances of different proteins contained in the same SDS-PAGE band were estimated based on the relative ion intensities of the three most abundant peptide ions associated with each protein by MS/MS analysis. The relative abundances of the protein families present in the venom were calculated as the ratio of the sum of the percentages of the individual proteins from the same toxin family to the total area of venom protein peaks in the reverse-phase chromatogram, and these figures were used to compile a compositional pie chart for the venom proteomes^{98,99}.

Characterization of venom-encoding genes and venom selection analyses

We used proteomic data (i.e. peptide sequences) to identifying the venom-encoding genes in each *Vipera* spp. reference genome. For *V. latastei*, we used the data produced in this study (Methods above), whereas for *V. seoanei* and *V. aspis* we used already published data^{101,102}. We translated into protein sequences all annotated isoforms of the three genomes with AGAT and searched for exact matches from each peptide database with the translated isoforms of each corresponding species using custom bash scripts. For all matching genes (i.e. candidate venom-encoding genes), we kept the isoform with the highest number of hits against unique peptide sequences as representative for that gene. We identified the function of each protein sequence with BLASTP, manually discarding candidate venom-encoding genes matching non-toxic proteins. For candidate genes with more than one isoform with the same number of hits, we kept the one with the higher resemblance to known toxins (i.e. a higher e-value in BLASTP). We checked that results from proteomic evidence were similar to the ones based on differential gene expression for *V. latastei*. For the main toxin families (i.e. SVMs, SVSPs, PLA₂s, CTLs, Kunitz peptides), we also checked that the chromosomal location of each gene cluster agreed between the three *Vipera* species (see Macrosynteny results) and the sister subfamily Crotalinae. Finally, once the chromosomal location of each of the above-mentioned toxin families was known, we conducted chromosome-wide blasts of all isoforms annotated in such chromosome against custom databases of Viperidae toxins. In that way, we were able to identify putative candidate genes that were not expressed in the sampled venoms/venom glands.

Selection signals on venom-encoding genes

Once the toxin-encoding genes were identified, we extracted the coding sequences (CDS) with AGAT and aligned with all copies belonging to each toxin family (n=11, families with at least 3 copies present in 2 species) using the MUSCLE algorithm implemented on MEGA 11. We trimmed the terminal stop codons and performed d_N/d_S (ω) ratio tests in each family of toxins with BUSTED¹⁰³ and FEL¹⁰⁴ analyses implemented in Datamonkey 2¹⁰⁵. BUSTED tests for episodic diversification, whereas FEL looks for pervasive diversifying and purifying selection. For FEL tests, *p-value* was left as default (0.1). Later, to estimate the relative ubiquity or pervasiveness of both types of selection in each family, we divided the number of sites under selection estimated with FEL tests by the number of codons in the alignment and the number of copies in the family, since the more copies or the more codons, the more likely is to find selection signals. This selection ubiquity values were normalized and used to explore the relationship between the ubiquity of diversifying and purifying selection with the number of copies within a family or its relative expression in the venom. The expression was measured in relative abundance of the family in the venom, taken from Giribaldi et al.¹⁰². The disintegrin expression percentage was added to SVMs, as they derived from the same genes. Regressions were performed in base R and plotted with ggplot2.

References

1. Carøe, C. *et al.* Single-tube library preparation for degraded DNA. *Methods Ecol Evol* **9**, 410–419 (2018).
2. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10–12 (2011).
3. De Coster, W., D’Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
4. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).
5. Hu, J. *et al.* An efficient error correction and accurate assembly tool for noisy long reads. *bioRxiv* 2023.03.09.531669 (2023) doi:10.1101/2023.03.09.531669.
6. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).
7. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
8. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
9. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: 1303.3997v2* (2013) doi:10.48550/arXiv.1303.3997.
10. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963- (2014).
11. Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
12. Ghurye, J., Pop, M., Koren, S., Bickhart, D. & Chin, C. S. Scaffolding of long read assemblies using long range contact information. *BMC Genomics* **18**, (2017).
13. Alonge, M. *et al.* Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol* **23**, 258 (2022).
14. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**, 540–546 (2019).
15. Kundu, R., Casey, J. & Sung, W.-K. HyPo: Super Fast & Accurate Polisher for Long Read Genome Assemblies. *bioRxiv* (2019) doi:10.1101/2019.12.19.882506.
16. Harry, E. PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps. <https://github.com/sanger-tol/PretextView> (2022).
17. Meyne, J., Ratliff, R. L. & Moyzis, R. K. Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proceedings of the National Academy of Sciences* **86**, 7049–7053 (1989).
18. Girgis, H. Z. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinformatics* **16**, 227 (2015).

