RESEARCH | Article in Press

Population genomics unravel structural variation of venom-encoding genes across the desert viper genus *Cerastes*

Gabriel Mochales-Riaño · Adrián Talavera · Bernat Burriel-Carranza · Johannes Els · Mohammed Shobrak · Salem Busais · Maria Estarellas · Jiří Šmíd · Fernando Martínez-Freiría · Mark J. Margres · Salvador Carranza

Received: 31 March 2026 / Accepted: 9 September 2026

Abstract

Understanding the molecular mechanisms underlying phenotypic differentiation is central to evolutionary biology. Yet most traits are shaped by complex, polygenic architectures, making links between genomic change and phenotype difficult to resolve. Venom provides an advantageous model for genotype–phenotype research because it is produced in a specialized secretory tissue, governed by a modular regulatory system, and largely composed of proteins from a limited set of recurrent toxin gene families. In parallel, genomics enables reconstruction of species evolutionary histories and detection of structural variation within and between species. Here we used whole genome sequencing data from 27 individuals to investigate the evolutionary history of the three species of the desert viper genus *Cerastes* (*C. cerastes*, *C. gasperettii* and *C. vipera*), a group of arid-adapted Palearctic snakes. Whole genome produced broadly concordant inferences of population structure, phylogenomic relationships, and introgression, with previous studies based on genome-wide datasets, although whole genomes provided finer resolution. Conservation genomic analyses further revealed pronounced genomic consequences of long-term isolation in relict Arabian populations of *C. cerastes*. We also detected extensive interspecific and intraspecific structural variation across major toxin gene regions, suggesting that genomic structural changes may contribute to variation in venom composition. These findings indicate that structural variation is consistent with previously reported differences in venom composition among *Cerastes*. Overall, our study underscores the value of integrative genomic approaches for disentangling the multiple evolutionary processes shaping complex adaptive traits such as venom.

Keywords Snake · Toxin evolution · Genomics · Structural variants

https://doi.org/10.1186/s12864-026-13357-8

Published online: 17 September 2026

Background

Identifying the molecular pathways underpinning phenotypic differentiation is crucial to understanding how species evolve, and therefore linking variation in genotype to phenotype [1, 2]. Several mechanisms have been proposed to drive genetic variation [3, 4], including structural changes in the genome as well as changes in gene regulation (such as transcriptional regulation or epigenetics) [5–11]. However, structural variants can also have an impact on the genome, with potential phenotypic consequences [12–14]. Structural variation is defined as large-scale differences in the structure of DNA, including deletions, duplications, insertions, inversions, and translocations of genomic segments [15]. Structural variation has been mostly studied in humans, as they can cause genomic disorders [15–17]. However, recent studies in non-model organisms have suggested a major role in adaptation and speciation [18–20]. Among structural variation mechanisms, gene duplications and deletions play an important role in phenotypic variation [8, 21–25]. Whilst gene duplication can increase gene dosage as well as neo- and/or subfunctionalization, gene deletions can lead to partial or complete loss of coding or regulatory regions, potentially resulting in reduced gene function, altered regulation, or in some cases complete gene loss and gene rearrangements [26, 27]. In addition to structural variation, regulatory evolution represents a major and often dominant driver of phenotypic diversity, particularly in venom systems. Changes in gene expression, mediated by transcriptional regulation, chromatin accessibility, and epigenetic modifications, have been shown to play a central role in shaping venom composition and its ecological variation across snake lineages [5, 6, 11, 27]. Nevertheless, most phenotypic traits arise from complex pathways involving multiple genes [28], making the link between genomic changes, including structural changes, and polygenic traits an ongoing unresolved issue.

Venom is a complex phenotype in which diverse gene families give rise to its protein composition [29, 30]. Its study serves as an effective system for mapping the relationships between genotypic and phenotypic variation [11, 31, 32]. Venomous organisms include multiple animal groups such as cnidarians, molluscs, arthropods, squamates, and even mammals [29, 30, 33]. Among those, snakes are the fundamental model system in venom research, being one of the most life-threatening animal groups to humans [34]. Snake genes that encode venom arose through gene duplication of non-toxin paralogs followed by neo- and/or subfunctionalization, as well as venom gland-specific expression [29, 30, 35]. Additional tandem duplication events often occurred and expanded the number of toxin genes for each family, making them the main components of snake venom [22, 36]. However, deletions impacting venom genes have also been reported (e.g [8, 21, 22, 37]). For that, intraspecific and interspecific venom variation provides an excellent opportunity to study genomic changes at shallow scales of evolutionary divergence [10, 31, 36].

In parallel, genomic approaches allow a plethora of analyses to characterize the evolutionary history of species [38, 39]. Species may experience different evolutionary trajectories, with an impact on phenotypic traits, including venom [8, 13, 40]. For that, characterizing the evolutionary history of species is also important when studying phenotypic variation. In fact, the rise of genomics has led to an increase in the number of high-quality reference genomes available for venomous species in the last years [8, 32, 41–45], allowing the study of species' evolutionary history and venom from a genomic perspective.

Snakes from the family Viperidae are one of the most commonly studied venomous groups [29, 30, 36]. However, studies on viperids are highly biased towards the Crotalinae subfamily, which contains rattlesnakes and other pit vipers (subfamily Crotalinae, e.g. [8, 32, 41]). The Viperinae subfamily, with some of the most medically relevant genera such as *Echis*, *Daboia* or *Bitis*, have been historically less studied (especially from a genomic perspective) most likely due to the difficulty of sampling in remote regions as well as their large distributions [46]. Within Viperinae, the genus *Cerastes* is starting to emerge as a genomic system to study venom evolution [42]. It is a small monophyletic group of medically important venomous snakes [47–50] adapted to arid conditions [51–53] and widely distributed across the Sahara Desert and the Arabian Peninsula [52, 54, 55]. While C. *vipera* and *C. cerastes* are mostly present in North Africa (although the latter is also found in southwestern Arabia), *C. gasperettii* is exclusively found in Arabia. Although proteomic studies in *Cerastes* are limited,

interspecific variation has been observed [42, 56–59]. In *C. gasperettii*, Snake Venom Serine Proteases (*SVSPs*) and Snake Venom Metalloproteinases (*SVMPs*) together comprise more than half of the venom composition [42]. Other toxins present in lower proportions include Phospholipases A_2 (PLA_2), L-amino-acid oxidases (*LAAO*), C-type lectins (*CTL*) and disintegrins (*DISI*). *Cerastes cerastes* displays a similar venom profile to *C. gasperettii*, although only *SVMPs* account for more than half of its venom. However, significant intraspecific variation, particularly in *SVMPs*, *SVSPs*, *CTLs* and *LAAOs* has been reported [56–59]. Lastly, venom characterization in *C. vipera* is limited to a single study, which revealed lower venom complexity compared to *C. cerastes* [56]. While proteomic studies have advanced our understanding of venom composition and variation in this genus, genomic data can also provide complementary insights into the evolution and diversification of venom [60].

In this study, we aimed to investigate the evolutionary history, genome-wide diversity, and venom evolution of the genus *Cerastes*. To this end, we generated whole-genome resequencing data for 27 samples covering most of the geographic range of the three species within the genus, along with one outgroup (*Echis omanensis*) (Table 1). We reconstructed the evolutionary history of the genus using population genomics and phylogenomics analyses, estimated standard conservation genomics parameters to assess genome-wide diversity and inbreeding via Runs of Homozygosity (ROHs), and studied fluctuations in population effective size through time as well as historical gene flow. Finally, once we characterized the evolutionary histories of each species, we focused on the main toxin families to study their evolutionary dynamics and identify intra- and interspecific structural variation. Overall, our work provides insights into the evolutionary history of *Cerastes*, its conservation status, and the genomic mechanisms shaping viper venom variation.

Table 1 All individuals sampled in this study, with their corresponding species name, id and coverage levels. Individuals with high-coverage (>25x) are shown in bold (as well as their genome-wide heterozygosity is indicated) and individuals used for PSMC analysis are underlined

Species	Code	ID	Coverage	Heterozygosity	Country
Cerastes cerastes cerastes	1	**12.953**	29.0498	0.0026	Chad
Cerastes cerastes cerastes	2	BEV10313	11.2182		Morocco
Cerastes cerastes cerastes	3	9059	11.4745		Morocco
Cerastes cerastes cerastes	4	9076	10.5075		Western Sahara
Cerastes cerastes cerastes	5	92	9.90682		Egypt
Cerastes cerastes cerastes	6	**CCC**	27.4703	0.0042	Mauritania
Cerastes cerastes cerastes	7	SPM0002585	12.29		Egypt
Cerastes cerastes cerastes	8	SPM000783	9.81		Algeria
Cerastes cerastes cerastes	9	A20487	6.66		Israel
Cerastes cerastes hoofieni	10	**CN15035**	26.0386	0.0013	Saudi Arabia
Cerastes cerastes hoofieni	11	CN15888	9.35198		Yemen
Cerastes cerastes hoofieni	12	JIR250	9.2771		Yemen
Cerastes vipera	13	**17V080**	25.3617	0.005	Western Sahara
Cerastes vipera	14	BEV10185	26.8555	0.0035	Algeria
Cerastes vipera	15	**BEV9125**	27.1994	0.004	Morocco
Cerastes vipera	16	A17272	1.67		Israel
Cerastes vipera	17	SPM3923	6.33		Morocco
Cerastes gasperettii gasperettii	18	AO141	11.8433		Oman
Cerastes gasperettii gasperettii	19	**BEV10077**	30.1788	0.0022	Kuwait
Cerastes gasperettii gasperettii	20	**CN15045**	27.0268	0.0025	Saudi Arabia
Cerastes gasperettii gasperettii	21	CN15650	11.8272		Saudi Arabia
Cerastes gasperettii gasperettii	22	**CN2698**	30.5271	0.0024	Oman
Cerastes gasperettii gasperettii	23	CN3856	11.079		Oman
Cerastes gasperettii gasperettii	24	CN3923	10.6037		Oman
Cerastes gasperettii gasperettii	25	**CG1**	78.21	0.0024	United Arab Emirates
Cerastes gasperettii cf. *mendelssohni*	26	A17182	7.37		Israel
Echis omanensis	27	CN11991	18.5953		United Arab Emirates

Materials and methods

Data collection

A total of 27 samples representing the distribution range of *C. cerastes* (including nine *C. c. cerastes* and three *C. c. hoofieni*), *C. gasperettii* (eight *C. g. gasperettii* and one individual for the putative subspecies *C. g. mendelssohni*) and five *C. vipera* were included in this study (Fig. S1 and Table 1). *Cerastes boehmei* was not included as it has been recently synonymized within *C. vipera* [50]. From those, a total of 15 individuals were the same ones used in [48]. In addition, we included one individual of *Echis omanensis* as an outgroup. Samples consisted of tail tips or scale clips collected during previous field expeditions, preserved in absolute ethanol, and stored at − 20 °C until processing. Genomic DNA was extracted using the MagAttract HMW Kit (Qiagen) following the manufacturer's instructions. Then, Illumina libraries were prepared and were paired-end sequenced on an Illumina NovaSeq 6000 (2 × 150 bp) following the manufacturer's protocol and obtaining around 15 to 20 Gb of raw data for medium-coverage samples and between 50 and 60 Gb of raw data for high-coverage samples. Additionally, Illumina reads for one high-coverage sample (CG1) were obtained from [42].

