

Supplementary information

The roles of balancing selection and recombination in the evolution of rattlesnake venom

In the format provided by the authors and unedited

Supplementary Results

Sampling and filtration of copy-number variation

Our sampling includes genome resequencing data for 68 individuals, which we filtered and mapped to the *C. viridis* reference genome¹. A mean of 97% \pm standard deviation of 1.7% reads aligned uniquely, corresponding to $28.7\times \pm 16.9\times$ read depth per sample. We called variants using GATK v4.0.8.1^{2,3} and filtered to remove low quality genotypes, sites with extremely low or high read depths, indels, and repeats from further analysis. This procedure yielded 27,749,933 SNPs passing filtering criteria. Phased non-singleton variants were also available for CV1 and CO1 populations from⁴, which we used in analyses requiring haplotypes.

Because inferences of selection can be confounded by structural and copy number variants (CNVs), particularly in tandemly duplicated gene arrays, we conducted extensive assessments, filtering approaches, and post-hoc validations (Extended Data Fig. 1a) to ensure bioinformatic artifacts from these sources of variation would not bias our inferences. For analyses focused on venom gene clusters, this included conservatively masking genotypes in genomic regions with evidence of CNVs within individuals and populations to avoid interpretations based on spurious paralogous mappings (see Methods; Extended Data Fig. 1a-b). We further removed SNPs with significant departures from Hardy-Weinberg equilibrium due to excess heterozygosity driven by potential bioinformatic artifacts. We would expect persistent bioinformatic artifacts of CNVs, especially in the case of collapsed duplications, to produce multimodal allele frequency spectra within distinct spikes of high allele frequency counts. Instead, post-filtered minor allele frequency spectra for venom genes are diffuse and lack punctuated spikes of high minor allele counts (Extended Data Fig. 1c), consistent with effective removal of artifacts. As further validation that our filtering approaches removed false variation, there are no significant differences in genotype quality scores for SVMP, SVSP, and PLA2 venom gene clusters and other genes (Mann-Whitney U, $P > 0.05$). For context, we also visualize population genetic statistics in venom gene regions together with variation in the proportion of individuals per population with masked genotypes in CNVs (% CNV).

Population structure and demographic history

We inferred population structure within and between *C. viridis* and *C. oregonus* using ADMIXTURE⁵ across different numbers of genetic clusters (K) values. These analyses clearly distinguish *C. viridis* from *C. oregonus* under a $K = 2$ model (Fig. 1c). Additional subdivision between CO1 (California *C. o. helleri*) and CO2 (Idaho *C. o. oregonus*) is inferred under $K = 3$. This is the best-supported model based on the cross-validation procedure and is consistent with the phylogeographic break at the Sacramento-San Joaquin River Delta detected in⁶. Under the $K = 4$ model, which has similar support to $K = 3$ (Supplementary Fig. 1), each of the four populations largely corresponds to a distinct cluster. This model distinguishes Montana and Colorado *C. viridis* populations (CV2 and CV1, respectively), though there are a number of Colorado individuals with low-probability assignment to the Montana cluster, suggesting weaker population structure between *C. viridis* populations as a potential consequence of northern range expansion in the recent past⁷.

We used the pairwise sequentially Markovian coalescent (PSMC⁸) to infer the demographic history of each population and to inform forward-time simulations of evolutionary mechanisms (described below). PSMC estimates indicate that each species experienced multiple expansion and contraction events within the last ten million years (Fig. 1d), with major population contractions within the last 100,000 years. Inferred population sizes coalesce around three million years ago (i.e., the estimated divergence time for the two species; Fig. 1a). Size estimates for *C. oregonus* and *C. viridis* population pairs are similar up until the very recent past, with divergent patterns inferred for *C. oregonus* populations occurring roughly 100,000 years ago and around 30,000 years ago for *C. viridis* populations, suggesting comparatively recent population separations within each species. Estimates of the recent past suggest that population genetic diversity has been shaped by population contractions coincident with Pleistocene glacial cycles (Fig. 1d). Recent declines are steepest for the northern populations in *C. viridis* and *C. oregonus*, consistent with founder effects on genetic diversity in northern populations⁶.

Genomic patterns of genetic diversity and differentiation

We examined genome-wide patterns of population genetic diversity and differentiation to provide comparative context for patterns in venom gene regions (Supplementary Fig. 3a-c; Supplementary Table 9). Mean intra-population nucleotide diversity (π) ranges from 0.0019 – 0.0027 across genomic regions, with numerous peaks and valleys scattered across chromosomes in each population. Inter-population sequence divergence (d_{xy}) varies similarly across the genome (Supplementary Fig. 3b) and is strongly correlated with π (Supplementary Fig. 3d; Supplementary Table 10; Spearman's $\rho = 0.83$ to 0.99; $P < 2.2$

$\times 10^{-16}$). Mean d_{xy} is generally higher than π (mean $d_{xy} = 0.0026$ to 0.0052), indicating greater sequence differences between populations than within populations and consistent with inferences of population genetic structure (Fig. 1c). Relative genetic differentiation (F_{st}) differs substantially among intra- and inter-species comparisons (Supplementary Fig. 3c) and is weakly correlated across comparisons (Supplementary Fig. 3; Supplementary Table 10; $\rho = 0.066$ to 0.16 , $P < 3.7 \times 10^{-14}$). F_{st} and π are negatively correlated in all comparisons (Supplementary Table 10; $\rho = -0.12$ to -0.77 , $P < 2.2 \times 10^{-16}$). F_{st} and d_{xy} are also negatively correlated across the genome for each pair of populations ($\rho = -0.09$ to -0.47 , $P < 2.2 \times 10^{-16}$).

The conserved landscape of genetic diversity across populations may be explained by conserved recombination rate variation across the genome^{9,10}, which is predicted to influence the sorting of ancestral polymorphism under background selection and hitchhiking^{11,12}. Consistent with this prediction, π and d_{xy} are correlated with population-scaled recombination rate ($\rho = 4N_e r$; Supplementary Table 10; $r = 0.47$ to 0.62 , $P < 2.2 \times 10^{-16}$). F_{st} and ρ are negatively correlated ($\rho = -0.09$ to -0.52 , $P < 2.2 \times 10^{-16}$), consistent with greater differentiation due to reduced ancestral polymorphism in low recombination regions.

Diversity and differentiation in major venom gene regions

The SVMP and SVSP regions exhibit substantial heterogeneity in estimates of π , d_{xy} , and F_{st} , with fine-scale variation in patterns at sites within and linked to venom genes in *C. viridis* and *C. oreganus* populations. To dissect this heterogeneity, we compared diversity estimates in venom genes to intergenic regions. These comparisons show that venom gene π is either similar to intergenic regions (e.g., SVMP region in CV1 and CV2; Mann-Whitney U, $P > 0.5$) or higher (e.g., SVMP in CO1 and CO2 and SVSP in all populations; $P < 0.05$). Similarly, d_{xy} is on average 1.5-fold higher in SVMP and SVSP venom genes than intergenic regions, with significant differences in each comparison ($P < 0.05$) except for the SVMP region between CV1 and CV2 ($P = 0.97$; Supplementary Table 4). These patterns of π and d_{xy} indicate a concentration of genetic diversity in coding regions of the SVMP and SVSP venom gene clusters.

Broadly, patterns in the SVMP, SVSP, and PLA2 venom gene regions are not consistent with the hypothesis that venom has been predominantly shaped by ‘extreme positive selection’ (e.g.,¹³), which would predict decreased genetic diversity and high relative differentiation surrounding targets of selection. Patterns suggesting positive directional selection are mainly present in the PLA2 venom cluster (e.g., PLA2 B1 in *C. viridis*), which we explored further using additional test statistics (see Main Article text). Our results instead highlight exceptional levels of genetic diversity across major venom regions

(SVMPs and SVSPs) and specific paralogs (e.g., PLA2 A1), supporting the alternative hypothesis of balancing selection.

Signatures of selection in other venom genes

For comparison to patterns in major venom gene tandem arrays, we analyzed population genetic variation in venom gene families outside of the three major venom gene regions. These ‘other’ venom components represent an interesting comparison to SVMPs, SVSPs, and PLA2s because they are encoded on different chromosomes¹ (Supplementary Table 11) and are not present in large multigene tandem arrays in the *C. viridis* genome, yet are highly expressed in the venom gland¹⁴. These ‘other’ venom gene families include C-type lectin (CTL), cysteine-rich secretory proteins (CRISPs), exonucleases, L-amino acid oxidases (LAAOs), vascular endothelial growth factors (VEGFs), glutaminyl cyclases (GCs), and vespryn. The majority of these venom genes have patterns similar to the genomic background, lack distinguishable peaks of genetic diversity or differentiation, and do not show strong signatures of directional or balancing selection (Supplementary Figs. 4-5), suggesting that many of these components are evolving in a nearly neutral fashion, and likely under purifying selection. However, a small number of these genes do show signatures of selection in one or both species, including extreme $|iHS|$ and or β values exceeding the genome-wide 95th quantile. For example, CTL shows signals of directional positive selection in *C. oreganus* (e.g., low π , negative Tajima’s D , high $|iHS|$; Supplementary Figs. 4-5). We also find signatures of positive directional selection across vespryn in *C. viridis* (e.g., low π , high Fst , negative Tajima’s D , high $|iHS|$, low β). In contrast, CRISP 1, CRISP 2, and LAAO 3 each show signatures consistent with balancing selection in both species (e.g., high π and d_{xy} , low Fst and d_f , and high β); CRISP 2 and LAAO3 exhibit particularly extreme β peaks (Supplementary Fig. 5; Supplementary Table 11). These results highlight heterogeneity in selection regimes impacting different components of the venom phenotype, including variation between species, further underscoring the complexity of shallow timescale venom evolution.

