HHS Public Access

Author manuscript

Nat Ecol Evol. Author manuscript; available in PMC 2023 September 01.

Published in final edited form as:

Nat Ecol Evol. 2022 September ; 6(9): 1367–1380. doi:10.1038/s41559-022-01829-5.

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

Drew R. Schield[1,2,✉], **Blair W. Perry**[1,3], **Richard H. Adams**[4], **Matthew L. Holding**[5], **Zachary L. Nikolakis**[1], **Siddharth S. Gopalan**[1], **Cara F. Smith**[6], **Joshua M. Parker**[7], **Jesse M. Meik**[8], **Michael DeGiorgio**[9], **Stephen P. Mackessy**[6], **Todd A. Castoe**[1,✉]

[1]Department of Biology, University of Texas at Arlington, Arlington, TX, USA.

[2]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA.

[3]School of Biological Sciences, Washington State University, Pullman, WA, USA.

[4]Department of Biological and Environmental Sciences, Georgia College and State University, Milledgeville, GA, USA.

[5]Life Science Institute, University of Michigan, Ann Arbor, MI, USA.

[6]School of Biological Sciences, University of Northern Colorado, Greeley, CO, USA.

[7]Life Science Department, Fresno City College, Fresno, CA, USA.

[8]Department of Biological Sciences, Tarleton State University, Stephenville, TX, USA.

[9]Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL, USA.

Abstract

The origin of snake venom involved duplication and recruitment of non-venom genes into venom systems. Several studies have predicted that directional positive selection has governed this process. Venom composition varies substantially across snake species and venom phenotypes are locally adapted to prey, leading to coevolutionary interactions between predator and prey. Venom origins and contemporary snake venom evolution may therefore be driven by fundamentally different selection regimes, yet investigations of population-level patterns of selection have been

✉**Correspondence and requests for materials** should be addressed to Drew R. Schield, drew.schield@colorado.edu or Todd A. Castoe. todd.castoe@uta.edu.

Author contributions

D.R.S. and T.A.C. designed the study. D.R.S., B.W.P., Z.L.N., S.S.G., C.F.S., J.M.P., J.M.M., S.P.M. and T.A.C. collected samples and generated data. D.R.S., B.W.P., R.H.A. and M.D. performed analyses. D.R.S. and T.A.C. wrote the manuscript with contributions from B.W.P., R.H.A., M.L.H. and M.D. All authors provided edits to the manuscript.

Code availability

Analysis scripts are available on GitHub (https://github.com/drewschield/venom_population_genomics).

Competing interests

The authors declare no competing interests.

Extended data is available for this paper at https://doi.org/10.1038/s41559-022-01829-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01829-5.

limited. Here, we use whole-genome data from 68 rattlesnakes to test hypotheses about the factors that drive genomic diversity and differentiation in major venom gene regions. We show that selection has resulted in long-term maintenance of genetic diversity within and between species in multiple venom gene families. Our findings are inconsistent with a dominant role of directional positive selection and instead support a role of long-term balancing selection in shaping venom evolution. We also detect rapid decay of linkage disequilibrium due to high recombination rates in venom regions, suggesting that venom genes have reduced selective interference with nearby loci, including other venom paralogues. Our results provide an example of long-term balancing selection that drives trans-species polymorphism and help to explain how snake venom keeps pace with prey resistance.

Venom is a key trophic adaptation of many snakes and has become an important model system for understanding ecological and evolutionary processes that underlie adaptive traits[1–3]. Genes that encode snake venoms originated through a process of gene duplication of non-toxin genes followed by neofunctionalization and venom gland-specific expression[4–8]. Subsequent tandem duplication events have expanded many of these venom gene families, and differentially expanded families represent major axes of variation in snake venom composition[9–12]. These processes have been hypothesized to involve episodes of directional positive selection, especially during their early stages. Indeed, numerous studies have documented evidence for episodic positive selection in snake venom evolution in comparisons of venom and non-venom paralogues, often using codon-model-based approaches (for example, refs.[13,14]). However, after the establishment of venom gene families, it is reasonable to expect that fundamentally different evolutionary forces may drive allelic variation in venom genes within populations, as venom evolution reaches a new equilibrium state dominated by ongoing predator–prey interactions. In other words, selective processes acting on venom genes may differ substantially over evolutionary timescales, between processes impacting their early origins and those shaping their contemporary maintenance (sensu ref.[15], ‘ultimate’ causes). This potential contrast between processes underlying venom adaptation at deep versus shallow timescales remains largely unexplored. While previous studies have focused primarily on deep-time processes, understanding processes acting on more recent timescales requires analyses of population genetic variation, which have been used in relatively few studies (for example, refs.[16–18]).