Data processing

Raw reads for each of the 27 samples (26 *Cerastes* and one *E. omanensis*) were filtered, and adapters removed with fastp v.0.23.3 [61]. A minimum base quality score was set to 30, and adapter detection for paired-end sequencing was activated, with a required fragment length of 50 bp. 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 v.0.12.1 [62] to ensure data quality and absence of adapters. Filtered reads were mapped against the Arabian horned viper (*Cerastes gasperettii*) reference genome (GenBank assembly accession: JAZHBS000000000; [42]) with bwa-mem v0.7.17 [63]. Mapped reads were sorted with Samtools v1.9 [64]. Duplicated reads were marked and removed with PicardTools v.2.28.0 [65] and only uniquely mapped reads with mapping quality higher than 30 were kept. We downsampled high-coverage samples aiming for an average of 10x coverage, to have similar coverage levels of medium-coverage samples. SNP calling was carried out with HaplotypeCaller from GATK 4.1.7 [66] 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 non-variant sites. Finally, the whole dataset split by chromosome was concatenated with bcftools concat [67] keeping only the autosomes. Then, we split the dataset in two, one containing only high-coverage samples (dataset 1) and a second one with medium-coverage samples that included the downsampled high-coverage samples (dataset 2). Then, we filtered the raw datasets by excluding variants matching at least one of the following criteria: Quality by Depth (QD) < 10, Mapping Quality (MQ) < 50, Fisher Strand test (FS) > 10, StrandOddsRatio (SOR) > 4, MQRankSum < − 5 && MQRankSum > 5 and ReadPosRankSum < − 5 && ReadPosRankSum > 5. Later, dataset 2 was filtered using vcftools v.0.1.16 [68] with a minimum variant quality of 30, removing indels, keeping only biallelic SNPs, allowing 10% missing data and filtering for minor allele frequency (MAF) of 0.001, filtering out monomorphic sites and keeping only variable positions. Repetitive regions were identified from the softmasked Arabian horned viper reference genome and removed. A third dataset was created (dataset 3) for population genetic analyses keeping only unlinked SNPs through bcftools [67] using a maximum value of $R^2 = 0.5$.

Population structure analyses

We performed a Principal Component Analysis (PCA) with dataset 3 using Plink v1.9 [69]. Visualization of results from these analyses was performed with R v.3.6.3 [70] and the R package ggplot2 [71].

Genomic phylogeny

We reconstructed the phylogenomic relationships of all samples (including the outgroup) using only autosomal chromosomes selected from bam files with the view function from Samtools v1.9 [64]. These bam files were used to generate individual pseudohaploid consensus sequences with ANGSD v.0.933 [72] taking a consensus-based sampling approach (-doFasta 3) in non-overlapping sliding windows of 1 Mbp. Maximum likelihood trees for each window along the Arabian horned viper reference genome were calculated with IQ-TREE2 v.2.4.0 [73] applying a GTR + I+G substitution model, with the Oman saw-scaled viper (*Echis omanensis*) sample as an outgroup. Windows where one individual had > 50% missing data were removed, leaving a total of 1,105 non-overlapping windows. A maximum clade-credibility tree was created with TreeAnnotator v2.6.4. from BEAST2 v2.6.6 [74]. We tested the effect of window size by repeating the analyses for smaller window sizes (500 Kbp) and similar topologies were obtained.

Demographic history

We inferred the demographic history of our samples with the Pairwise Sequential Markovian Coalescent (PSMC v.0.6.5) software [75] for which we used three high-coverage samples (individuals underlined in Table 1). Heterozygous positions were obtained from bam files with Samtools v1.9 and data were filtered for low mapping (< 30) and base quality (< 30). Minimum and maximum depths were set at half and double the average coverage for each sample, respectively. Only autosomal data were considered. A rate of 2.4×10^{-9} substitutions/site/generation and a generation time of 3 years were used, following [42, 76, 77]. Other parameters were set as default. For each sample, 10 bootstraps were calculated. Final results were plotted with the psmc_plot.pl function from PSMC (https://github.com/lh3/psmc).

Gene flow analyses

We ran the ABBA-BABA test, as implemented in Dsuite [78] to test for gene flow among the different species of *Cerastes* following previous research using genome-wide markers [48]. For this analysis, we used the filtered VCF file containing the outgroup (31.10 million SNPs across 27 individuals), with the Oman saw-scaled viper as outgroup. We explored possible introgression between *C. gasperettii* with *C. vipera* and both subspecies of *C. cerastes*, as well as between *C. vipera* with *C. c. cerastes*. Finally, in order to delve into the evolution of venom for this group, we created a unique dataset for each of the three main toxin families (*SVMP*, *SVSP* and PLA_2) and explored if introgression occurred within those regions, using the same methodology as described above. For that, we created a new VCF containing only SNPs present in the genomic region for each of the three main toxin families. To reduce potential biases associated with paralogous sequences and mapping ambiguity in structurally variable regions, only uniquely mapped reads with quality higher than 30 were retained for downstream analyses. In addition, repetitive regions identified in the reference genome were excluded from all datasets used for D-statistic analyses. These filtering steps collectively reduce the inclusion of reads originating from non-unique genomic locations and therefore mitigate, although do not fully remove, the impact of paralogous gene copies in highly duplicated gene families.

Genomic diversity and ROHs

We used dataset 1 to calculate average genome heterozygosity for high-coverage individuals. We generated non-overlapping sliding windows of 100 Kbp for the Arabian horned viper reference genome and took only sites (both variant and invariant) with site quality higher than 30 (QUAL field in a VCF file from GATK). Only windows containing more than 60,000 unfiltered sites were considered. ROHs were calculated per individual based on the density of heterozygous sites in the genome using the implemented Hidden Markov Model (HMM) in bcftools

roh function [79] with –AF-dflt 0.4 (following [80]) on the filtered dataset. The minimum ROH length threshold of 100 Kbp was selected to reduce the detection of spurious short homozygous tracts potentially arising from linkage disequilibrium, genotyping errors, or uneven sequencing coverage, and to ensure robustness given the sequencing depth of the dataset. This threshold is consistent with previous population genomic studies in non-model organisms with comparable coverage [81, 82]. The Phred score threshold of 70 was used to retain only high-confidence ROH calls supported by reliable genotype likelihoods and to minimize false positives in HMM-based inference. We kept ROHs with a Phred Score of at least 70 and with a minimum length of 100 Kbp. Finally, following [81], we split the ROHs per length in short (< 500 Kbp), medium (> 500 Kbp and < 1 Mbp) and large (> 1 Mbp). Visualization for all analyses was carried out with ggplot2 [71].

Identify genomic regions under introgression

Following the approach of [45], we investigated in detail the strongest signal of introgression detected between *C. vipera* and *C. cerastes* by exploring the local window topologies across the genome. To do so, we selected two high-coverage individuals per species and/or subspecies ($n = 7$, as there was only one high-coverage individual was available for *C. c. hoofieni*) and phased the dataset using Beagle v.5.4 [83]. We then calculated Neighbor-Joining trees in non-overlapping sliding windows containing at least 50 SNPs using PhyML v3.3 [84]. This analysis was implemented using the script phyml_sliding_windows.py (https://github.com/simonhmartin/genomics_general). We calculated complete topology weighting with the resulting local trees in TWISST v.18 [85]. Local topologies where the introgressed pattern was fully supported, i.e., (*Echis*, (*C. gasperettii*,(*C. c. hoofieni*,(*C. c. cerastes*, *C. vipera*)))), were highlighted across the genome.

Interspecific venom variation

We analyzed coverage variation in venom-coding genes by selecting nine high-coverage individuals (mean genome-wide coverage > 20x) and comparing the coverage read-depth profiles across intronic and exonic regions for the main toxin families (i.e., SVMPs, SVSPs and PLA_2s). Only uniquely mapped reads with a mapping quality higher than 30 were retained. We converted the mapped files into bedgraph and plotted the coverage levels on the specific genes using Spark v.2.6.1 (https://github.com/harbourlab/SparK). Summary plots were produced with ggplot2 [71] in R [70]. In addition, we used CNVnator v.0.4.1 [86] to identify putative duplications and deletions affecting toxin gene families with a bin size of 100 bp, which allowed us to complement the coverage-based analysis with explicit detection of structural variants. Read depth normalization was performed internally by CNVnator, which estimates copy number from read depth normalized to the genome-wide average coverage after GC correction. For duplications, we retained events overlapping toxin gene regions with a minimum length of 1,000 bp, a normalized read depth (RD) ≥ 1.5, an RD p-value $\leq 1 \times 10^{-5}$, a t-test p-value $\leq 1 \times 10^{-5}$, and a q0 value (fraction of zero-mapping reads, used to filter repetitive regions) ≤ 0.3. Filtering criteria for deletions were identical except for RD < 0.7. The RD p-value assesses the statistical significance of coverage deviations relative to the genomic background, whereas the t-test p-value evaluates the significance of coverage shifts at the predicted CNV boundaries. The genomic location of each CNV was determined by intersecting CNV coordinates with the toxin gene annotations, allowing us to classify events as overlapping exonic, intronic, or intergenic regions.