Recombination rate variation in major venom gene regions

We investigated the relationship between recombination rate and signatures of selection to clarify their roles in venom adaptation and to test the alternative hypothesis that high recombination rates are themselves artifacts of balancing selection, as balanced polymorphism can mimic recombination hotspots due to local reductions in LD¹⁵. If high recombination estimates in venom regions were driven by balancing selection, we would expect strong positive relationships between ρ/π and estimates of allele frequency correlation, β . Instead, we find no significant associations between these variables in any of the

major venom gene regions (Spearman's $\rho = -0.25$ to -0.04 , $P = 0.14$ to 0.81). While we cannot rule out that balanced polymorphism has impacted recombination rate estimates in some genomic intervals, these results reject the hypothesis that high recombination rate estimates are a broad artifact of balancing selection. We further compared ρ/π and $|iHS|$, predicting a positive relationship under a scenario of greater selection efficiency due to recombination. We find significant positive correlations between ρ/π and $|iHS|$ in the SVMP and SVSP venom regions ($\rho = 0.47$ to 0.66 , $P < 0.05$) and the PLA2 region of *C. oreganus* ($\rho = 0.39$, $P = 0.048$). Generally consistent with this expectation, ρ/π and $|iHS|$ are also positively correlated in chromosome-wide comparisons ($\rho = 0.2$ to 0.42 , $P < 0.001$). Finally, we investigated the possibility that recombination rates were influenced by potential bioinformatic artifacts of CNVs by comparing ρ/π and the proportion of individuals in each population within evidence of a CNV across venom regions. We find no significant associations in any of the venom regions in *C. viridis* or *C. oreganus* ($P > 0.05$), suggesting that if any artifacts persisted after our strict filtering approaches, then they do not strongly influence recombination rate inferences.

Supplementary Discussion

Removal of artifacts of copy-number variation

An important consideration in our analysis was that signatures of balancing selection can be misled by bioinformatic artifacts of CNVs being misinterpreted as allelic variation (i.e., from collapsed duplicates). For this reason, we conducted stringent filtering and masking of putative CNV regions and *post hoc* analysis to confirm that allele frequency spectra were not biased by such artifacts (Extended Data Fig. 1). Our findings are consistent with an effective removal of spurious variation, and several venom gene regions with the strongest evidence of balancing selection show little or no evidence of CNVs within or between species. We also find among the strongest signals of balancing selection in the population in which we expect the least structural variation compared to the reference genome (i.e., CV1, the population from which the reference genome was sampled). These results, combined with *post hoc* allele frequency analyses indicating no persistent post-filtering effects of collapsed duplicates, confirm that inferences of balancing selection are robust to bioinformatic artifacts and represent legitimate evolutionary features of major venom gene regions.

A role of balancing selection in venom evolution

While signatures of balancing selection are most pronounced in the major multigene venom families, we also find evidence of balanced polymorphism across CRISPs 1 and 2, and in LAAO 3 (Supplementary Figs. 4-5). These findings support the hypothesis that balancing selection has maintained diversity in venom genes outside of the three dominant venom gene families, for which less is known about the biological relevance of their activities in envenomation^{16,17}. Our results do not wholly reject the role of directional selection in venom adaptation, and a subset of venom genes show signatures of directional selection (e.g., PLA2 B1 and vespryn in *C. viridis* and CTL in *C. oreganus*). We also acknowledge the possibility that directional selection may have operated in the past, or may occasionally impact specific SVMP and SVSP genes, and that its footprints have been ‘overwritten’ by the dominant effects of balancing selection at nearby genes. Still other venom genes had population genetic signatures that were indistinguishable from the genomic background and are likely sorting neutrally or evolving under purifying selection. These results broadly support that the maintenance of diversity in large venom gene complexes affords a greater ability for evolution to constantly tune their allelic composition, while other minor venom components remain more static. This hypothesis aligns with less dynamic evolution of minor venom gene families in pitvipers, generally^{13,18-22}. Collectively, our data provide evidence for the multiple forms of selection among gene families, and likely even within gene families, contributing to the evolutionary dynamics shaping snake venom genotypic and phenotypic diversity.

Predator-prey coevolution and trans-species polymorphism

A major role of balancing selection in venom evolution is logically consistent with predictions of antagonistic predator-prey coevolution, including that an outcome of adaptive evolution may be a genetically diverse set of segregating alleles rather than a single optimal genotype²³. In contrast, directional positive selection may lead to evolutionary ‘dead ends’ in which alleles with high fitness at certain points in time and space become fixed, but subsequently have reduced fitness as prey evolve effective resistance. In extreme cases, evolved resistance in prey could render fixed venom alleles completely ineffective, at which point the snake predator population must wait for new beneficial mutations to evolve or arrive through gene flow from other populations. Evidence for balancing selection in major venom gene regions therefore provides a plausible explanation for the ability of snake predators to keep pace with coevolving prey through selective processes that maintain venom allelic diversity. This hypothesis further predicts that genomic regions underlying venom resistance mechanisms in prey may

show similar signatures of balancing selection, and thus detailed population genomic investigations of prey populations may have the potential to identify genomic regions that coevolve with venom.

There are multiple ways in which local prey diversity, population size, population structure, and natural history may interact with patterns of selection on snake venom. For example, prey population structure may be a driver of spatially variable selection on venom and large prey population sizes may impose stronger reciprocal selection pressure on snake predators due to more efficient selection for effective resistance mechanisms. Holding et al.¹⁸ also showed that pitviper venom complexity was strongly predicted by the phylogenetic diversity of prey, where species with diverse diets tend to have more complex venoms. This trend suggests that the relative role of balancing selection mechanisms may depend on local phylogenetic diversity of prey, in which case high prey diversity would impose balancing selection through a larger number of coevolutionary interactions with multiple resistant prey species.

Recombination and selection shape venom adaptation

The genetic architecture of a trait can have a profound impact on the persistence of balanced polymorphism due to selective interference related to recombination rate^{24,25}. Major venom gene regions present an intriguing set of examples in which balancing selection may be especially efficient despite the potential for high selective interference. Specifically, the genetic architecture of major venom gene families consists of tandem arrays of paralogs in relatively confined genomic regions^{1,20,26,27}. While each venom gene may contribute to the venom phenotype (depending on gene-by-gene regulation and expression²⁸), epistatic interactions among genes are not well understood. As such, it has remained unclear how such compact tandem arrays of fitness-relevant loci could experience independent selection to target distinct molecules and prey through vastly different biological functions (e.g., PLA2 genes) while circumventing pronounced hitchhiking effects of close physical linkage.

Our results reveal high local recombination rates and the presence of recombination hotspots in major venom tandem arrays, providing key insight on how selection can operate efficiently on individual venom loci (Fig. 6; Extended Data Fig. 10). Consequently, the rapid decay of LD between venom gene paralogs suggests reduced selective interference among loci, allowing natural selection to operate more efficiently on individual loci and novel allele combinations²⁹. This erosion of LD aligns with our inferences of balancing selection on venom, as reduced LD may be expected in cases of recombination between long-term balanced polymorphisms³⁰, further explaining how the genetic architecture of venom tandem arrays can circumvent selective interference due to tight physical linkage. Another consequence of high

recombination may be increased exposure of slightly deleterious alleles to selection³¹, reducing the build-up of genetic load in gene-dense venom regions. Previous studies have also shown that ancestral duplication and deletion events have been driven by non-allelic homologous recombination between venom gene-associated repeat elements^{26,32}. This mechanism is a logical explanation for variable venom gene structure among species. Together with our demonstration of high allelic recombination rates within populations, these observations indicate that various forms of recombination have played central roles in the ancestral origins and contemporary adaptation of venom.

Supplementary Methods

Training a classifier to predict evolutionary mechanism

To better explore potential evolutionary mechanisms shaping haplotypic variation in venom gene clusters, we trained predictive classification models with simulated data designed to mimic the evolutionary history of *C. viridis* and *C. oreganus*. Specifically, using the forward-time simulator SLiM 3.7.1³³, we generated $L = 10$ kb long sequences under demographic histories inferred by PSMC⁸ for CO1 and CV1 populations and respectively sampled 34 and 36 haplotypes for CO1 and CV1 simulations to match empirical sample sizes. To further match the empirical data, we assumed a per-site per-generation neutral mutation rate of $\mu = 8.4 \times 10^{-9}$ ³⁴ and species-specific per-site per-generation recombination rates of $r = 6.08 \times 10^{-8}$ and 1.79×10^{-8} for *C. viridis* and *C. oreganus*, respectively⁴. We trained each model using simulations spanning two million generations, which is six million years assuming a generation time of three years³⁵. To ensure that sufficient variation was present in the simulated sequences, we allowed for an initial burnin of $10N_e$ generations³³, assuming PSMC-estimated effective sizes of $N_e = 41,104$ and $64,281$ diploid individuals for CO1 and CV1, respectively. These parameters were used to model neutral variation for CO1 and CV1. As is common practice when using forward-time simulators^{36,37}, for computational efficiency we scaled simulations by multiplying mutation rates, recombination rates, and selection coefficients by a factor of $\lambda = 50$ while also dividing population sizes and numbers of generations by a factor of λ .

For simulations including selection, we introduced a mutation at a site in the center of each simulated sequence, with the site subsequently undergoing one of two balancing selection mechanisms: heterozygote advantage or negative frequency-dependent selection. For both settings, we drew the per-generation selection coefficient s uniformly at random on a \log_{10} scale over the interval $[0.001, 0.1]$, allowing for a range of weak to strong balancing selection. We further introduced the selected allele at T

generations in the past, where T was drawn uniformly between 1.5 million and 4.5 million years, corresponding to 0.5 to 1.5 million generations, spanning the estimated split time of *C. viridis* and *C. oreganus* (~ 3 MYA⁴). At the selected site, the relative genotype fitnesses from time T to the present are $w_{aa} = 1$, $w_{Aa} = 1 + hs$, $w_{AA} = 1 + s$, where A is the derived (mutant) allele and h is the dominance coefficient. We also considered a range of values for the equilibrium frequency p of the derived allele at the selected site, sampling p uniformly at random on the interval $(0.5, 0.9]$. Because we simulated balancing selection, we also ensured that the selected allele was neither lost (frequency of 0) nor fixed (frequency of 1) within the population, and therefore remained a balanced polymorphism until sampling. Simulated replicates in which the balanced allele was lost or fixed were rerun. To simulate heterozygote advantage, we set $h = p/(2p - 1)$, which derives from the equilibrium frequency $p = h/(2h - 1)$ under this fitness scheme³⁸. Under negative frequency-dependent selection, we assumed an additive model for genotype fitness in which $h = 0.5$, but instead assumed that the relative fitness $w_A(t) = 1 + p - p(t)$ of the derived (mutant) allele at the selected site changed each generation t , and depended on the difference of the equilibrium frequency p of the derived allele and its frequency $p(t)$ at generation t . Under this setting, assuming the relative fitness of $w_a(t) = 1$, the derived allele has an advantage when $w_A(t) > w_a(t)$ (i.e., when $p(t)$ is rare) and has a disadvantage when $w_A(t) < w_a(t)$ (i.e., when $p(t)$ is common).