Empirical studies suggest that contemporary (shallow timescale) snake venom evolution is driven by complex predator–prey coevolutionary dynamics, in which toxicity to prey is fundamental to snake fitness[2,19–21]. In response to intense selection pressure imposed by snake predators, molecular mechanisms have evolved in prey species to resist the biological activities of venom components[22]. Studies characterizing local adaptation in venom and resistance phenotypes[19,23–25] and mechanisms for resistance in prey[26–30] have demonstrated coevolution between snake and prey populations across space and between species, underscoring the potential for frequency-dependent coevolution between these complex traits.

Despite the likelihood that these coevolutionary dynamics shape venom evolution, analyses between distantly related snake species have echoed the prevailing hypothesis that

venom evolves via strong directional positive selection (often referred to as 'positive Darwinian selection' in the literature; for example, refs.[13,31–33]). Most previous studies have compared venom genes in a phylogenetic context and interpreted excesses of non-synonymous substitutions as evidence for positive selection. Previous work has largely neglected alternative hypotheses, such as persistent versus episodic positive selection (but see ref.[13]) and has not clarified whether patterns of selection on venom are solely directional. In a recent study[16], the authors compared interspecific substitutions to intraspecific polymorphisms at synonymous and non-synonymous sites using McDonald–Kreitman tests[34] and found that venom genes have a higher frequency of non-synonymous polymorphisms than do 'housekeeping' genes. Taken together, prior studies have galvanized the view that venom evolution is driven by directional selection, often to the exclusion of other mechanisms such as balancing selection (but see ref.[18]). Accordingly, directional selection remains the presumed dominant process shaping venom variation at both deep and shallow evolutionary timescales, spanning the neofunctionalization of ancient gene duplicates[5,13] to contemporary adaptation in populations[16,17,35].

Directional positive selection leaves characteristic signatures in genomic variation at sites nearby targets of selection[36–38], providing testable predictions for patterns of genetic diversity in regions under directional selection. The analogy of 'selective sweeps' describes the effects of directional positive selection on linked neutral variation[39]. Here, neutral variation is reduced as haplotypes with the target of selection rapidly increase in frequency[40,41], creating 'valleys' of genetic diversity within populations surrounding targets of selection and associated 'peaks' of genetic differentiation between populations. If directional positive selection is the dominant force shaping venom phenotypes among closely related lineages, then we would expect to observe these patterns in venom gene regions. However, no prior studies have examined signatures of selection at linked sites across venom-encoding loci using population genomic data, leaving open the question of precisely which processes drive variation in venom genes at recent evolutionary timescales.

Although not previously demonstrated in venom genes, other gene complexes involved in coevolutionary dynamics have been shown to evolve under various forms of balancing selection[42,43]. Genes in the vertebrate major histocompatibility complex are arguably the best known examples of balancing selection maintaining adaptive genetic diversity over long time periods[44,45]. Others include genes underlying pathogen resistance[46] and self-incompatibility in plants[47]. Many prominent examples of genes under balancing selection also occur in tandem arrays similar to the architecture of multiple venom gene clusters[45,46,48]. In contrast to the prevailing assumption of directional selection, both the ecological context and genetic architecture of venom raise the possibility that venom gene regions may be under balancing selection. Indeed, maintenance of polymorphism in snake venom genes by balancing selection is a reasonable expectation if snake venom systems are involved in complex frequency-dependent predator–prey dynamics[49,50].