Toxin phylogenies

We used phylogenetic inference to study the evolutionary history for the main groups of toxins (i.e., SVMPs, PLA_2s, and SVSPs; [42, 56–58]. For the three main toxin families, we selected available toxin genes and their non-toxin paralogs from the consensus sequences generated from uniquely mapped reads of the high-coverage samples in this study. We complemented the dataset with other genes from venomous snake species provided in [42] and containing data from [21, 22, 41]. Following [22], we built a phylogeny for each of the toxin groups

using PhyML v3.3 [84], implementing the Dayhoff substitution model and validating our inferred tree with aBayes support.

Results

Population structure

We generated whole genome resequencing data for a total of 26 individuals representing the three studied *Cerastes* species, along with one outgroup. One individual (A17272) was excluded due to low coverage. The retained samples exhibited coverage levels ranging from 6.33x to 78.21x (Table 1). The final dataset included representations of all currently recognized taxa within the genus. With this genomic dataset, we investigated the population structure of the genus *Cerastes*, using a PCA with a total of 1.36 million polymorphic positions, after LD pruning (Fig. 1A). PC1, explaining 17% of the total genomic variation, separated *C. gasperettii* from *C. cerastes* and *C. vipera* whilst PC2, with a total of 10.4% of the total variation, split *C. cerastes* from *C. vipera*. These axes likely capture genome-wide patterns of divergence among species, potentially reflecting a combination of phylogenetic history and accumulated genetic differences across multiple loci rather than specific gene families. Interestingly, *C. cerastes* individuals from Western Sahara clustered closer to *C. vipera* (consistent with introgression results, see below).

Genomic phylogeny

To explore the evolutionary relationships of *Cerastes* from a genomic perspective, we inferred the phylogenomic relationships of the genus using a total of 1,105 non-overlapping 1-Mbp sliding windows, covering around 70%

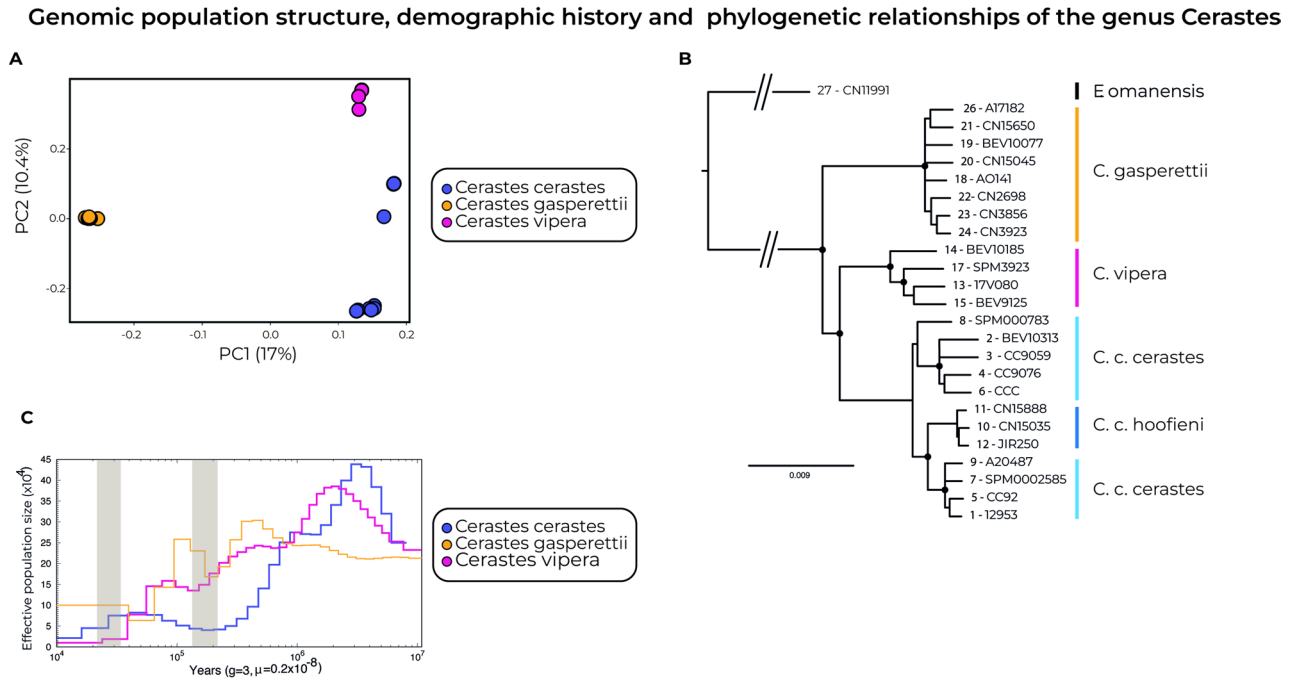

Fig. 1 **A** Genomic PCA for the three species within the genus *Cerastes*, using more than 1.36 M SNPs. **B** Genomic phylogeny for the genus *Cerastes*, using *Echis omanensis* as an outgroup. Branches with support values higher than 85 are indicated by black circles. **C** Demographic history for the three species of *Cerastes*. Vertical grey lines are indicative of the Last glacial and Penultimate glacial periods

of the whole genome. Our resulting phylogenomic tree showed three well-supported clades, one for each species. Results did not show any geographical clustering for *C. gasperettii* or *C. vipera*. However, we report three different clusters for *C. cerastes* (Fig. 1B). Cluster one contained individuals from North-Western Africa (*C. c. cerastes*), cluster two contained individuals from North-Eastern Africa, including the two samples from the Sinai Peninsula (*C. c. cerastes*) and the third cluster contained the Arabian populations (*C. c. hoofieni*). Some intraspecific polytomies were present, especially within *C. gasperettii*, but most were resolved when using a lower sliding window threshold (Fig. S2).

Demographic history

When studying the changes in effective population size across time for all three species, we found a similar trend between *C. cerastes* and *C. vipera* (Fig. 1C), as we recovered high levels of effective population size in the past (>1 Million years ago (Mya)) with a constant decline towards the present. On the other hand, for *C. gasperettii* the demographic history remained relatively constant between 10 and 1 Mya, followed by several population expansions and contractions in the last 400 Kya (Fig. 1C).

Introgression

Introgression analyses revealed gene flow between *C. c. cerastes* and *C. vipera* (Fig. 2A). Additionally, our whole-genome resequencing data indicated introgression between the Arabian subspecies *C. c. hoofieni* and the Arabian *C. gasperettii*, as well as between the latter and the ancestral *C. cerastes* lineage (Fig. 2A), However, only the signal involving *C. c. cerastes* and *C. vipera* was associated with a substantial f4 ratio estimate (around 0.053 with *C. c. cerastes* as P2 and *C. vipera* as P3), whereas the remaining comparisons yielded values close to zero, indicating either very limited introgressed ancestry or signals that are statistically detectable but quantitatively negligible (Table S1). When introgression was assessed after splitting *C. c. cerastes* into geographic groups, a larger number of significant comparisons emerged (Fig. S3). However, as in the species level analysis, the only comparison with an f4-ratio exceeding 1% involved *C. vipera* and the western populations of *C. c. cerastes* (Table S2). An additional significant signal was detected between the East African populations of *C. c. cerastes* and *C. c. hoofieni*, but this pattern is best interpreted as a sampling artifact, since *C. c. hoofieni* originates from East Africa and is therefore expected to share more alleles with those populations than with western ones in the absence of recent introgression. Interestingly, we did not identify signals of introgression when focusing specifically on the venom-coding genes, except for the PLA_2 gene family, where introgression was observed between *C. c. hoofieni* and *C. vipera* with around 40% of the sites being introgressed (Fig. 2 and Tables S3-5). Although these two taxa are allopatric and genome-wide analyses did not support gene flow between them, the PLA_2 signal is consistent with similar copy number variation (CNV) observed for this gene in both species (see below). When investigating the specific genomic regions exhibiting signals of introgression between *C. cerastes* and *C. vipera*, we observed sparse and scattered patterns of introgression across the genome (Fig. S4).