For each simulated setting and for each population, we generated 10,000 neutral, 4,000 negative frequency-dependent selection, and 4,000 heterozygote advantage replicates. We used these $N = 18,000$ simulated replicates to train a machine learning classifier to discriminate among the three classes. To extract features for input into our machine learning models, we used CoMuStats³⁹ to generate 20 summary statistics that assess diverse characteristics of haplotype variation in a genomic region, and which have been used for identifying evolutionary settings with greatest support to observed haplotype data. We considered and evaluated five different classifier models⁴⁰ that took the 20 summary statistics as input features: one linear classifier using multinomial regression, and four non-linear classifiers that used a neural network architecture with one hidden layer and $M \in \{2, 4, 8, 16\}$ nodes in the hidden layer that predicted three output class probabilities with a softmax activation function. We trained these five models using the N training observations with the R package `nnet`⁴¹. Classification models typically exhibit best performance when given equal numbers of training observations for each class, such that the trained model does not favor the dominant class over the others⁴⁰. However, we wanted to be conservative with our classifiers, and therefore chose to have a greater number of training observations for the neutral class compared to the balancing selection classes, ensuring that any bias in the model would favor of evolutionary neutrality.

We evaluated the predictive accuracy of the five models with increasing model complexity; the least complex of our classification models is multinomial regression with the fewest parameters, whereas the neural network with $M = 16$ nodes in its hidden layer is the most complex with the greatest number of parameters. A key consideration when training classifiers is that more complex models will fit the training data better than less complex ones, potentially leading to overfitting⁴⁰. Hence, to identify the best-fit model with adequate complexity, we conducted 10-fold cross validation, and selected the model with the smallest validation error determined by the categorical cross entropy measure⁴⁰, which is akin to the negative log likelihood. The best model for both CO1 and CV1 was the neural network with $M = 2$ nodes in its hidden layer (Supplementary Fig. 2), which is neither the least nor most complex model, echoing the principle that the optimal model typically has intermediate complexity that balances the bias-variance tradeoff⁴⁰. We then retrained the neural network models with $M = 2$ hidden nodes on all $N = 18,000$ simulated observations in each species, for application to the empirical dataset.

We explored evidence for the three simulated evolutionary scenarios across windows of the PLA2, SVMP, and SVSP venom gene clusters. Specifically, we considered SNPs falling within the genomic interval spanning 50 kb immediately upstream and downstream of each cluster. Only SNPs with no missing data within the span of this region were considered for downstream analysis. We then used CoMuStats to compute 20 summary statistics in 10 kb windows along each region, shifting windows with a step size of 100 bp. Windows were skipped if there were too few SNPs, either due to missing or masked data, gaps, or fewer SNPs than would be expected from a single diploid individual. That is, if a window had fewer than $\theta = 4N_e\mu L$ SNPs, which is the expected number of SNPs when sampling two haplotypes⁴², then it was removed. The summary statistics at each 10 kb region were then used as input to the trained neural network model to predict the probability of neutrality, negative frequency dependence, or heterozygote advantage classes.

Supplementary References

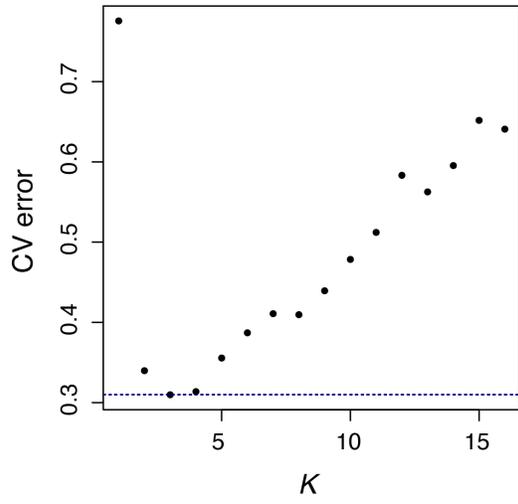
1. Schield, D. R. *et al.* The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome Res.* **29**, 590–601 (2019).
2. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
3. Van der Auwera, G. A. *et al.* From FastQ data to high-confidence variant calls: the genome

- analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 10–11 (2013).
4. Schield, D. R. *et al.* Snake recombination landscapes are directed by PRDM9 but concentrated in functional regions. *Mol. Biol. Evol.* **37**, 1272–1294 (2020).
 5. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
 6. Holding, M. L., Sovic, M. G., Colston, T. J. & Gibbs, H. L. The scales of coevolution: comparative phylogeography and genetic demography of a locally adapted venomous predator and its prey. *Biol. J. Linn. Soc.* **132**, 297–317 (2021).
 7. Schield, D. R. *et al.* Allopatric divergence and secondary contact with gene flow: a recurring theme in rattlesnake speciation. *Biol. J. Linn. Soc.* **128**, 149–169 (2019).
 8. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
 9. Noor, M. A. & Bennett, S. M. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Hered.* **103**, 439–444 (2009).
 10. Cruickshank, T. E. & Hahn, M. W. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23**, 3133–3157 (2014).
 11. Cutter, A. D. & Payseur, B. A. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat. Rev. Genet.* **14**, 262 (2013).
 12. Burri, R. Interpreting differentiation landscapes in the light of long-term linked selection. *Evol. Lett.* **1**, 118–131 (2017).
 13. Aird, S. D. *et al.* Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. *Genome Biol. Evol.* **9**, 2640–2649 (2017).
 14. Perry, B. W. *et al.* Snake venom expression is coordinated by novel regulatory architecture and the integration of multiple co-opted vertebrate pathways. *Genome Res.* (2022).

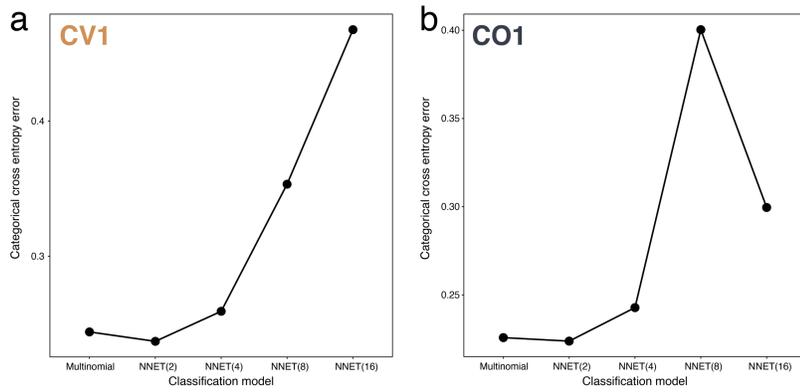
15. DeGiorgio, M., Lohmueller, K. E. & Nielsen, R. A model-based approach for identifying signatures of ancient balancing selection in genetic data. *PLoS Genet.* **10**, e1004561 (2014).
16. Peichoto, M. E. & Santoro, M. L. Reptile venom cysteine-rich secretory proteins. in *Handbook of Venoms and Toxins of Reptiles* (ed. Mackessy, S. P.) 225–239 (CRC Press, 2021).
17. Zuliani, J. P. *et al.* Reptile Venom L-Amino Acid Oxidases—Structure and Function. in *Handbook of Venoms and Toxins of Reptiles* (ed. Mackessy, S. P.) 413–430 (CRC Press, 2021).
18. Holding, M. L. *et al.* Phylogenetically diverse diets favor more complex venoms in North American pitvipers. *Proc. Natl. Acad. Sci.* **118**, (2021).
19. Rautsaw, R. M. *et al.* Intraspecific sequence and gene expression variation contribute little to venom diversity in sidewinder rattlesnakes (*Crotalus cerastes*). *Proc. R. Soc. B* **286**, 20190810 (2019).
20. Margres, M. J. *et al.* The Tiger Rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. *Proc. Natl. Acad. Sci.* **118**, (2021).
21. Mason, A. J. *et al.* Trait differentiation and modular toxin expression in palm-pitvipers. *BMC Genomics* **21**, 1–20 (2020).
22. Mackessy, S. P. Venom composition in rattlesnakes: trends and biological significance. in *The Biology of Rattlesnakes* (eds. Hayes, W. K., Beaman, K. R., Cardwell, M. D. & Bush, S. P.) (Loma Linda University Press, 2008).
23. Clarke, B. C. The evolution of genetic diversity. *Proc. R. Soc. B* **205**, 453–474 (1979).
24. Lenormand, T. & Otto, S. P. The evolution of recombination in a heterogeneous environment. *Genetics* **156**, 423–438 (2000).
25. Charlesworth, D. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* **2**, e64 (2006).
26. Dowell, N. L. *et al.* The deep origin and recent loss of venom toxin genes in rattlesnakes. *Curr.*

- Biol.* **26**, 2434–2445 (2016).
27. Casewell, N. R., Wagstaff, S. C., Harrison, R. A., Renjifo, C. & Wüster, W. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. *Mol. Biol. Evol.* **28**, 2637–2649 (2011).
 28. Rokyta, D. R., Margres, M. J. & Calvin, K. Post-transcriptional mechanisms contribute little to phenotypic variation in snake venoms. *G3 Genes, Genomes, Genet.* **5**, 2375–2382 (2015).
 29. Hill, W. G. & Robertson, A. The effect of linkage on limits to artificial selection. *Genet. Res. (Camb)*. **8**, 269–294 (1966).
 30. Leffler, E. M. *et al.* Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science* **339**, 1578–1582 (2013).
 31. Llaurens, V., Whibley, A. & Joron, M. Genetic architecture and balancing selection: the life and death of differentiated variants. *Mol. Ecol.* **26**, 2430–2448 (2017).
 32. Dowell, N. L. *et al.* Extremely divergent haplotypes in two toxin gene complexes encode alternative venom types within rattlesnake species. *Curr. Biol.* **28**, 1016–1026 (2018).
 33. Haller, B. C. & Messer, P. W. SLiM 3: forward genetic simulations beyond the Wright–Fisher model. *Mol. Biol. Evol.* **36**, 632–637 (2019).
 34. Green, R. E. *et al.* Three crocodylian genomes reveal ancestral patterns of evolution among archosaurs. *Science* **346**, 1254449 (2014).
 35. Castoe, T. A., Spencer, C. L. & Parkinson, C. L. Phylogeographic structure and historical demography of the western diamondback rattlesnake (*Crotalus atrox*): A perspective on North American desert biogeography. *Mol. Phylogenet. Evol.* **42**, 193–212 (2007).
 36. Yuan, X., Miller, D. J., Zhang, J., Herrington, D. & Wang, Y. An overview of population genetic data simulation. *J. Comput. Biol.* **19**, 42–54 (2012).
 37. Ruths, T. & Nakhleh, L. Boosting forward-time population genetic simulators through genotype