19. Kobel, H. R. Morphometrische Karyotypanalyse einiger Schlangen-Arten. *Genetica* **38**, 1–31 (1967).
20. Challis, R., Richards, E., Rajan, J., Cochrane, G. & Blaxter, M. BlobToolKit – Interactive quality assessment of genome assemblies. *G3 Genes|Genomes|Genetics* **10**, 1361–1374 (2020).
21. Saint Girons, H. & Duguy, R. Écologie et position systématique de *Vipera seoanei* Lataste, 1879. *Bulletin de la Société Zoologique de France* **101**, 325–339 (1976).
22. Formenti, G. et al. Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics* **38**, 4214–4216 (2022).
23. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
24. Keilwagen, J. et al. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res* **44**, e89–e89 (2016).
25. Margres, M. J. et al. The Tiger Rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. *PNAS* **118**, (2021).
26. Hogan, M. P. et al. The Chemosensory Repertoire of the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*) Reveals Complementary Genetics of Olfactory and Vomeronasal-Type Receptors. *J Mol Evol* **89**, 313–328 (2021).
27. Vonk, F. J. et al. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proceedings of the National Academy of Sciences* **110**, 20651–20656 (2013).
28. Ullate-Agote, A., Milinkovitch, Michel C. & Tzika, Athanasia C. The genome sequence of the corn snake (*Pantherophis guttatus*), a valuable resource for EvoDevo studies in squamates. *Int J Dev Biol* **58**, 881–888 (2014).
29. Tang, S., Lomsadze, A. & Borodovsky, M. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res* **43**, e78–e78 (2015).
30. Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767–769 (2016).
31. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
32. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform* **3**, lqaa108 (2021).
33. Kriventseva, E. V et al. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res* **47**, D807–D811 (2019).
34. Gabriel, L., Hoff, K. J., Brůna, T., Borodovsky, M. & Stanke, M. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics* **22**, 566 (2021).
35. Tang, H. et al. Synteny and collinearity in plant genomes. *Science* (1979) **320**, 486–488 (2008).

36. Pasquesi, G. I. M. *et al.* Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat Commun* **9**, (2018).
37. Dainat, J. *et al.* AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. (Version v0.8.0). *Zenodo* (2023) doi:10.5281/zenodo.8178877.
38. Kiełbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res* **21**, 487–493 (2011).
39. Tang, H. *et al.* jcv: JCVI utility libraries. *Zenodo* (2015) doi:10.5281/zenodo.31631.
40. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
41. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 201178 (2018) doi:10.1101/201178.
42. Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A. & RoyChoudhury, A. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol Biol Evol* **29**, 1917–1932 (2012).
43. Bouckaert, R. *et al.* BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput Biol* **10**, e1003537- (2014).
44. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
45. 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).
46. Stange, M., Sánchez-Villagra, M. R., Salzburger, W. & Matschiner, M. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. *Syst Biol* **67**, 681–699 (2018).
47. Šmíd, J. & Tolley, K. A. Calibrating the tree of vipers under the fossilized birth-death model. *Sci Rep* **9**, (2019).
48. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol* **67**, 901–904 (2018).
49. Bouckaert, R. DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372–1373 (2010).
50. Jin, J.-J. *et al.* GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol* **21**, 241 (2020).
51. Gao, D., Liu, C. & Hao, S. Characterization of the complete mitochondrial genome of the European adder (*Vipera berus*) using Illumina sequencing technology. *Conserv Genet Resour* **10**, 793–796 (2018).
52. Allio, R. *et al.* MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol Ecol Resour* **20**, 892–905 (2020).
53. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* **38**, 3022–3027 (2021).
54. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).

55. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol* **34**, 772–773 (2017).
56. Malinsky, M., Matschiner, M. & Svardal, H. Dsuite-Fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour* **21**, 584–595 (2021).
57. Koppetsch, T., Malinsky, M. & Matschiner, M. Among-species rate variation produces false signals of introgression. *bioRxiv* 2023.05.21.541635 (2023) doi:10.1101/2023.05.21.541635.
58. Malinsky, M. *et al.* Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat Ecol Evol* **2**, 1940–1955 (2018).
59. Pickrell, J. & Pritchard, J. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* **8**, e1002967 (2012).
60. Fitak, R. R. OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biol Methods Protoc* **6**, bpab017 (2021).
61. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**, 2611–2620 (2005).
62. Browning, B. L., Tian, X., Zhou, Y. & Browning, S. R. Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics* **108**, 1880–1890 (2021).
63. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307–321 (2010).
64. Martin, S. H. & Van Belleghem, S. M. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429–438 (2017).
65. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
66. Pasquesi, G. I. M. *et al.* Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat Commun* **9**, 2774 (2018).
67. Green, R. E. *et al.* Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science (1979)* **346**, 1254449 (2014).
68. Barbanera, F. *et al.* Molecular phylogeography of the asp viper *Vipera aspis* (Linnaeus, 1758) in Italy: Evidence for introgressive hybridization and mitochondrial DNA capture. *Mol Phylogenet Evol* **52**, 103–114 (2009).
69. 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).
70. Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M. & Yang, T.-L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788 (2019).
71. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575 (2007).
72. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655–1664 (2009).

73. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>. (2021).
74. QGIS.org. QGIS Geographic Information System. QGIS Association. <http://www.qgis.org>. (2022).
75. Hausdorf, B. & Hennig, C. Species delimitation and geography. *Mol Ecol Resour* **20**, 950–960 (2020).
76. Meirmans, P. G. genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Mol Ecol Resour* **20**, 1126–1131 (2020).
77. Hennig, C. & Hausdorf, B. Package ‘prabclus’, version 2.3-1. Functions for clustering of presence-absence, abundance and multilocus genetic data. <http://cran.r-project.org/web/packages/prabclus/>. (2019).
78. Wickham, H. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>. (2016).
79. Mochales-Riaño, G. *et al.* Genomics reveals introgression and purging of deleterious mutations in the Arabian leopard (*Panthera pardus nimr*). *iScience* **26**, (2023).
80. Narasimhan, V. *et al.* BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* **32**, 1749–1751 (2016).
81. Dyson, C. J. *et al.* Genome of the endangered Guatemalan Beaded Lizard, *Heloderma charlesbogerti*, reveals evolutionary relationships of squamates and declines in effective population sizes. *G3* **12**, jkac276 (2022).
82. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: Windows into population history and trait architecture. *Nat Rev Genet* **19**, 220–234 (2018).
83. SIARE. Servidor de Información de anfibios y reptiles de España. Available at: <https://siare.herpetologica.es/>. (2024).
84. Chamorro, D., Martínez-Freiría, F., Real, R. & Muñoz, A. R. Understanding parapatry: How do environment and competitive interactions shape Iberian vipers’ distributions? *J Biogeogr* **48**, 1322–1335 (2021).
85. Martínez-Freiría, F. *et al.* Integrative taxonomy reveals two species and intraspecific differentiation in the *Vipera latastei–monticola* complex. *Journal of Zoological Systematics and Evolutionary Research* **59**, 2278–2306 (2021).
86. Fick, S. E. & R.J. Hijmans. Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* (2017).
87. ESRI. ArcGIS Desktop: Release 10. Environmental Systems Research Institute, Redlands. (2011).
88. Martínez-Freiría, F. *et al.* Climatic refugia boosted allopatric diversification in Western Mediterranean vipers. *J Biogeogr* **47**, 1698–1713 (2020).
89. Blonder, B., Lamanna, C., Violle, C. & Enquist, B. J. The n-dimensional hypervolume. *Global Ecology and Biogeography* **23**, 595–609 (2014).
90. Mochales-Riaño, G. *et al.* Hidden in the sand: Phylogenomics unravel an unexpected evolutionary history for the desert-adapted vipers of the genus *Cerastes*. *Mol Phylogenet Evol* **191**, 107979 (2024).

91. Simó-Riudalbas, M., Tarroso, P., Papenfuss, T., Al-Sariri, T. & Carranza, S. Systematics, biogeography and evolution of *Asaccus gallagheri* (Squamata, Phyllodactylidae) with the description of a new endemic species from Oman. *Syst Biodivers* **16**, 323–339 (2018).
92. Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E. & Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943–2945 (2022).
93. Love, M., Anders, S. & Huber, W. Differential analysis of count data—the DESeq2 package. *Genome Biol* **15**, 10–1186 (2014).
94. Gu, Z. Complex heatmap visualization. *iMeta* **1**, e43 (2022).
95. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).
96. Coutinho-Neto, A. *et al.* ESI-MS/MS identification of a bradykinin-potentiating peptide from Amazon *Bothrops atrox* snake venom using a hybrid Qq-*oa* TOF mass spectrometer. *Toxins (Basel)* **5**, 327–335 (2013).
97. Menin, L. *et al.* High throughput screening of bradykinin-potentiating peptides in *Bothrops moojeni* snake venom using precursor ion mass spectrometry. *Toxicon* **51**, 1288–1302 (2008).
98. Calvete, J. J. Next-generation snake venomomics: protein-locus resolution through venom proteome decomplexation. *Expert Rev Proteomics* **11**, 315–329 (2014).
99. Eichberg, S., Sanz, L., Calvete, J. J. & Pla, D. Constructing comprehensive venom proteome reference maps for integrative venomomics. *Expert Rev Proteomics* **12**, 557–573 (2015).
100. Calderón-Celis, F., Cid-Barrio, L., Encinar, J. R., Sanz-Medel, A. & Calvete, J. J. Absolute venomomics: Absolute quantification of intact venom proteins through elemental mass spectrometry. *J Proteomics* **164**, 33–42 (2017).
101. Avella, I. *et al.* One size fits all—venomics of the Iberian adder (*Vipera seoanei*, Lataste 1878) reveals low levels of venom variation across its distributional range. *Toxins (Basel)* **15**, 371 (2023).
102. Giribaldi, J. *et al.* Venomomics of the asp viper *Vipera aspis aspis* from France. *J Proteomics* **218**, 103707 (2020).
103. Murrell, B. *et al.* Gene-wide identification of episodic selection. *Mol Biol Evol* **32**, 1365–1371 (2015).
104. Kosakovsky Pond, S. L. & Frost, S. D. W. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**, 1208–1222 (2005).
105. Weaver, S. *et al.* Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol* **35**, 773–777 (2018).