Genomic diversity and inbreeding

Overall, we observed relatively high levels of genome-wide genetic diversity across the three *Cerastes* species, especially when compared to reported genomic diversity levels in other venomous snake species (Fig. 3A). However, patterns of genetic diversity differed depending on the level of comparison (individual versus species-level summaries). At the intraspecific level, *Cerastes cerastes* exhibited substantial variation among the individuals. One individual from Mauritania (code CCC) showed relatively high levels of genomic diversity, comparable to other congenerics. However, other two individuals showed lower diversity levels, the specimen from Chad (12953) showed intermediate diversity, and the Arabian individual (CN15035) exhibited the lowest genomic diversity recorded across the entire genus. In contrast, *Cerastes vipera* showed relatively little intraspecific

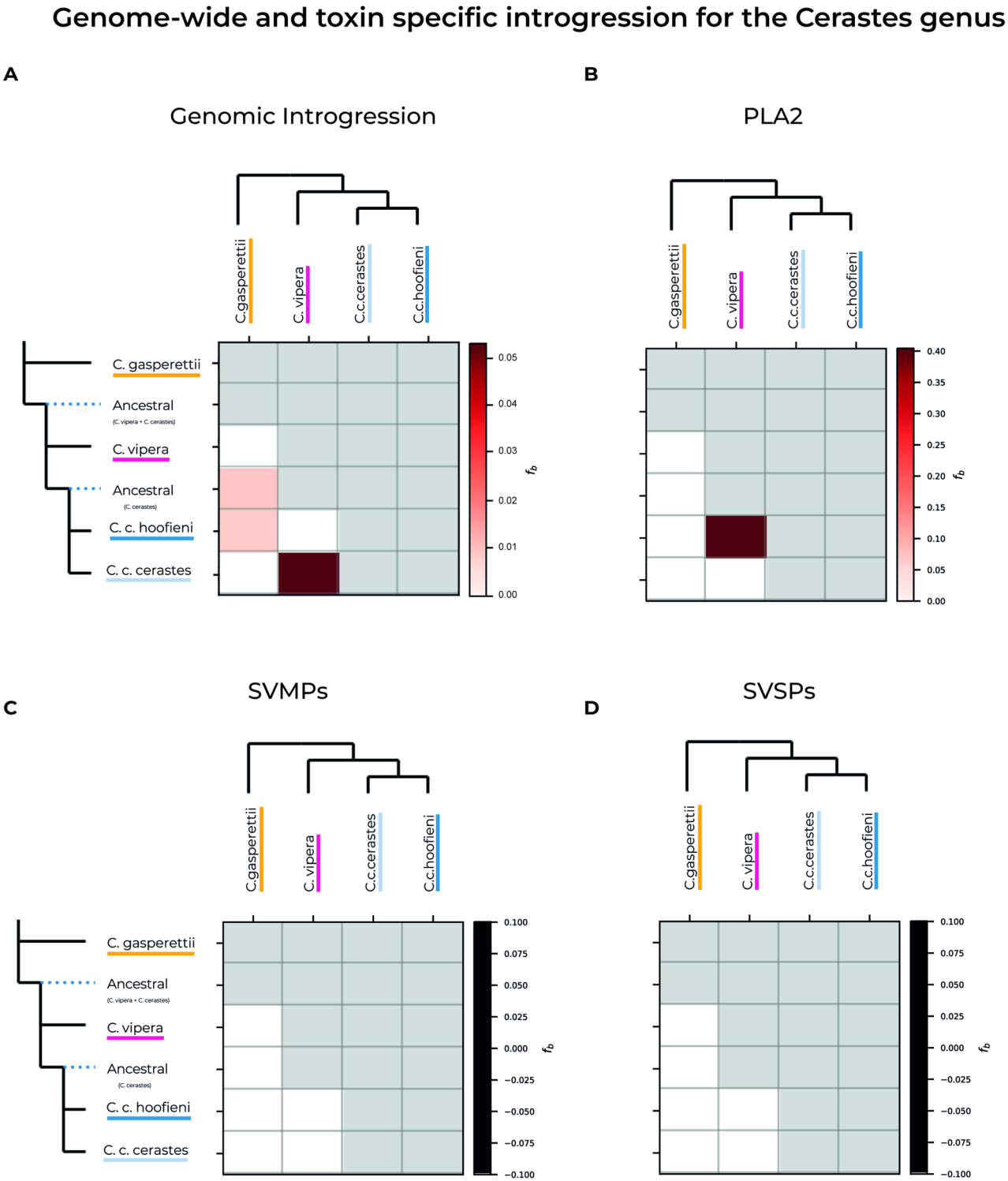

Fig. 2 Branch-based excess of introgression (fb) inferred by Dsuite. Excess introgression is shown at the (**A**) genome-wide scale and (**B**-**D**) for toxin-gene regions containing PLA_2, SVMPs and SVSPs. Values and legend colors represent the branch-specific fb statistic, summarizing excess introgression associated with each phylogenetic branch. Although not shown, *Echis omanensis* was used as an outgroup. Note that legend may vary between figures

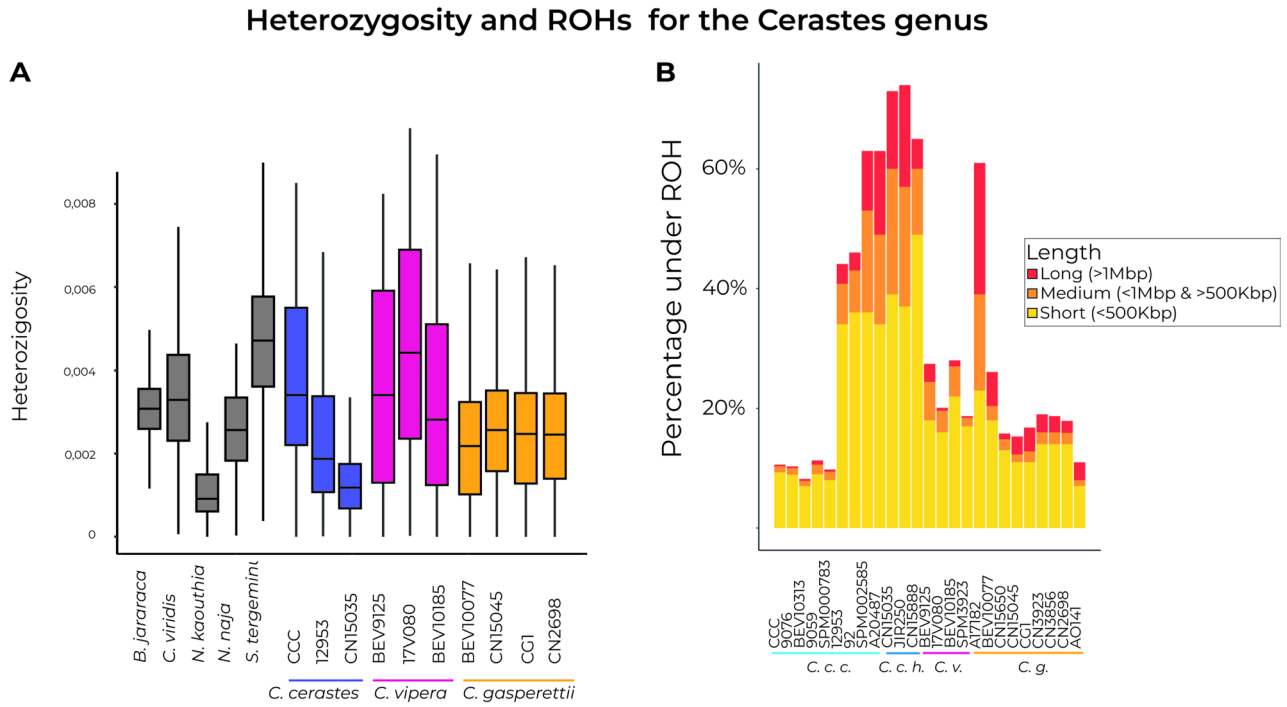

Fig. 3 **A** Genome-wide heterozygosity for several venomous snake species (in grey), *C. cerastes* (in purple), *C. vipera* (in pink) and *C. gasperettii* (in yellow). **B** Percentage of the genome under ROHs for the three studied species, including subspecies when possible. ROHs longer than 1 Mbp are in red, ROHs between 1 Mbp and 500 Kbp are in orange and ROHs smaller than 500 Kbp are in yellow. *C.c.c.* stand for *Cerastes cerastes cerastes*, *C.c.h.* for *Cerastes cerastes hoofieni*, *C.v.* for *Cerastes vipera* and *C.g.* for *Cerastes gasperettii*

variation between individuals, but higher average genomic diversity across sampled individuals compared to the other species. *Cerastes gasperettii* displayed intermediate and relatively homogeneous levels of genomic diversity among individuals. ROHs varied considerably among species and geographic groups. Populations of *C. c. cerastes* from Western Africa exhibited the lowest levels of ROHs, which were predominantly short ROHs (Fig. 3B). In contrast, two individuals from Northeastern Africa had around 40% of their genomes composed of short ROHs, suggesting a historical population bottleneck. A similar signal was detected in *C. c. hoofieni*, although with a higher proportion of medium and long ROHs, consistent with possible recent inbreeding due to geographic isolation. Individuals of *C. vipera* showed low to moderate levels of ROHs, with no evident geographic variation among samples, while *C. gasperettii* individuals shared a similar ROHs profile regardless of geographic origin, with the exception of a single individual from the Sinai Peninsula.

Interspecific venom variation

To explore structural variation in the toxin-encoding genes for the main venom families, we compared the different coverage levels within the main toxin regions for the three species in order to identify duplication or deletion events (Fig. 4 and Tables S6-8 and S10-11). Coverage levels for PLA_2 showed an important drop for PLA_2-*gC* within *C. vipera* (Fig. 4). Interestingly, this drop of coverage was also observed in *C. c. hoofieni* but it was absent from *C. c. cerastes* (Fig. 4 and Table S6). A region between PLA_2-*gC* and PLA_2*g-2 F* also showed drops of coverage in both *C. vipera* and *C. cerastes*, but we could not identify any specific gene or transposon element. For *SVMPs*, we observed constant coverage levels for almost all *SVMPs* (mainly 2, 8, 9, 10, 11, 12, 13 and including their non-toxic paralog *ADAM28*). However, there was a mild drop of coverage in *SVMP1* in *C. c. hoofieni* when compared to *C. c. cerastes* and the rest of the studied species. There was another drop of coverage for *SVMPs*