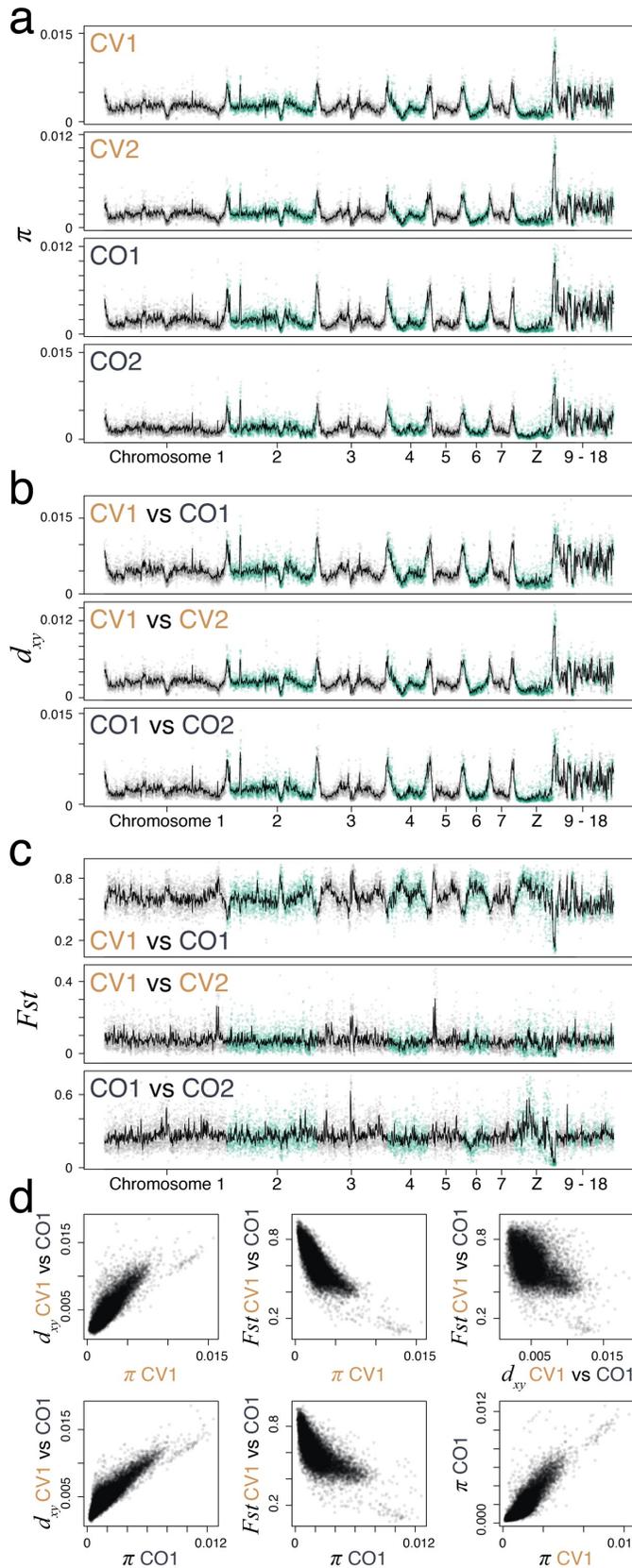
- compression. *BMC Bioinformatics* **14**, 1–12 (2013).
38. Gillespie, J. H. *Population genetics: a concise guide*. (JHU press, 2004).
 39. Papadantonakis, S., Poirazi, P. & Pavlidis, P. CoMuS: simulating coalescent histories and polymorphic data from multiple species. *Mol. Ecol. Resour.* **16**, 1435–1448 (2016).
 40. Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*. (Springer, 2009).
 41. Ripley, B., Venables, W. & Ripley, M. B. Package ‘nnet’. *R Packag. version 7*, 700 (2016).
 42. Watterson, G. A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276 (1975).



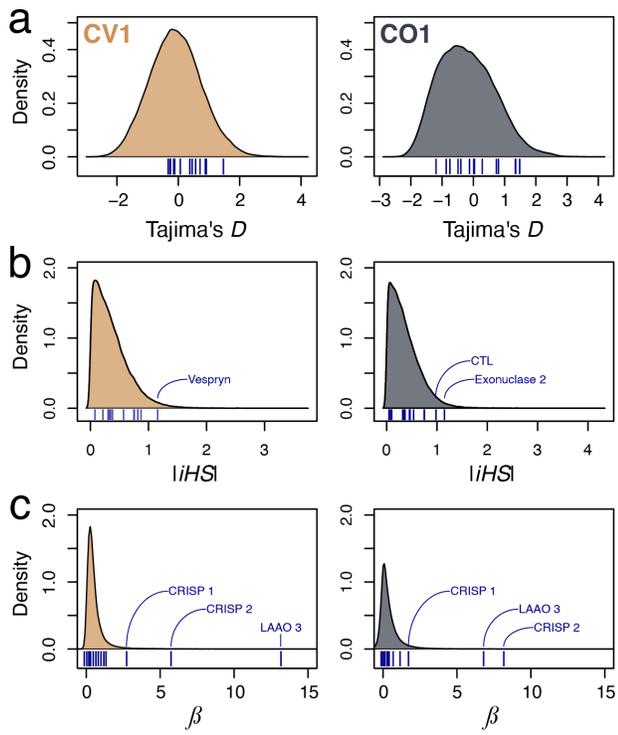
Supplementary Figure 1. Cross-validation (CV) error under K genetic clusters models 1-16 in ADMIXTURE analyses. The dashed line corresponds with the model with the lowest CV error, $K = 3$, which had a similar CV error value to $K = 4$.



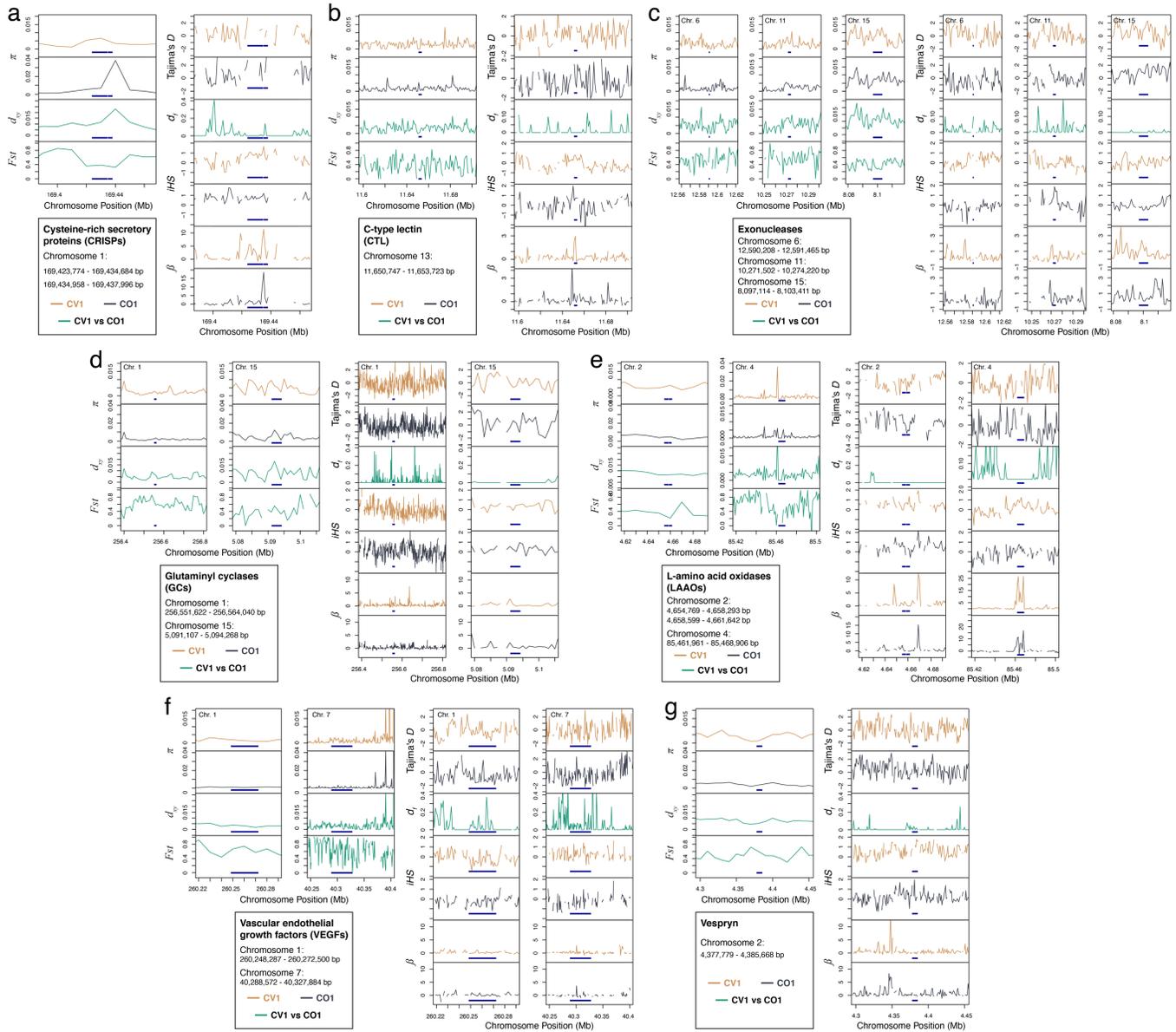
Supplementary Figure 2. Results of cross-validation procedure used to select the best-fit classification model to predict probabilities of evolutionary mechanisms in venom regions of CV1 (a) and CO1 (b) populations. The nnet(2) model had the lowest categorical cross entropy error in both populations and was selected as the best-fit classification model.



Supplementary Figure 3. Genome-wide patterns of genetic diversity within populations and differentiation between populations. Genome scans of (a) nucleotide diversity (π) in *C. viridis* (CV1 and CV2 populations) and *C. oregonus* (CO1 and CO2 populations), (b) sequence divergence (d_{xy}) between CV1 and CO1, CV1 and CV2, and CO1 and CO2, and (c) relative differentiation (F_{st}) between CV1 and CO1, CV1 and CV2, and CO1 and CO2. Shaded points represent estimates in 100 kb sliding windows and bold black lines represent 1 Mb windowed averages. **d** Correlations between measurements of genetic diversity and differentiation. Individual shaded points represent estimates from 100 kb sliding windows.