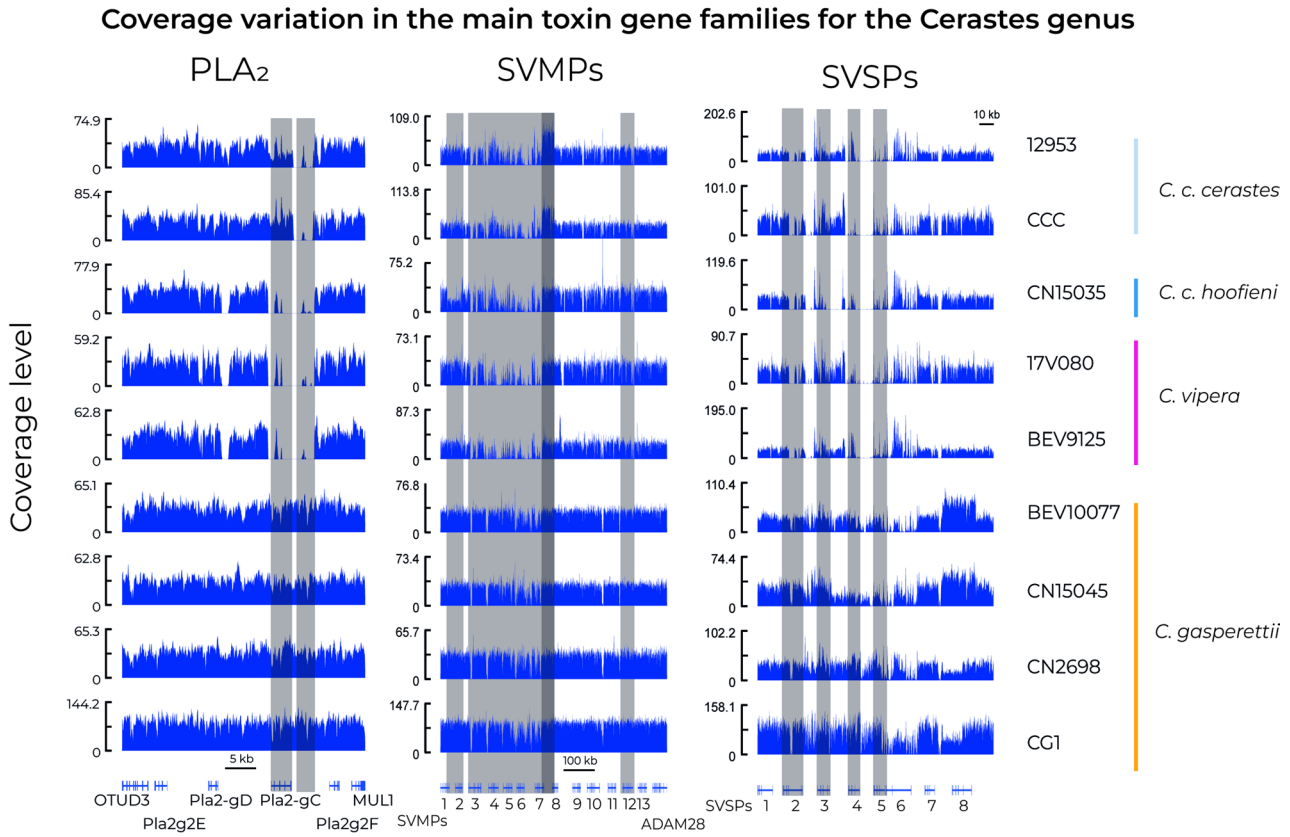

Fig. 4 Coverage variation for the exonic and intronic regions for the main toxin families (PLA_2, SVMPs and SVSPs from left to right) within the genus *Cerastes*. PLA_2: Coverage variation for PLA_2, including two flanking genes (OTUD3 and MUL1), two non-toxin paralogs (PLA_2-g2E and PLA_2-g2F) and the two toxin-genes identified in *C. gasperettii* (PLA_2-gD and PLA_2-gC). SVMPs: Coverage variation for SVMPs, indicating all the 13 different genes identified in the genome of *C. gasperettii*. Putative gene duplication is shown in darker gray. SVSPs: Coverage variation for SVSPs, containing the eight different genes identified in the genome of *C. gasperettii*. For the three toxin families, vertical gray lines show main differences between the studied species. Individual and species labels are indicated at the right of the figure

3, 4, 5, 6 and 7 in both *C. cerastes* and *C. vipera* compared to *C. gasperettii* (Fig. 4 and Table S7). Interestingly, we noticed average coverage levels for all the studied species in *SVMP12*, a new gene recently found in *C. gasperettii* [42]. We observed an increase of coverage specific to C. *c. cerastes* (not present in *C. c. hoofieni*) between *SVMP7* and *SVMP8* (Fig. 4 and Table S8). Blast search identified it as a Bov-B LINE retrotransposon, and although the query coverage was low, the percentage of identity was high (Table S9). Overall, intraspecific coverage differences were not found for *C. gasperettii*. Finally, *SVSPs* showed lower coverage levels in *SVSP1*, *SVSP2*, *SVSP3* and *SVSP4* of *C. cerastes* and *C. vipera* when compared with *C. gasperettii* (Fig. 4). Interestingly, there seemed to be an increase of coverage for the region between *SVSP7* and *SVSP8* (unidentified by blast), as well as within *SVSP8* for some individuals of *C. gasperettii* (Fig. 4). To complement the coverage-based analyses, we performed read-depth CNV detection using CNVnator across all individuals. CNVnator identified putative copy-number events overlapping toxin gene regions, including both duplications and deletions affecting intronic and exonic segments (Tables S6-8 and S10-11). Within the SVMP cluster, duplications were detected only in *C. c. cerastes*, whereas deletions were detected in all studied species and in multiple individuals. For the SVSP gene family, a single duplication event was detected in *C. gasperettii*, while deletions were identified in *C. c. cerastes*, *C. gasperettii*, *C. vipera*, and *C. c. hoofieni*. In the PLA_2 cluster, CNVnator detected deletions in all species, including multiple individuals per taxon. However, these putative CNVs were consistent with the observed

read-depth patterns, although we acknowledge that complex toxin loci may still be affected by mapping ambiguity and collapsed paralogous regions.

Toxin phylogenies

We obtained similar results for the main toxin groups (i.e., *SVMPs*, *SVSPs* and PLA_2) as was previously reported for *C. gasperettii* [42]. Within PLA_2, we reported as outgroups a clade containing both non-toxic PLA_2-*g2E* and PLA_2-*g2F* (Fig. S5). We also found monophyly for all other PLA_2 groups reported in previous studies: PLA_2-*gC*, PLA_2-*gB*, PLA_2-*gK*, PLA_2-*gD* and PLA_2-*gA* [21, 42, 87]. For most of our samples, we reported the two groups previously found in *C. gasperettii*: PLA_2-*gD* and PLA_2-*gC*. However, for *C. vipera* and *C. c. hoofieni*, the putative PLA_2-*gC* genes were not found within the other PLA_2-*gC*. Lastly, the putative PLA_2-*gD* gene for one *C. cerastes* sample (12953) clustered within the PLA_2-*gC* group. Within *SVMPs*, we reported a supported clade comprising *ADAM28* peptides, the non-toxic paralogous gene for this toxin family (Fig. S6). We also found a second clade of orthologous toxins present both in elapid and viperid families. Interestingly, *SVMP12* gene (a gene uniquely found so far in *C. gasperettii*) was found in all individuals, with the sample of *E. omanensis* found as an external group. Finally, our phylogenetic inference analysis for *SVSPs* showed two groups, both containing *Cerastes* genes (Fig. S7). The most basal group mostly contained Crotalinae genes, although there was the presence of some Viperinae species and two of our studied genes (*SVSP1* and *SVSP7*) (Fig. S7). The second group was highly expanded in true vipers and most of our studied genes were located there, although support was low.

Discussion

In this study, we used whole-genome sequencing data for 27 samples to investigate the different evolutionary histories of the three species of the genus *Cerastes*. This group of arid-adapted Palearctic vipers, distributed across North Africa and Arabia, represents a valuable model to study diversification in desert environments and across a transcontinental range, while also being of medical relevance. Our work represents one of the first to use whole-genome data within the true vipers (see also [45]), a group of medically important snakes such as the genera *Echis*, *Daboia* or *Bitis* for which genomic data have historically been limited or lacking. Our results are consistent with previous population structure findings using genome-wide data and further support a revised phylogenomic relationship for the genus *Cerastes* (see [48]). Moreover, we applied standard conservation genomic analyses to characterize patterns of genetic variation, revealing the lowest genome-wide heterozygosity and high ROHs in the geographically isolated Arabian populations of *C. c. hoofieni*. Finally, we investigated the evolutionary history of venom for the three studied species as well as the differences in structural variation within the main toxin gene families, suggesting that genomic variation might be promoting phenotypic changes. Overall, our study highlights the value of genomic data for investigating patterns of venom evolution, providing a foundation for future work on the underlying mechanisms.

Evolutionary history of the genus *Cerastes*

Cerastes gasperettii showed marked genomic differentiation from both *C. cerastes* and *C. vipera* (Fig. 1A), consistent with findings based on reduced-representation genome-wide data [48]. Notably, fine-scale genomic data employed here identified some *C. cerastes* samples from Western Africa closer to *C. vipera*, in agreement with evidence of introgression (see below). Whole-genome phylogenomic analyses further supported *C. gasperettii* as the external clade within the genus (Fig. 1B), also revealing low differentiation within *C. gasperettii*. In line with mitochondrial evidence [88], we detected a high genomic similarity between *C. g. gasperettii* and the putative subspecies *C. g. mendelssohni* (sample A17182). Within *C. cerastes*, we identified three different lineages, as previously described [48]. Two individuals from the Sinai Peninsula (samples A20487 and SPM0002585) grouped

with the Eastern African samples of *C. cerastes*, while *C. c. hoofieni* remained a distinct, geographically isolated taxon confined to the Tihama desert in southwestern Arabia.

Demographic reconstructions based on PSMC analyses suggested recent population declines in all three *Cerastes* species (Fig. 1C), although these inferences are derived from a limited number of high-coverage individuals and should be interpreted as indicative of long-term trends rather than precise species-wide histories. The population expansion observed in *C. gasperettii* may be consistent with favorable conditions during the last glacial periods [42], whereas the inferred steady decline in *C. cerastes* and *C. vipera* may be attributed to long-term reductions in climatic suitability (for *C. cerastes*, see [53]). Interestingly, a broad convergence in effective population size between *C. cerastes* and *C. vipera* around 7 Mya aligns with previous speciation estimates (Fig. 1C; [48]). We also detected ancestral introgression between *C. cerastes* and *C. vipera* (Fig. 2A). A weaker signal of gene flow was observed between the two Arabian taxa, *C. c. hoofieni* and *C. gasperettii* (Fig. 2A), despite their current allopatric distribution. While these two taxa are separated by the Asir Mountains of Saudi Arabia and Yemen [55], sporadic dispersal events by *C. gasperettii* towards the Tihama region may have facilitated occasional introgression. With WGS data we detected more signals of introgression than previously detected [48], highlighting the intricate evolutionary history of this genus.

Analysis of venom-coding genes revealed limited evidence of introgression across the genus, with the exception of PLA_2 gene family. Here, we found introgression between *C. vipera* and *C. c. hoofieni*, consistent with the reduced coverage for the coding-gene PLA_2-*gC* in taxa groups (Figs. 2B and 4). As both species are allopatric, this pattern should be interpreted cautiously and may alternatively reflect convergent evolution, retained ancestral variation, incomplete lineage sorting or mapping artifacts rather than ongoing gene flow. Although stringent filtering of uniquely mapped reads and repetitive regions reduces the influence of paralogous sequences, gene families such as PLA_2 remain challenging in short-read based analyses due to their frequent duplication and high sequence similarity among copies. Consequently, introgression signals detected in toxin-associated regions should be interpreted with caution, as they may reflect a combination of true historical gene flow and residual mapping ambiguity across duplicated loci. Further venom characterization and long-read data in *C. c. hoofieni* is needed to determine whether genomic similarities are phenotypically expressed. Using the *C. gasperettii* reference genome, we mapped the specific regions introgressed between *C. cerastes* and *C. vipera*, identifying a scattered and episodic pattern of introgression through the genome (Fig. S4), suggestive of ancient hybridization events. Altogether, our whole-genome resequencing data reinforce prior findings based on ddRAD-seq [48], supporting the utility of genome-wide data for inferring evolutionary relationships and population structure in non-model organisms, particularly in the absence of a reference genome.

Although our whole-genome dataset includes 27 *Cerastes* individuals, several analyses, particularly those requiring high-coverage genomes (e.g., demographic inference, runs of homozygosity, heterozygosity estimates, and toxin-gene copy number variation), were conducted on smaller subsets of samples. In addition, some taxa and geographic lineages, such as *C. c. hoofieni*, are represented by relatively few individuals. Therefore, these results should be interpreted with caution, as uneven sampling and differences in sequencing coverage may influence the accuracy and generality of the inferred demographic patterns and genomic variation.

Conservation genomics of horned and sand vipers

Standard conservation genomic analyses (i.e., genome-wide heterozygosity and ROHs) revealed notable differences among the three *Cerastes* species (Fig. 3). In particular, the Arabian representative of *C. c. hoofieni* showed the lowest genome-wide heterozygosity recorded across the whole genus, with elevated ROHs. This pattern suggests that the Arabian population has experienced prolonged periods of limited gene flow, likely due to geographic and ecological barriers, which has led to the accumulation of homozygosity over time. Such isolation may have important implications for their adaptive potential and conservation, as reduced genetic diversity could increase susceptibility to environmental change and stochastic events. Both *C. vipera* and *C. gasperettii* did not show high intraspecific variation in relation to heterozygosity and ROHs. However, one *C. gasperettii* sample

(A17182) from the Sinai Peninsula showed around 60% of the genome under ROHs, comparable to *C. cerastes* individuals from the same region (SPM0002585 and A20487). The Sinai Peninsula represents the geographical convergence zone of all three *Cerastes* species and is characterized by restricted accessibility from both Africa and Arabia. This geographic positioning likely contributes to elevated isolation and limited gene flow. The high levels of ROHs in *C. gasperettii* from Sinai may reflect not only geographic isolation but also possible anthropogenic barriers. Recent studies have demonstrated the impact of urban areas on species inbreeding, which can increase rapidly under strong anthropogenic pressures (e.g [89]). Notably, the *C. gasperettii* individual sampled from Sinai was located near the coast and possibly close to urbanized areas, which could further hamper dispersal. Unfortunately, the *C. vipera* sample collected in the Sinai Peninsula was unsuccessfully sequenced, but we encourage future studies to focus on the role of this peninsula in the isolation and inbreeding of this *C. vipera* population as well as how recent urban development may affect gene flow and long-term viability in these peripheral populations.

Venom evolution

Using whole-genome resequencing data, we investigated the evolutionary history and structural variation of the main toxin families in the genus *Cerastes* (Fig. 4 and Fig. S5-7), supported by available phenotypic venom data from previous proteomic studies [42, 56–59]. Geographic variation in venom composition has only been described in *C. cerastes*, with significant variation mainly in SVMPs, SVSPs, CTLs and LAAOs (see [57]). Proteomic data are scarce for the other two species. To our knowledge, the only study providing proteomic data for *C. vipera* [56] indicated a simpler venom composition compared to *C. cerastes*, with differences mainly in SVMPs and PLA_2. Notably, two different PLA_2 toxin peptides were present within *C. cerastes*, whereas no such variation was observed in *C. vipera*. For *C. gasperettii*, no data are available regarding geographic or ontogenetic variation, although one venom proteome has been published [42], and other functional properties of its venom have been studied [90, 91].

We observed a significant reduction in coverage levels of the PLA_2-gC gene in *C. vipera* and *C. c. hoofieni* (Fig. 4 and Table S6), located in both CDS and intergenic regions. Phylogenetic analyses also supported this result, as PLA_2-gC sequences from these two taxa did not show homology with the rest of the clade (Fig. S5) and exhibited high levels of missing data. Previous proteomic results from [56] already reported two distinct PLA_2 sequences in *C. cerastes*, but not in *C. vipera*, suggesting that structural variation may contribute to this phenotypic variation. Given that PLA_2 toxins are major venom components involved in diverse biological activities, including membrane disruption and prey immobilization [92], changes in PLA_2 gene content may potentially influence venom phenotype. Interestingly, some intraspecific phenotypic venom variation has also been reported in *C. c. cerastes* [93], suggesting that the absence of one of the two PLA_2 peptides seems a variable trait within both species. Future genomic and proteomic data, especially from *C. c. hoofieni*, may help to disentangle the origin of this phenotypic variation.

For SVMPs, the toxins from all three *Cerastes* species clustered together with previously identified SVMPs from the *C. gasperettii* genome (Fig. S6; [42]). However, we did not find toxin groups clustering with class P-I SVMPs, even though proteomic data report their presence in *C. cerastes* [58]. On the other hand, class P-II SVMPs were found in *C. gasperettii* but have not been detected through proteomics in the other two species. These differences could be explained by the variation in coverage for the SVMP genes 3, 4, 5, 6 and 7, as we found differences in coverage levels for *C. cerastes* and *C. vipera* (Fig. 4 and Tables S7 and S8). SVMPs are the most common toxin family for both *C. cerastes* and *C. vipera* and this toxin family is known for fragmenting themselves and creating new genes [22]. Interspecific variation within *Cerastes* could have originated from a putative fragmentation of some of the aforementioned genes. Interestingly, coverage data for the SVMP12 gene was found within the three species of the genus *Cerastes*, suggesting that this new gene is present in all *Cerastes* species. Moreover, we observed a duplication event within two SVMP genes in *C. cerastes*, where a Bov-B LINE retrotransposon was identified (Table S8-9). These types of retrotransposons have been studied within snakes

and have been suggested as promoters of gene copy number variation [94]. The presence of this retrotransposon in proximity to SVMP genes suggests a potential mechanism for generating structural variation in toxin genes, which could contribute to functional diversification within the venom repertoire. Such duplications may facilitate the rapid evolution of venom components, and highlight the dynamic role of mobile genetic elements in shaping snake genomes. However, with our data we cannot establish any direct mechanistic involvement in toxin diversification.

Finally, we did not recover new groups for SVSPs when studying the evolutionary history of *C. cerastes* and *C. vipera* (Fig. S7), as all genes were orthologous between *Cerastes* species. Conversely, we did observe significant coverage differences in toxin-coding SVSP genes 2, 3, 4 and 5, with both *C. cerastes* and *C. vipera* presenting low or even absent coverage levels (Fig. 4 and Tables S10 and S11). These findings are consistent with proteomic studies, as SVSPs are significant components of the venom for *C. cerastes* and *C. vipera* [57] but they are not the main component as in *C. gasperettii* [42], in line with a putative decrease in the number of gene copies.

Venom systems evolve through multiple interacting mechanisms, including transcriptional regulation, ontogenetic expression shifts, epigenetic modulation, post-transcriptional processes, ecological selection, and stochastic variation [11, 27, 31, 32, 35, 36, 41]. Although our study does not include gene regulatory information (i.e., expression data as RNA-seq or epigenetics), our CNV results are consistent with previously reported venom differences and may represent candidate genomic mechanisms underlying phenotypic variation in venom composition within *Cerastes*. However, the use of a single reference genome (*C. gasperettii*) may introduce reference bias, particularly in more divergent lineages such as *C. vipera*, which could affect the detection and interpretation of structural variation in highly repetitive toxin gene families. In particular, reduced coverage in toxin loci may reflect mapping divergence, paralog collapse, repetitive sequence structure, or assembly limitations of the reference genome, rather than true gene absence or copy-number change.

The development of high-quality reference genomes for *C. cerastes* and *C. vipera*, together with additional transcriptomic or proteomic data, will be critical to fully understand the genomic basis of venom evolution within this genus.

Conclusions

Whole-genome sequencing data for a total of 26 individuals confirmed the previously reported phylogenomic relationships within the genus *Cerastes* using genome-wide data. Our analyses revealed evidence of ancient introgression between *C. c. cerastes* and *C. vipera*, identifying the genomic regions where such introgression persists. Additionally, we characterized the genomic status of the three species within the genus, highlighting a low genome-wide heterozygosity and high ROHs for the isolated Arabian populations of *C. c. hoofieni*. We also investigated interspecific structural variation in the main toxin gene families within the genus *Cerastes* and integrated these genomic patterns with their evolutionary history of the genus and existing proteomic data. Our whole-genome data suggests that structural variation in toxin genes is widespread across *Cerastes* species and may contribute to differences in venom composition. Altogether, this study demonstrates the utility of whole-genome data to disentangle evolutionary, demographic, and functional processes in non-model venomous snakes. We encourage future research to further explore venom evolution and population history using integrated genomic, transcriptomic and proteomic approaches.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-13357-8.