Supplementary Figure 4. Point estimates of Tajima's D (a), $|iHS|$ (b), and β (c) for venom genes outside the three major venom families (blue vertical dashes) compared to genome-wide distributions in CV1 and CO1 populations. Genes with mean estimates that exceeded the 95th percentile of genome-wide $|iHS|$ and β are labeled.



Supplementary Figure 5. Genomic scans of population genetic parameters across venom genes distributed throughout the genome, but which do not form tandem arrays. Lines show mean estimates in 10 kb and 1 kb sliding windows. Gaps in lines represent windows where there was insufficient data to calculate a mean estimate. Orange lines show estimates for the CV1 population, blue-grey lines show estimates for CO1, and blue-green lines show values for CV1 versus CO1 comparisons. Chromosomal positions are shown in Mb. Dark blue segments in each panel show venom gene locations. **a** Cysteine-rich secretory proteins (CRISPs), **b** C-type lectin (CTL), **c** Exonucleases, **d** Glutaminyl cyclases (GCs), **e** L-amino acid oxidases (LAAOs), **f** Vascular endothelial growth factors (VEGFs), **g** Vespryn.

Supplementary Table 1. Sampling used in this study for whole genome resequencing and population genomic analyses. Coverage was calculated assuming a genome size of 1.5 Gb.

Sample ID	Museum or Collector Number	Species (population)	Sex	Locality	Reads	Coverage	Reference
CA0346	CAS 229233	<i>Crotalus atrox</i>	Male	San Miguel Co., NM	183,088,910	18.3	Schild et al. 2020
CV0005	SPM-W09001	<i>Crotalus viridis</i> (CV1)	Male	Weld Co., CO	607,247,022	60.7	Schild et al. 2020
CV0007	SPM-W08152	<i>Crotalus viridis</i> (CV1)	Male	Weld Co., CO	493,612,334	49.4	Schild et al. 2020
CV0008	SPM-W09002	<i>Crotalus viridis</i> (CV1)	Male	Weld Co., CO	1,081,907,838	108.2	Schild et al. 2020
CV0009	–	<i>Crotalus viridis</i> (CV1)	Male	Weld Co., CO	103,878,358	10.4	Schild et al. 2021
CV0628	DRS207	<i>Crotalus viridis</i> (CV1)	Male	Weld Co., CO	623,428,734	62.3	Schild et al. 2020
CV0632	DRS212	<i>Crotalus viridis</i> (CV1)	Male	Larimer Co., CO	271,766,036	27.2	Schild et al. 2020
CV0641	DRS219	<i>Crotalus viridis</i> (CV1)	Male	Weld Co., CO	339,338,066	33.9	Schild et al. 2020
CV0642	DRS220	<i>Crotalus viridis</i> (CV1)	Male	Weld Co., CO	631,042,546	63.1	Schild et al. 2020
CV0643	DRS221	<i>Crotalus viridis</i> (CV1)	Male	Weld Co., CO	440,484,200	44.0	Schild et al. 2020
CV0648	DRS225	<i>Crotalus viridis</i> (CV1)	Male	Weld Co., CO	905,861,476	90.6	Schild et al. 2020
CV0004	–	<i>Crotalus viridis</i> (CV1)	Female	Weld Co., CO	430,179,708	39.5	Schild et al. 2020
CV0006	–	<i>Crotalus viridis</i> (CV1)	Female	Weld Co., CO	595,118,858	54.7	Schild et al. 2020
CV0010	–	<i>Crotalus viridis</i> (CV1)	Female	Weld Co., CO	473,910,582	38.7	Schild et al. 2020
CV0011	–	<i>Crotalus viridis</i> (CV1)	Female	Weld Co., CO	629,971,926	57.4	Schild et al. 2020
CV0629	DRS208	<i>Crotalus viridis</i> (CV1)	Female	Weld Co., CO	325,201,850	30.9	Schild et al. 2020
CV0634	DRS214	<i>Crotalus viridis</i> (CV1)	Female	Weld Co., CO	351,684,726	29.4	Schild et al. 2020
CV0636	DRS151	<i>Crotalus viridis</i> (CV1)	Female	Weld Co., CO	326,721,986	27.4	Schild et al. 2020
CV0646	DRS224	<i>Crotalus viridis</i> (CV1)	Female	Weld Co., CO	229,649,748	21.8	Schild et al. 2020
CV0650	DRS210	<i>Crotalus viridis</i> (CV1)	Female	Weld Co., CO	559,341,904	52.9	Schild et al. 2020
CV0853	UTAR 65363	<i>Crotalus viridis</i> (CV2)	Male	Chouteau Co., MT	214,070,746	21.4	Schild et al. 2021
CV0854	UTAR 65372	<i>Crotalus viridis</i> (CV2)	Male	Chouteau Co., MT	293,358,764	29.3	Schild et al. 2021
CV0856	UTAR 65366	<i>Crotalus viridis</i> (CV2)	Male	Chouteau Co., MT	210,833,550	21.1	Schild et al. 2021
CV0860	UTAR 65372	<i>Crotalus viridis</i> (CV2)	Male	Chouteau Co., MT	311,940,018	31.2	Schild et al. 2021
CV0865	UTAR 65374	<i>Crotalus viridis</i> (CV2)	Male	Chouteau Co., MT	284,405,154	28.4	Schild et al. 2021
CV0867	UTAR 65376	<i>Crotalus viridis</i> (CV2)	Male	Chouteau Co., MT	92,906,802	9.3	Schild et al. 2021
CV0870	UTAR 65379	<i>Crotalus viridis</i> (CV2)	Male	Chouteau Co., MT	237,632,590	23.8	Schild et al. 2021
CV0858	UTAR 65367	<i>Crotalus viridis</i> (CV2)	Female	Chouteau Co., MT	284,671,840	26.5	This study
CV0862	UTAR 65369	<i>Crotalus viridis</i> (CV2)	Female	Chouteau Co., MT	236,153,902	21.9	This study
CV0863	UTAR 65373	<i>Crotalus viridis</i> (CV2)	Female	Chouteau Co., MT	331,951,534	28.3	This study
CV0864	UTAR 65370	<i>Crotalus viridis</i> (CV2)	Female	Chouteau Co., MT	196,275,972	16.9	This study
CV0868	UTAR 65377	<i>Crotalus viridis</i> (CV2)	Female	Chouteau Co., MT	297,928,908	27.9	This study
CV0869	UTAR 65378	<i>Crotalus viridis</i> (CV2)	Female	Chouteau Co., MT	429,867,086	40.8	This study
CV0859	UTAR 65371	<i>Crotalus viridis</i> (CV2)	Female	Chouteau Co., MT	95,255,394	8.8	This study
CV0087	CAS 205756	<i>Crotalus oreganus</i> (CO1)	Male	Alameda Co., MT	185,029,038	18.5	Schild et al. 2020
CV0094	CAS 208761	<i>Crotalus oreganus</i> (CO1)	Male	Fresno Co., CA	406,368,080	40.6	Schild et al. 2020
CV0096	CAS 208764	<i>Crotalus oreganus</i> (CO1)	Male	Fresno Co., CA	153,473,574	15.3	Schild et al. 2020
CV0105	CAS 209200	<i>Crotalus oreganus</i> (CO1)	Male	Mariposa Co., CA	279,548,320	28.0	Schild et al. 2020
CV0136	CAS 224859	<i>Crotalus oreganus</i> (CO1)	Male	Fresno Co., CA	401,904,228	40.2	Schild et al. 2020
CV0145	CAS 228193	<i>Crotalus oreganus</i> (CO1)	Male	Santa Clara Co., CA	414,417,224	41.4	Schild et al. 2020
CV0150	CAS 235854	<i>Crotalus oreganus</i> (CO1)	Male	Alameda Co., CA	128,358,806	12.8	Schild et al. 2020
CV0151	CAS 235855	<i>Crotalus oreganus</i> (CO1)	Male	Alameda Co., CA	251,835,256	25.2	Schild et al. 2020
CV0085	CAS 202983	<i>Crotalus oreganus</i> (CO1)	Female	Monterey Co., CA	308,494,718	18.0	Schild et al. 2020
CV0093	CAS 206480	<i>Crotalus oreganus</i> (CO1)	Female	Tulare Co., CA	275,995,910	24.1	Schild et al. 2020
CV0095	CAS 208762	<i>Crotalus oreganus</i> (CO1)	Female	Fresno Co., CA	249,494,118	21.2	Schild et al. 2020
CV0098	CAS 208785	<i>Crotalus oreganus</i> (CO1)	Female	Fresno Co., CA	238,932,864	19.4	Schild et al. 2020
CV0148	CAS 234626	<i>Crotalus oreganus</i> (CO1)	Female	Butte Co., CA	256,879,906	22.8	Schild et al. 2020
CV0152	CAS 236038	<i>Crotalus oreganus</i> (CO1)	Female	Alameda Co., CA	267,339,058	23.8	Schild et al. 2020
CV0153	CAS 236216	<i>Crotalus oreganus</i> (CO1)	Female	Kern Co., CA	295,339,668	17.9	Schild et al. 2020
CV0155	CAS 241772	<i>Crotalus oreganus</i> (CO1)	Female	Alameda Co., CA	267,401,784	23.8	Schild et al. 2020
CV0157	CAS 252894	<i>Crotalus oreganus</i> (CO1)	Female	Fresno Co., CA	306,179,992	20.1	Schild et al. 2020
CV0764	UTAR 65292	<i>Crotalus oreganus</i> (CO2)	Male	Nez Perce Co., ID	322,894,588	32.3	Schild et al. 2021
CV0766	UTAR 65298	<i>Crotalus oreganus</i> (CO2)	Male	Nez Perce Co., ID	153,398,436	15.3	Schild et al. 2021
CV0770	UTAR 65297	<i>Crotalus oreganus</i> (CO2)	Male	Nez Perce Co., ID	135,075,102	13.5	Schild et al. 2021
CV0772	UTAR 65299	<i>Crotalus oreganus</i> (CO2)	Male	Nez Perce Co., ID	285,987,580	28.6	Schild et al. 2021
CV0775	UTAR 65302	<i>Crotalus oreganus</i> (CO2)	Male	Nez Perce Co., ID	407,383,892	40.7	Schild et al. 2021
CV0781	UTAR 65310	<i>Crotalus oreganus</i> (CO2)	Male	Nez Perce Co., ID	282,050,038	28.2	Schild et al. 2021
CV0790	UTAR 65319	<i>Crotalus oreganus</i> (CO2)	Male	Nez Perce Co., ID	174,159,484	17.4	Schild et al. 2021
CV0780	UTAR 65309	<i>Crotalus oreganus</i> (CO2)	Female	Nez Perce Co., ID	159,964,308	14.6	This study
CV0783	UTAR 65312	<i>Crotalus oreganus</i> (CO2)	Female	Nez Perce Co., ID	336,933,840	31.3	This study
CV0784	UTAR 65313	<i>Crotalus oreganus</i> (CO2)	Female	Nez Perce Co., ID	404,449,490	37.5	This study
CV0786	UTAR 65315	<i>Crotalus oreganus</i> (CO2)	Female	Nez Perce Co., ID	135,327,384	12.4	This study
CV0787	UTAR 65303	<i>Crotalus oreganus</i> (CO2)	Female	Nez Perce Co., ID	233,182,368	21.6	This study
CV0793	UTAR 65324	<i>Crotalus oreganus</i> (CO2)	Female	Nez Perce Co., ID	292,582,498	27.2	This study
CV0796	UTAR 65305	<i>Crotalus oreganus</i> (CO2)	Female	Nez Perce Co., ID	247,413,118	22.6	This study
CV0798	UTAR 65326	<i>Crotalus oreganus</i> (CO2)	Female	Nez Perce Co., ID	250,669,310	23.5	This study
CV0800	UTAR 65327	<i>Crotalus oreganus</i> (CO2)	Female	Nez Perce Co., ID	193,568,674	18.0	This study

Supplementary Table 2. Population genomic summary statistics (mean \pm standard deviation) and results of Mann-Whitney U tests comparing distributions in major venom gene regions to chromosome-specific backgrounds within and between populations.

Family	Statistic	Population	Venom Region	Chromosome	Non-venom homologs
SVMP	π	CV1	0.0041 \pm 0.0079	0.0033 \pm 0.0121*	0.0037 \pm 0.0017
		CV2	0.0043 \pm 0.021	0.0026 \pm 0.0094**	0.0029 \pm 0.0016
		CO1	0.009 \pm 0.022	0.003 \pm 0.0085**	0.003 \pm 0.0015**
		CO2	0.0101 \pm 0.02	0.0027 \pm 0.0067**	0.0031 \pm 0.0018**
	d_{xy}	CV1 vs CO1	0.013 \pm 0.019	0.0068 \pm 0.0076**	0.0063 \pm 0.0021**
		CV1 vs CV2	0.005 \pm 0.017	0.0031 \pm 0.0066**	0.0035 \pm 0.0017
		CO1 vs CO2	0.01 \pm 0.0179	0.0036 \pm 0.006**	0.0041 \pm 0.0021**
	F_{st}	CV1 vs CO1	0.45 \pm 0.21	0.499 \pm 0.221**	0.53 \pm 0.14*
		CV1 vs CV2	0.071 \pm 0.092	0.066 \pm 0.081	0.076 \pm 0.066
		CO1 vs CO2	0.12 \pm 0.13	0.189 \pm 0.142**	0.27 \pm 0.13**
SVSP	π	CV1	0.0077 \pm 0.032	0.0039 \pm 0.008**	0.0029 \pm 0.002*
		CV2	0.006 \pm 0.028	0.0032 \pm 0.0107*	0.0024 \pm 0.0018
		CO1	0.0054 \pm 0.0086	0.0036 \pm 0.0151**	0.0019 \pm 0.0013**
		CO2	0.0071 \pm 0.0151	0.0033 \pm 0.0104**	0.0024 \pm 0.0019**
	d_{xy}	CV1 vs CO1	0.0094 \pm 0.0096	0.0071 \pm 0.01**	0.005 \pm 0.0023**
		CV1 vs CV2	0.0067 \pm 0.0246	0.0037 \pm 0.0072**	0.0028 \pm 0.0019
		CO1 vs CO2	0.007 \pm 0.011	0.0042 \pm 0.0077**	0.0026 \pm 0.0018**
	F_{st}	CV1 vs CO1	0.38 \pm 0.235	0.472 \pm 0.2544**	0.6 \pm 0.2**
		CV1 vs CV2	0.045 \pm 0.089	0.0621 \pm 0.0962**	0.074 \pm 0.071**
		CO1 vs CO2	0.148 \pm 0.17	0.189 \pm 0.1693**	0.23 \pm 0.13**
PLA2	π	CV1	0.0034 \pm 0.0034	0.0038 \pm 0.0062	0.0018 \pm 0.0012
		CV2	0.0033 \pm 0.0047	0.003 \pm 0.006	0.0016 \pm 0.0011
		CO1	0.0062 \pm 0.0041	0.0034 \pm 0.0033**	0.002 \pm 0.0014**
		CO2	0.0017 \pm 0.0024	0.0029 \pm 0.0038*	0.0013 \pm 0.0012
	d_{xy}	CV1 vs CO1	0.0122 \pm 0.0082	0.0071 \pm 0.007**	0.0044 \pm 0.0021**
		CV1 vs CV2	0.0038 \pm 0.0043	0.0036 \pm 0.0061	0.0017 \pm 0.0011
		CO1 vs CO2	0.0075 \pm 0.0058	0.0041 \pm 0.004**	0.0027 \pm 0.0018*
	F_{st}	CV1 vs CO1	0.55 \pm 0.199	0.47 \pm 0.2208*	0.6 \pm 0.22
		CV1 vs CV2	0.12 \pm 0.12	0.062 \pm 0.0924**	0.042 \pm 0.098*
		CO1 vs CO2	0.32 \pm 0.216	0.21 \pm 0.1798**	0.45 \pm 0.24

* $P < 0.05$; ** $P < 0.001$; Abbreviations: CV1 (Colorado *C. viridis*), CV2 (Montana *C. viridis*), CO1 (California *C. oregonus*), CO2 (Idaho *C. oregonus*), SVMP (snake venom metalloproteinase), SVSP (snake venom serine protease), PLA2 (phospholipase A2).

Supplementary Table 3. Mean population genetic diversity and differentiation estimates for venom genes in the snake venom metalloproteinase (SVMP), snake venom serine protease (SVSP), and phospholipase A2 (PLA2) gene families.

Gene	Chr.	Start (bp)	End (bp)	π				d_{xy}			F_{st}		
				CV1	CV2	CO1	CO2	CV1 vs CO1	CV1 vs CV2	CO1 vs CO2	CV1 vs CO1	CV1 vs CV2	CO1 vs CO2
SVMP 1	9	13,901,005	14,014,581	0.0049	0.0035	0.0091	0.0093	0.0131	0.0046	0.0099	0.43	0.08	0.10
SVMP 2	9	14,021,684	14,075,705	0.0043	0.0031	0.0067	0.0074	0.0082	0.0038	0.0075	0.28	0.05	0.07
SVMP 3	9	14,091,593	14,113,017	0.0042	0.0027	0.0117	0.0127	0.0146	0.0038	0.0130	0.43	0.09	0.09
SVMP 4	9	14,147,738	14,170,909	0.0037	0.0033	0.0135	0.0149	0.0148	0.0036	0.0153	0.39	0.07	0.12
SVMP 5	9	14,174,795	14,190,699	0.0042	0.0038	0.0112	0.0160	0.0136	0.0043	0.0140	0.50	0.06	0.16
SVMP 6	9	14,211,286	14,242,814	0.0047	0.0045	0.0099	0.0143	0.0148	0.0049	0.0120	0.45	0.06	0.05
SVMP 7	9	14,248,544	14,273,221	0.0027	0.0028	0.0106	0.0126	0.0190	0.0028	0.0120	0.57	0.05	0.08
SVMP 8	9	14,281,167	14,301,343	0.0026	0.0039	0.0114	0.0134	0.0176	0.0037	0.0125	0.56	0.08	0.05
SVMP 9	9	14,310,448	14,338,786	0.0027	0.0034	0.0117	0.0119	0.0324	0.0036	0.0119	0.60	0.16	0.05
SVMP 10	9	14,368,022	14,394,021	0.0029	0.0031	0.0093	0.0120	0.0131	0.0033	0.0106	0.47	0.07	0.05
SVMP 11	9	14,401,246	14,424,729	0.0024	0.0017	0.0037	0.0043	0.0059	0.0021	0.0050	0.46	0.04	0.19
SVSP 1	10	8,568,727	8,575,991	0.0076	0.0037	0.0087	0.0118	0.0152	0.0058	0.0102	0.35	0.03	0.02
SVSP 2	10	8,587,665	8,593,660	0.0070	0.0043	0.0046	0.0063	0.0081	0.0058	0.0056	0.26	0.04	0.04
SVSP 3	10	8,627,539	8,636,885	0.0044	0.0024	0.0065	0.0088	0.0132	0.0034	0.0078	0.61	0.06	0.10
SVSP 4	10	8,664,603	8,670,797	0.0053	0.0025	0.0038	0.0054	0.0064	0.0040	0.0048	0.21	0.09	0.07
SVSP 10	10	8,707,633	8,717,662	0.0178	0.0152	0.0059	0.0096	0.0143	0.0162	0.0080	0.26	0.00	0.10
SVSP 5	10	8,739,986	8,745,649	0.0124	0.0087	0.0066	0.0081	0.0123	0.0107	0.0075	0.26	0.03	0.06
SVSP 6	10	8,751,755	8,759,449	0.0096	0.0058	0.0074	0.0077	0.0138	0.0078	0.0082	0.37	0.03	0.07
SVSP 11	10	8,786,671	8,797,377	0.0031	0.0031	0.0040	0.0068	0.0054	0.0031	0.0060	0.25	0.00	0.14
SVSP 7	10	8,864,549	8,879,456	0.0044	0.0034	0.0053	0.0073	0.0096	0.0040	0.0075	0.50	0.03	0.26
SVSP 8	10	8,937,064	8,948,017	0.0082	0.0082	0.0066	0.0107	0.0129	0.0082	0.0103	0.40	0.01	0.14
SVSP 9	10	8,959,437	8,981,362	0.0046	0.0043	0.0036	0.0040	0.0087	0.0045	0.0076	0.43	0.01	0.40
PLA2 B1	15	3,027,530	3,029,405	0.0023	0.001	0.0082	0.0015	0.0122	0.0029	0.0075	0.55	0.33	0.09
PLA2 K	15	3,031,033	3,033,792	0.0008	0.001	0.0037	0.0043	0.0100	0.0009	0.0056	0.80	0.03	0.15
PLA2 C1	15	3,037,041	3,038,745	0.0053	0.006	0.0144	0.0003	0.0249	0.0061	0.0163	0.58	0.05	0.49
PLA2 A1	15	3,041,892	3,043,778	0.0133	0.017	0.0121	0.0023	0.0240	0.0165	0.0156	0.47	0.07	0.45

Supplementary Table 4. π and d_{xy} statistics (mean \pm standard deviation) and results of Mann-Whitney U tests comparing estimates in venom genes to intergenic regions in SVMP, SVSP, and PLA2 regions.