Acknowledgements FM-F acknowledge CIBIO-BIODESERTS members during fieldwork campaigns. We also want to acknowledge Carlos de la Calle for his helpful insights in graphic visualization as well as four anonymous reviewers for their helpful feedback.

Authors' contributions G.M.-R. conceived the study, performed the data analyses, and wrote the manuscript. S.C., M.J.M., A.T., and B.B.-C. contributed to the conceptualization and study design. A.T. and B.B.-C. also contributed to data curation and methodological development. J.E., M.S., S.B., M.E., J.Š., and F.M.-F. contributed to sample collection, data generation, and/or provided key resources. All authors contributed to the interpretation of results, critically revised the manuscript, and approved the final version.

Funding GM-R was supported by an FPI grant from the Ministerio de Ciencia, Innovación y Universidades, Spain (PRE2019-088729), BB-C is supported by FPU grant from Ministerio de Ciencia, Innovación y Universidades, Spain (FPU18/04742), AT is supported by "la Caixa" doctoral fellowship program (LCF/BQ/DR20/11790007). ME is funded by an FPI fellowship from the *Ministerio de Ciencia*, *Innovación y Universidades*, Spain (PRE2022-101473). This work was funded by grants PID2021-128901NB-I00 (MCIN/AEI/10.13039/501100011033 and by ERDF, A way of making Europe), Spain, and grant 2021-SGR-00420 from the Departament de Recerca i Universitats from the Generalitat de Catalunya, Spain to SC. JŠ was supported by the Czech Science Foundation (GAČR) under grant number 25–17736 S.

Data availability The data generated in this study is available in NCBI under the Bioproject PRJNA1416256. Reviewer's link can be available here: (https://dataview.ncbi.nlm.nih.gov/object/PRJNA1416256?reviewer=6ushs63vovvenu8c6pjmb3v7fi).

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Bragg JG, Supple MA, Andrew RL, Borevitz JO. Genomic variation across landscapes: insights and applications. New Phytol. 2015;207(4):953–967. https://doi.org/10.1111/nph.13410.
2. Wood ZT et al. Sep., Meta-analysis: congruence of genomic and phenotypic differentiation across diverse natural study systems. Evol Appl. 2021;14(9):2189–2205. https://doi.org/10.1111/eva.13264.
3. Pierce BA. Genetics: a conceptual approach, Seventh edition. New York, NY: Macmillan International, Higher Education. 2020.
4. Snustad DP, Simmons MJ. Principles of genetics. 6th ed. Hoboken, NJ: Wiley; 2012.
5. Amazonas DR, et al. Molecular mechanisms underlying intraspecific variation in snake venom. J Proteom. Jun. 2018;181:60–72. https://doi.org/10.1016/j.jprot.2018.03.032.
6. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003;33(3):245–254. https://doi.org/10.1038/ng1089.
7. Lewis JJ, Reed RD. Genome-wide regulatory adaptation shapes population-level genomic landscapes in *Heliconius*. Mol Biol Evol. 2019;36(1):159–73. https://doi.org/10.1093/molbev/msy209.
8. Margres MJ et al. The tiger rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. Proc Natl Acad Sci. 2021;118(4):e2014634118. https://doi.org/10.1073/pnas.2014634118.
9. O'Brien RM, Granner DK. Regulation of gene expression by insulin. Physiol Rev 1996;76(4):1109–1161. https://doi.org/10.1152/physrev.1996.76.4.1109.

10. Rokyta DR, Margres MJ, Calvin K. Post-transcriptional mechanisms contribute little to phenotypic variation in snake venoms. G3 GenesGenomesGenetics. 2015;5(11):2375–2382. https://doi.org/10.1534/g3.115.020578.
11. Zancolli G, Casewell NR. Venom systems as models for studying the origin and regulation of evolutionary novelties. Mol Biol Evol. 2020;37(10):2777–2790. https://doi.org/10.1093/molbev/msaa133.
12. Chiang C, et al. The impact of structural variation on human gene expression. Nat Genet. 2017;49(5):692–9. https://doi.org/10.1038/ng.3834.
13. Faria R et al. Mar., Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. Mol Ecol. 2019;28(6):1375–13930. https://doi.org/10.1111/mec.14972.
14. Weischenfeldt J, Symmons O, Spitz F, Korbel JO. Phenotypic impact of genomic structural variation: insights from and for human disease. Nat Rev Genet. 2013;14(2):125–138. https://doi.org/10.1038/nrg3373.
15. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2006;7(2):85–97. https://doi.org/10.1038/nrg1767.
16. Carvalho CMB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet. 2016;17(4):224–38. https://doi.org/10.1038/nrg.2015.25.
17. Currall BB, Chiangmai C, Talkowski ME, Morton CC. Mechanisms for structural variation in the human genome. Curr Genet Med Rep. 2013;1(2):81–90. https://doi.org/10.1007/s40142-013-0012-8.
18. Mérot C, Oomen RA, Tigano A, Wellenreuther M. A Roadmap for understanding the evolutionary significance of structural genomic variation. Trends Ecol Evol. 2020;35(7):561–572. https://doi.org/10.1016/j.tree.2020.03.002.
19. Zhang L, Reifová R, Halenková Z, Gompert Z. How important are structural variants for speciation? Genes. 2021. https://doi.org/10.3390/genes12071084.
20. Wellenreuther M, Mérot C, Berdan EL, Bernatchez L. Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification. Mol Ecol. 2019;28:1203–1209. https://doi.org/10.1111/mec.15066.
21. Dowell NL, Giorgianni MW, Kassner VA, Selegue JE, Sanchez EE, Carroll SB. 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.
22. Giorgianni MW, Dowell NL, Griffin S, Kassner VA, Selegue JE, Carroll SB. The origin and diversification of a novel protein family in venomous snakes. Proc Natl Acad Sci. 2020;117(20):10911–10920. https://doi.org/10.1073/pnas.1920011117.
23. Posadas-García YS, Espinosa-Soto C. Early effects of gene duplication on the robustness and phenotypic variability of gene regulatory networks. BMC Bioinformatics. 2022;23(1):509. https://doi.org/10.1186/s12859-022-05067-1.
24. Smith CF, et al. The best of both worlds? Rattlesnake hybrid zones generate complex combinations of divergent venom phenotypes that retain high toxicity. Biochimie Jul. 2023. https://doi.org/10.1016/j.biochi.2023.07.008.
25. Yang M, Ohnuki S, Ohya Y. Unveiling nonessential gene deletions that confer significant morphological phenotypes beyond natural yeast strains. BMC Genomics. 2014;15(1):932. https://doi.org/10.1186/1471-2164-15-932.
26. Birchler JA, Yang H. The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. Plant Cell. 2022;34(7):2466–2474. https://doi.org/10.1093/plcell/koac076.
27. Margres MJ, Bigelow AT, Lemmon EM, Lemmon AR, Rokyta DR. Selection to increase expression, not sequence diversity, precedes gene family origin and expansion in rattlesnake venom. Genetics. 2017;206(3):1569–1580. https://doi.org/10.1534/genetics.117.202655.
28. Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. Nat Rev Genet. 2013;14(11):807–820. https://doi.org/10.1038/nrg3522.
29. Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG. Complex cocktails: the evolutionary novelty of venoms. Trends Ecol Evol. 2013;28(4):219–229. https://doi.org/10.1016/j.tree.2012.10.020.
30. Fry BG, et al. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. Annu Rev Genomics Hum Genet. 2009;10:483–511. https://doi.org/10.1146/annurev.genom.9.081307.164356.
31. Gopalan SS, et al. Diverse gene regulatory mechanisms alter rattlesnake venom gene expression at fine evolutionary scales. Genome Biol Evol. 2024;16:evae110. https://doi.org/10.1093/gbe/evae110.
32. Schield DR et al. Apr., 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.
33. Jenner RA, Casewell NR, Undheim EAB. What is animal venom? Rethinking a manipulative weapon. Trends Ecol Evol. 2025;40(9):852–61. https://doi.org/10.1016/j.tree.2025.05.009.
34. Williams DJ, 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.
35. Reyes-Velasco J, et al. Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom. Mol Biol Evol. 2015;32(1):173–83. https://doi.org/10.1093/molbev/msu294.
36. Casewell NR, Jackson TNW, Laustsen AH, Sunagar K. Causes and consequences of snake venom variation. Trends Pharmacol Sci. 2020;41(8):570–581. https://doi.org/10.1016/j.tips.2020.05.006.
37. Nachtigall PG, et al. A Segregating structural variant defines novel venom phenotypes in the eastern diamondback rattlesnake. Mol Biol Evol. 2025;42(4):msaf058. https://doi.org/10.1093/molbev/msaf058.