Family	Statistic	Population	Venom Genes	Intergenic Regions
SVMP	π	CV1	0.0039 \pm 0.0071	0.0044 \pm 0.0092
		CV2	0.0033 \pm 0.0066	0.0059 \pm 0.0331
		CO1	0.0095 \pm 0.0077	0.0082 \pm 0.0339**
		CO2	0.0108 \pm 0.0088	0.0091 \pm 0.0316**
	d_{xy}	CV1 vs CO1	0.0144 \pm 0.0178	0.0107 \pm 0.0199**
		CV1 vs CV2	0.0039 \pm 0.0071	0.0058 \pm 0.0259
		CO1 vs CO2	0.0106 \pm 0.008	0.0091 \pm 0.0275**
SVSP	π	CV1	0.0072 \pm 0.0067	0.0079 \pm 0.0372**
		CV2	0.0055 \pm 0.0063	0.0062 \pm 0.0314*
		CO1	0.0055 \pm 0.0046	0.0053 \pm 0.0095*
		CO2	0.0076 \pm 0.0058	0.0069 \pm 0.0171**
	d_{xy}	CV1 vs CO1	0.0107 \pm 0.007	0.0089 \pm 0.0103**
		CV1 vs CV2	0.0064 \pm 0.0062	0.0068 \pm 0.0281**
		CO1 vs CO2	0.0077 \pm 0.0056	0.0067 \pm 0.0121**
PLA2	π	CV1	0.0054 \pm 0.0056	0.0027 \pm 0.0018
		CV2	0.0066 \pm 0.0075	0.0021 \pm 0.0023
		CO1	0.0091 \pm 0.0049	0.0051 \pm 0.0031*
		CO2	0.0024 \pm 0.0039	0.0016 \pm 0.0018
	d_{xy}	CV1 vs CO1	0.0172 \pm 0.0086	0.0102 \pm 0.0073*
		CV1 vs CV2	0.0066 \pm 0.0069	0.0028 \pm 0.0023
		CO1 vs CO2	0.0112 \pm 0.0083	0.0062 \pm 0.0043

* $P < 0.05$; ** $P < 0.001$; Abbreviations: CV1 (Colorado *C. viridis*), CV2 (Montana *C. viridis*), CO1 (California *C. oregonus*), CO2 (Idaho *C. oregonus*), SVMP (snake venom metalloproteinase), SVSP (snake venom serine protease), PLA2 (phospholipase A2).

Supplementary Table 5. Summary of selection statistics (mean \pm standard deviation) and results of tests comparing distributions in major venom gene regions to chromosome-specific and non-venom homolog backgrounds. Tajima's D and $|iHS|$ distributions were compared using Welch's two-sample t -tests. d_f and β distributions were compared using Mann-Whitney U tests.

Family	Statistic	Population	Venom region	Chromosome	P	Non-venom homologs	P
SVMP	Tajima's D	CV1	0.27 \pm 0.86	-0.01 \pm 0.76*	0.01312	0.03 \pm 0.73	0.1386
		CV2	0.93 \pm 0.69	0.87 \pm 0.87	0.4621	1.03 \pm 0.87	0.5123
		CO1	1.18 \pm 0.98	-0.07 \pm 0.79**	2.3×10^{-14}	-0.07 \pm 0.74**	9.53×10^{-11}
		CO2	1.51 \pm 0.74	1.1 \pm 1.05**	0.00012	1.04 \pm 1.12*	0.0241
	d_f	CV1 vs CO1	0.014 \pm 0.041	0.026 \pm 0.038**	1.07×10^{-6}	0.014 \pm 0.048**	0.0002
	$ iHS $	CV1	0.95 \pm 0.44	0.32 \pm 0.27***	$< 2.2 \times 10^{-16}$	0.31 \pm 0.24**	3.13×10^{-15}
		CO1	0.76 \pm 0.37	0.32 \pm 0.25**	4.12×10^{-14}	0.38 \pm 0.34**	1.84×10^{-6}
	β	CV1	2.71 \pm 4.24	0.62 \pm 0.74**	0.00015	0.8 \pm 0.69	0.1158
		CO1	3.42 \pm 2.52	0.55 \pm 0.77***	$< 2.2 \times 10^{-16}$	0.8 \pm 0.86**	4.4×10^{-10}
	SVSP	Tajima's D	CV1	0.36 \pm 0.94	0.06 \pm 0.78*	0.0309	-0.02 \pm 0.86*
CV2			0.65 \pm 0.93	0.64 \pm 0.95	0.9879	0.83 \pm 0.83	0.2631
CO1			0.54 \pm 0.78	0.01 \pm 0.89**	2.03×10^{-5}	-0.25 \pm 1.0**	3.24×10^{-6}
CO2			1.01 \pm 0.74	1.11 \pm 1.04	0.3234	0.87 \pm 1.44	0.4792
d_f		CV1 vs CO1	0.01 \pm 0.036	0.017 \pm 0.028**	2.67×10^{-5}	0.038 \pm 0.093*	0.0027
$ iHS $		CV1	0.94 \pm 0.37	0.34 \pm 0.3**	1.48×10^{-14}	0.36 \pm 0.25**	5.84×10^{-14}
		CO1	0.69 \pm 0.37	0.33 \pm 0.3**	4.78×10^{-7}	0.37 \pm 0.25**	1.37×10^{-5}
β		CV1	2.83 \pm 2.85	0.7 \pm 1.02**	3.44×10^{-8}	0.72 \pm 0.8**	2×10^{-6}
		CO1	2.04 \pm 1.84	0.59 \pm 1.05**	4.48×10^{-8}	0.54 \pm 0.7**	1.55×10^{-7}
PLA2		Tajima's D	CV1	-0.2 \pm 0.82	0.04 \pm 1.01	0.077	-0.25 \pm 0.88
	CV2		1.33 \pm 1.41	0.66 \pm 1.06*	0.03	-0.05 \pm 1.03**	6.13×10^{-6}
	CO1		0.82 \pm 1.19	-0.01 \pm 1.08**	5.74×10^{-5}	-0.29 \pm 1.11**	3.43×10^{-5}
	CO2		0.45 \pm 1.46	0.89 \pm 1.22*	0.0102	0.29 \pm 1.22	0.6052
	d_f	CV1 vs CO1	0.047 \pm 0.078	0.033 \pm 0.08*	0.0442	0.073 \pm 0.085	0.2417
	$ iHS $	CV1	0.7 \pm 0.41	0.29 \pm 0.24**	2.77×10^{-7}	0.63 \pm 0.3	0.5231
		CO1	0.48 \pm 0.34	0.27 \pm 0.21*	0.0011	0.57 \pm 0.41	0.4466
	β	CV1	0.77 \pm 2.41	0.51 \pm 0.75	0.2422	0.08 \pm 0.22	0.6283
		CO1	1.93 \pm 2.41	0.46 \pm 0.75**	0.00047	0.31 \pm 0.4*	0.0159

* $P < 0.05$; ** $P < 0.001$; *** $P < 2.2 \times 10^{-16}$; Abbreviations: CV1 (Colorado *C. viridis*), CV2 (Montana *C. viridis*), CO1 (California *C. oreganus*), CO2 (Idaho *C. oreganus*), SVMP (snake venom metalloproteinase), SVSP (snake venom serine protease), PLA2 (phospholipase A2).

Supplementary Table 6. Mean estimates for population genetic statistics used to test for signatures of natural selection on venom genes in the snake venom metalloproteinase (SVMP), snake venom serine protease (SVSP), and phospholipase A2 (PLA2) gene families.

Gene	Chr.	Start (bp)	End (bp)	Tajima's D				d_f	$ iHS $		β	
				CV1	CV2	CO1	CO2	CV1 vs CO1	CV1	CO1	CV1	CO1
SVMP 1	9	13,901,005	14,014,581	0.641	0.650	1.248	1.368	0.0119	0.800	0.660	2.418	2.011
SVMP 2	9	14,021,684	14,075,705	0.169	0.260	0.554	0.694	0.0008	1.147	0.864	2.551	1.557
SVMP 3	9	14,091,593	14,113,017	0.631	1.117	2.151	1.879	0.0127	0.751	0.601	-0.92	1.633
SVMP 4	9	14,147,738	14,170,909	0.296	0.719	1.002	1.648	0.0070	1.053	0.821	1.071	4.201
SVMP 5	9	14,174,795	14,190,699	-0.05	0.151	1.552	1.725	0.0135	1.140	0.565	-0.05	3.571
SVMP 6	9	14,211,286	14,242,814	0.365	0.358	1.414	1.253	0.0191	1.123	1.219	-0.51	4.790
SVMP 7	9	14,248,544	14,273,221	0.341	1.041	1.238	2.032	0.0504	1.306	0.800	0.762	1.470
SVMP 8	9	14,281,167	14,301,343	-0.698	0.437	1.061	1.162	0.0293	1.254	0.810	-0.44	2.753
SVMP 9	9	14,310,448	14,338,786	0.022	0.831	2.051	2.220	0.0352	1.253	0.624	-1.13	2.398
SVMP 10	9	14,368,022	14,394,021	-0.127	0.627	0.483	0.904	0.0100	0.694	0.745	-0.68	2.709
SVMP 11	9	14,401,246	14,424,729	-0.644	0.582	0.963	1.521	0.0058	0.329	0.22	-0.03	0.829
SVSP 1	10	8,568,727	8,575,991	-0.744	0.515	-0.19	0.720	0	1.249	-	-0.15	3.331
SVSP 2	10	8,587,665	8,593,660	-0.023	0.528	-0.29	-0.07	0	1.421	-	1.972	1.052
SVSP 3	10	8,627,539	8,636,885	-0.250	0.176	0.416	0.833	0.0734	0.645	0.256	0.156	1.980
SVSP 4	10	8,664,603	8,670,797	0.3795	-0.16	-0.16	0.173	0.0067	0.841	0.677	1.218	0.180
SVSP 10	10	8,707,633	8,717,662	1.0069	1.524	0.314	0.773	0.0011	0.558	0.739	3.946	1.608
SVSP 5	10	8,739,986	8,745,649	0.6137	1.677	0.157	1.344	0	1.114	0.918	4.252	1.443
SVSP 6	10	8,751,755	8,759,449	0.1516	0.888	0.618	0.575	0.0102	0.904	0.854	1.068	1.408
SVSP 11	10	8,786,671	8,797,377	0.0625	-0.23	0.143	0.131	0.0013	0.864	1.437	0.419	0.922
SVSP 7	10	8,864,549	8,879,456	0.1973	0.197	0.794	1.046	0.0100	0.940	0.582	3.812	1.080
SVSP 8	10	8,937,064	8,948,017	0.3645	0.382	0.685	1.353	0.0006	0.774	0.899	1.909	1.430
SVSP 9	10	8,959,437	8,981,362	0.6585	0.680	0.125	1.433	0.0180	0.944	0.945	1.639	1.092
PLA2 B1	15	3,027,530	3,029,405	-0.25	0.55	0.970	0.823	0.0623	0.842	0.698	-0.34	4.607
PLA2 K	15	3,031,033	3,033,792	-1.32	0.74	1.201	2.925	0.0991	0.663	0.122	0.215	5.879
PLA2 C1	15	3,037,041	3,038,745	0.745	2.88	1.497	-0.85	0.0621	1.019	0.688	0.124	4.920
PLA2 A1	15	3,041,892	3,043,778	0.351	3.09	-0.51	0.389	0	1.055	0.960	4.979	1.558

Entries denoted with "--" denote genes without sufficient data to calculate a mean estimate.

Supplementary Table 8. Distributions (mean \pm standard deviation) of population-scaled recombination rate ($\rho = 4N_e r$) corrected for effective population size (ρ/π) and results of Mann-Whitney U tests comparing distributions in major venom gene regions to chromosome-specific backgrounds within and between populations.

Family	Species	Venom Region ρ/π	Chromosome ρ/π	P	Non-venom homolog ρ/π	P
SVMP	<i>C. viridis</i>	18.0492 \pm 18.83205	1.581972 \pm 4.340345***	$< 2.2 \times 10^{-16}$	1.420825 \pm 1.818915**	2.15×10^{-10}
	<i>C. oreganus</i>	16.65499 \pm 20.59502	6.233693 \pm 20.76791**	2.57×10^{-9}	8.117879 \pm 21.14086**	1.36×10^{-5}
SVSP	<i>C. viridis</i>	6.18022 \pm 6.335619	3.415031 \pm 31.5544**	8.54×10^{-12}	0.7159652 \pm 1.457774**	2.41×10^{-12}
	<i>C. oreganus</i>	15.70336 \pm 23.24486	8.829342 \pm 60.75674**	1.11×10^{-7}	2.059375 \pm 5.006803**	7.09×10^{-10}
PLA2	<i>C. viridis</i>	6.825836 \pm 5.838167	3.28576 \pm 22.89899*	0.0047	2.067746 \pm 3.206935	0.1429
	<i>C. oreganus</i>	2.17419 \pm NA	6.516126 \pm 23.09414	0.7639	0.7321676 \pm 0.4525945	0.5

* $P < 0.05$; ** $P < 0.001$; *** $P < 2.2 \times 10^{-16}$. Abbreviations: SVMP (snake venom metalloproteinase), SVSP (snake venom serine protease), PLA2 (phospholipase A2).

Supplementary Table 9. Population genomic summary statistics (mean \pm standard deviation) across the genome within and between populations.

Statistic	Population	Genome	Macro	Micro	Z	PAR
π	CO <i>C. viridis</i> (CV1)	0.00270 \pm 0.00148	0.00260 \pm 0.00121	0.00372 \pm 0.00161	0.00229 \pm 0.00264	0.00495 \pm 0.00200
	MT <i>C. viridis</i> (CV2)	0.00214 \pm 0.00128	0.00204 \pm 0.00100	0.00296 \pm 0.00134	0.00193 \pm 0.00251	0.00395 \pm 0.00164
	CA <i>C. oreganus</i> (CO1)	0.00210 \pm 0.00150	0.00197 \pm 0.00125	0.00335 \pm 0.00167	0.00158 \pm 0.00222	0.00517 \pm 0.00246
	ID <i>C. oreganus</i> (CO2)	0.00199 \pm 0.00140	0.00189 \pm 0.00117	0.00298 \pm 0.00162	0.00162 \pm 0.00221	0.00410 \pm 0.00226
d_{xy}	CV1 vs CO1	0.00516 \pm 0.00208	0.00502 \pm 0.00180	0.00700 \pm 0.00233	0.00374 \pm 0.00243	0.00950 \pm 0.00236
	CV1 vs CV2	0.00256 \pm 0.00142	0.00246 \pm 0.00115	0.00355 \pm 0.00154	0.00216 \pm 0.00250	0.00468 \pm 0.00184
	CO1 vs CO2	0.00260 \pm 0.00171	0.00245 \pm 0.00146	0.00406 \pm 0.00195	0.00187 \pm 0.00223	0.00605 \pm 0.00249
F_{st}	CV1 vs CO1	0.61314 \pm 0.11885	0.61868 \pm 0.10814	0.56096 \pm 0.11457	0.63437 \pm 0.18520	0.50444 \pm 0.11042
	CV1 vs CV2	0.07160 \pm 0.04987	0.07258 \pm 0.04988	0.07466 \pm 0.04193	0.05841 \pm 0.05659	0.06691 \pm 0.05416
	CO1 vs CO2	0.26171 \pm 0.09864	0.25971 \pm 0.09316	0.25948 \pm 0.07455	0.28409 \pm 0.15800	0.26421 \pm 0.09590

Abbreviations: CO (Colorado), MT (Montana), CA (California), ID (Idaho), Macro (Macrochromosome), Micro (Microchromosome), Z (Z chromosome), PAR (pseudoautosomal region).

Supplementary Table 10. Spearman's rank order correlation coefficients for pairwise comparisons of genome-wide nucleotide diversity (π), sequence divergence (d_{xy}), relative differentiation (F_{st}), and recombination rate (ρ).

		π				d_{xy}			F_{st}			ρ	
		CV1	CV2	CO1	CO2	CV1 vs CO1	CV1 vs CV2	CO1 vs CO2	CV1 vs CO1	CV1 vs CV2	CO1 vs CO2	<i>C. viridis</i>	<i>C. oregonus</i>
π	CV1	-	0.95***	0.81***	0.71***	0.83***	0.99**	-	-0.77***	-0.12***	-	0.62***	-
	CV2		-	0.79***	0.7***	-	0.98***	-	-	-0.21***	-	0.6***	-
	CO1			-	0.82***	0.89***	-	0.98***	-0.71***	-	-0.22***	-	0.52***
	CO2				-	-	-	0.93***	-	-	-0.37***	-	0.47***
d_{xy}	CV1 vs CO1					-	0.8***	0.83***	-0.47***	-	-	0.54***	0.5***
	CV1 vs CV2						-	0.8***	-	-0.09***	-	0.62***	-
	CO1 vs CO2							-	-	-0.16***	-	0.52***	
F_{st}	CV1 vs CO1							-	-	0.07**	0.16***	-0.52***	-0.4***
	CV1 vs CV2								-	0.07**	0.016	-	-
	CO1 vs CO2									-	-	-0.09***	-
ρ	<i>C. viridis</i>										-	-	0.7***
	<i>C. oregonus</i>												-

* $P < 0.05$; ** $P < 0.001$; *** $P < 2.2 \times 10^{-16}$. Abbreviations: CV1 (Colorado *C. viridis*), CV2 (Montana *C. viridis*), CO1 (California *C. oregonus*), CO2 (Idaho *C. oregonus*). Population scaled recombination rates in *C. viridis* and *C. oregonus* were estimated in Schield et al. (2020).

Supplementary Table 11. Mean Tajima's D , $|iHS|$, and β estimates used to test for evidence of natural selection on venom genes outside of the three major venom gene families.

Gene	Chromosome	Start (bp)	End (bp)	Tajima's D		$ iHS $		β	
				CV1	CO1	CV1	CO1	CV1	CO1
CRISP 1	1	169423774	169434684	1.46809533	1.48518783	0.81586959	0.7473673	2.73109045*	1.72299732*
CRISP 2	1	169434958	169437996	0.565988	1.35402567	0.8721553	0.75029649	5.73674515*	8.16466193*
CTL	13	11650747	11653723	0.911765	0.037345	0.57106241	0.9805324*	1.18843846	0.28148556
Exonuclease 1	6	12590208	12591465	0.0651375	-1.1930435	0.74771177	0.0494522	-0.130813	-0.129036
Exonuclease 2	11	10271502	10274220	0.7097135	0.0126735	0.3480382	1.14773135*	0.6495825	0.32699759
Exonuclease 3	15	8097114	8103411	0.88239914	1.34493357	0.07828052	0.3172539	1.33397965	0.4143664
GC 1	1	256551622	256564040	-0.114097	-0.4021076	0.3249434	0.0473596	0.47790853	0.36929033
GC 2	15	5091107	5094268	-0.131117	-0.1245545	0.30510816	0.45135965	0.81441062	0.68863738
LAAO 1	2	4654769	4658293	0.37364334	0.7370306	0.30599423	0.46523523	0.99893237	1.1485214
LAAO 2	2	4658599	4661642	-0.25619	-0.750064	0.81853808	0.10109293	0.28963197	-0.0741022
LAAO 3	4	85461961	85468906	0.450976	0.80317233	0.75565685	0.36436517	13.1719324*	6.80484563*
VEGF 1	1	260248287	260272500	-0.2730748	-0.869612	0.3808586	0.3383085	0.16743545	0.13965021
VEGF 2	7	40288572	40327884	-0.1501861	-0.4902092	0.21405333	0.0813892	0.23186827	0.02239229
Vespryn	2	4377779	4385668	-0.3308535	0.2850291	1.15739012*	0.53385297	0.03390518	0.11654844

*Denotes values that exceeded the genome-wide 95th percentile. Abbreviations: CRISP = Cysteine-rich secretory protein, CTL = C-type lectin, GC = Glutaminyl cyclase, LAAO = L-amino acid oxidase, VEGF = Vascular endothelial growth factor.