38. Allendorf FW, Funk WC, Aitken SN, Byrne M, Luikart G, Antunes A. Conservation and the genomics of populations. Third ed. New York, NY: Oxford University Press; 2022.
39. Rajora OP, editor. Population genomics: concepts, approaches and applications. in Population genomics. Cham: Springer. 2019.
40. Almeida DD et al. May., Tracking the recruitment and evolution of snake toxins using the evolutionary context provided by the *Bothrops jararaca* genome. Proc Natl Acad Sci. 2021;118(20):e2015159118. https://doi.org/10.1073/pnas.2015159118.
41. Hogan MP et al. The genetic regulatory architecture and epigenomic basis for age-related changes in rattlesnake venom. Proc Natl Acad Sci. 2024;121(16):e2313440121. https://doi.org/10.1073/pnas.2313440121.
42. Mochales-Riaño G, et al. Chromosome-level reference genome for the medically important Arabian horned viper (*Cerastes gasperettii*). GigaScience. 2025;14. https://doi.org/10.1093/gigascience/giaf030.
43. Pardos-Blas JR, Irisarri I, Abalde S, Afonso CML, Tenorio MJ, Zardoya R. 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.
44. Suryamohan K, 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.
45. Talavera A, et al. Unveiling the evolutionary history of european vipers and their venoms from a multi-omic approach. Mol Ecol. 2025;70019:e. https://doi.org/10.1111/mec.70019.
46. Gutiérrez J, et al. Preclinical evaluation of the efficacy of antivenoms for snakebite envenoming: state-of-the-art and challenges ahead. Toxins. 2017;9(5):163. https://doi.org/10.3390/toxins9050163.
47. Alencar LRV, et al. Diversification in vipers: phylogenetic relationships, time of divergence and shifts in speciation rates. Mol Phylogenet Evol. 2016;105:50–62. https://doi.org/10.1016/j.ympev.2016.07.029.
48. 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. 2024;191:107979. https://doi.org/10.1016/j.ympev.2023.107979.
49. 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.
50. Waldhauser V et al. Unmasking the desert phantom: genetic evidence does not support the validity of the enigmatic Boehme's Horned Viper Zool Scr. 2025;54(6):817–822. https://doi.org/10.1111/zsc.70013.
51. Brito JC, Fahd S, Geniez P, Martínez-Freiría F, Pleguezuelos JM, Trape J-F. Biogeography and conservation of viperids from North-West Africa: an application of ecological niche-based models and GIS. J Arid Environ. 2011;75(11):1029–1037. https://doi.org/10.1016/j.jaridenv.2011.06.006.
52. Phelps T. Old World vipers: a natural history of the Azemiopinae and Viperinae. Frankfurt am Main: Edition Chimaira; 2010.
53. Liz AV, et al. Historical biogeography of the Saharan horned viper enlightens past dynamics of hyperarid desert habitats. J Arid Environ. 2024;222:105171. https://doi.org/10.1016/j.jaridenv.2024.105171.
54. Sindaco R, Venchi A, Grieco C. The reptiles of the Western Palearctic. 2. Annotated checklist and distributional atlas of the snakes of Europe, North Africa. Middle East Cent Asia Update To. 2013;1.
55. Šmíd J, et al. Diversity patterns and evolutionary history of Arabian squamates. J Biogeogr. 2021;48(5):1183–99. https://doi.org/10.1111/jbi.14070.
56. Bazaa A, Marrakchi N, El Ayeb M, Sanz L, Calvete JJ. Snake venomics: comparative analysis of the venom proteomes of the Tunisian snakesCerastes cerastes, Cerastes vipera andMacrovipera lebetina. PROTEOMICS. 2005;5(16):4223–4235. https://doi.org/10.1002/pmic.200402024.
57. Damm M, Hempel B-F, Süssmuth RD. Old world vipers—a review about snake venom proteomics of viperinae and their variations. Toxins. 2021;13(6):427. https://doi.org/10.3390/toxins13060427.
58. Fahmi L et al. Venomics and antivenomics profiles of North African Cerastes cerastes and C. vipera populations reveals a potentially important therapeutic weakness. J Proteomics. 2012;75(8):2442–2453. https://doi.org/10.1016/j.jprot.2012.02.021.
59. Ozverel CS et al. Investigating the cytotoxic effects of the venom proteome of two species of the Viperidae family (Cerastes cerastes and Cryptelytrops purpureomaculatus) from various habitats. Comp Biochem Physiol Part C Toxicol Pharmacol. 2019;220:20–30. https://doi.org/10.1016/j.cbpc.2019.02.013.
60. Zancolli G et al. When one phenotype is not enough: divergent evolutionary trajectories govern venom variation in a widespread rattlesnake species. Proc R Soc B Biol Sci. 2019;286(1898):20182735. https://doi.org/10.1098/rspb.2018.2735.
61. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90. https://doi.org/10.1093/bioinformatics/bty560.
62. Andrews S. FastQC: a quality control tool for high throughput sequence data.
63. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. Available: http://arxiv.org/abs/1303.3997.

64. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352.
65. Institute B. Picard Tools, Broad Institute, GitHub repository.
66. McKenna A et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–1303. https://doi.org/10.1101/gr.107524.110.
67. Danecek P, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10(2). https://doi.org/10.1093/gigascience/giab008.
68. Danecek P et al. The variant call format and VCFtools, Bioinformatics. 2011;27(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330.
69. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4(1). https://www.R-project.org/.
70. Team RC. R: a language and environment for statistical computing. Vienna, Austria. 2021. Available: https://www.R-project.org/.
71. Wickham H. ggplot2: elegant graphics for data analysis. Springer-Verlag New York. 2016. Available: https://ggplot2.tidyverse.org.
72. Korneliussen T, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. 2014. Available: http://www.biomedcentral.com/1471-2105/15/356.
73. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74. https://doi.org/10.1093/molbev/msu300.
74. Bouckaert R, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15(4). https://doi.org/10.1371/journal.pcbi.1006650.
75. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nat. 2011;475(7357):493–496. https://doi.org/10.1038/nature10231.
76. Green RE, et al. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. Science. 2014;346(6215):1254449. https://doi.org/10.1126/science.1254449.
77. Schield DR, 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.
78. Malinsky M, Matschiner M, Svardal H. Dsuite - Fast D -statistics and related admixture evidence from VCF files. Mol Ecol Resour. 2021;21(2):584–595. https://doi.org/10.1111/1755-0998.13265.
79. Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics. 2016;32(11):1749–51. https://doi.org/10.1093/bioinformatics/btw044.
80. Armstrong EE, et al. Long live the king: Chromosome-level assembly of the lion (Panthera leo) using linked-read, Hi-C, and long-read data. BMC Biol. 2020;18(1). https://doi.org/10.1186/s12915-019-0734-5.
81. Mochales-Riaño G, et al. Genomics reveals introgression and purging of deleterious mutations in the Arabian leopard (Panthera pardus nimr). iScience. 2023;26(9):107481. https://doi.org/10.1016/j.isci.2023.107481.
82. Talavera A et al. Lizards on a sky archipelago: genomic approaches to the evolution of the mountain genus Iberolacerta. Mol Phylogenet Evol. 2026;108527.
83. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. Am J Hum Genet. 2021;108(10):1880–90. https://doi.org/10.1016/j.ajhg.2021.08.005.
84. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(30:307–321. https://doi.org/10.1093/sysbio/syq010.
85. Martin SH, Van Belleghem SM. Exploring evolutionary relationships across the genome using topology weighting. Genetics. 2017;206(1):429–38. https://doi.org/10.1534/genetics.116.194720.
86. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011;21(6):974–984. https://doi.org/10.1101/gr.114876.110.
87. Myers EA, 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 Evol. 2022;14(7):evac082. https://doi.org/10.1093/gbe/evac082.
88. Carné A, Fathinia B, Rastegar-Pouyani E. Molecular phylogeny of the Arabian Horned Viper, Cerastes gasperettii (Serpentes: Viperidae) in the Middle East. Zool Middle East. 2020;66(1):13–20. https://doi.org/10.1080/09397140.2020.1711622.
89. Burriel-Carranza B, et al. Clinging on the brink: Whole genomes reveal human-induced population declines and severe inbreeding in the critically endangered emirati leaf-toed Gecko (*Asaccus caudivolvulus*). Mol Ecol. 2024;e17451. https://doi.org/10.1111/mec.17451.
90. Alonazi M et al. Functional characterization and anti-tumor effect of a novel group II secreted phospholipase A2 from snake venom of Saudi Cerastes cerates gasperetti. Molecules. 2023;28(18):6517. https://doi.org/10.3390/molecules28186517.

91. Zaeri S, Aghaei Z, Mashayekhi NR, Salemi A, Seyedian R. Pharmacoligical characterization of the iranian *Cerastes cerastes gasperettii* (Reptilia: Ophidia: Viperidae) venom. J Emerg Pract Trauma. 2021;7(2):123–126. https://doi.org/10.34172/jept.2021.22.
92. Gutiérrez J, Lomonte B. Phospholipases A1: unveiling the secrets of a functionally versatile group of snake venom toxins. Toxicon: official J Int Soc Toxinology. 2013;62:27–39. https://doi.org/10.1016/j.toxicon.2012.09.006.
93. Djebari FL, Martin-Eauclaire MF. Purification and characterization of a phospholipase A2 from Cerastes cerastes (horn viper) snake venom. Toxicon. 1990;28(6):637–46. https://doi.org/10.1016/0041-0101(90)90252-3.
94. Puinongpo W et al. Existence of Bov-B LINE Retrotransposons in snake lineages reveals recent multiple horizontal gene transfers with copy number variation. Genes. 2020;11(11):1241. https://doi.org/10.3390/genes11111241.

Authors and Affiliations

Gabriel Mochales-Riaño[1,2] · Adrián Talavera[1,3] · Bernat Burriel-Carranza[1,4] · Johannes Els[5] · Mohammed Shobrak[6] · Salem Busais[7] · Maria Estarellas[1] · Jiří Šmíd[8,9] · Fernando Martínez-Freiría[10] · Mark J. Margres[11] · Salvador Carranza[1]

✉ Gabriel Mochales-Riaño
gabri.mochales@gmail.com

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

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

3 CREAF, Cerdanyola del Vallès, Catalonia 08193, Spain

4 Museu de Ciències Naturals de Barcelona, P° Picasso s/n, Parc Ciutadella, Barcelona 08003, Spain

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

6 National Center for Wildlife, Prince Saud Al Faisal Wildlife Research Centre, Taif, Saudi Arabia

7 Department of Biology, Faculty of Education, Aden University, Aden, Yemen

8 Department of Zoology, Faculty of Science, Charles University, Vinicná 7, Prague, Czech Republic

9 Department of Zoology, National Museum, Cirkusová 1740, Prague, Czech Republic

10 CRETUS - Department of Zoology, Genetics and Physical Anthropology, Universidade de Santiago de Compostela, Rúa Lope Gómez de Marzoa, Santiago de Compostela 15782, Spain

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