Like directional positive selection, balancing selection leaves characteristic signatures in regional genetic variation. Notably, balancing selection mechanisms increase and retain genetic diversity at linked sites[42,51,52]. In extreme cases, long-term balancing selection can maintain shared variation between species ('trans-species polymorphism'[53–55]), although

this phenomenon is thought to be rare in nature because it requires the maintenance of alleles over potentially millions of years[51,56]. Inferences of directional positive selection may also be confounded by other forms of selection if alternatives are not considered. For example, phylogenetic comparisons using d_N/d_S ratios (ω) can identify excess non-synonymous substitution due to directional positive selection for specific amino acid replacements ($\omega > 1$), yet long-term balancing selection can also produce this pattern if polymorphic variants that are sorting within populations are incorrectly assumed to be fixed[44], highlighting the importance of measuring population-level variation. Similarly, long-term balancing selection may confound inferences of directional selection in related tests examining synonymous and non-synonymous divergence and polymorphism[43].

Given the importance of linked variation for testing hypotheses about selection, integrated analysis of the structural organization and recombination landscape of venom tandem arrays is highly relevant to understanding the processes by which they evolve. Recombination can interact with selection to facilitate adaptation[37] through the erosion of linkage between targets of selection and nearby regions under alternative selection regimes[57,58]. Recombination further creates complements of alleles previously untested by selection[59]. While a correspondence between recombination rate and the efficiency of selection is predicted to broadly shape genetic diversity across the genome[60,61], whether venom evolution has been facilitated by high recombination rates remains unknown.

Our study was motivated by two primary aims. The first was to characterize population genetic diversity and differentiation associated with major venom gene regions. The second was to combine this information with additional measures of selection to test alternative predictions of neutrality, directional selection and balancing selection. We analysed a population genomic dataset for two broadly distributed members of the Western Rattlesnake Species Complex, the Prairie Rattlesnake (*Crotalus viridis*) and Northern Pacific Rattlesnake (*C. oreganus*), which share a common ancestor 3 million years ago[62] (Fig. 1a). We focus primarily on three venom gene families: snake venom metalloproteinases (SVMPs), snake venom serine proteases (SVSPs) and phospholipases A_2 (PLA2s), which are organized in tandem arrays on microchromosomes[12] and represent major axes of functional venom variation in these snakes[63]. We further interpret findings in the context of the genomic recombination landscape to understand how recombination and selection together shape venom evolution within and between related rattlesnake lineages. Our results highlight the role of balancing selection in adaptation across populations that has been sufficiently strong to maintain long-term genetic diversity and trans-species polymorphisms among major venom genes.

Results

We sequenced and analysed the genomes of 68 rattlesnakes from populations of *C. viridis* in Colorado and Montana and *C. oreganus* in California and Idaho (Fig. 1a and Supplementary Table 1), along with an outgroup (*C. atrox*). Populations from California and Idaho represent the subspecies *C. o. helleri* and *C. o. oreganus* sensu ref.[64]. Hereafter, we refer to *C. viridis* populations as CV1 (Colorado) and CV2 (Montana) and *C. oreganus* populations as CO1 (California *C. o. helleri*) and CO2 (Idaho *C. o. oreganus*) (Fig. 1a).

We mapped genome resequencing data to the *C. viridis* reference genome[12], with variant calling and filtering steps yielding 27,749,933 high-quality single nucleotide polymorphisms (SNPs). 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. 1) to ensure that bioinformatic artifacts from these sources of variation did not bias our inferences (Supplementary Information). 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.

Inferences of population structure clearly distinguish *C. viridis* from *C. oreganus* under a $K = 2$ model (Fig. 1c and Supplementary Information). Additional subdivision between CO1 (California *C. o. helleri*) and CO2 (Idaho *C. o. oreganus*) is inferred under $K = 3$, the best-supported model based on the cross-validation procedure. Under the $K = 4$ model, which has similar support (Supplementary Fig. 1), each of the four populations corresponds to a distinct cluster. This model distinguishes Montana and Colorado *C. viridis* populations (CV2 and CV1, respectively), although there are some Colorado individuals with low-probability assignment to the Montana cluster, suggesting weaker structure between *C. viridis* populations as a consequence of recent northern range expansion[65].

We used the pairwise sequentially Markovian coalescent (PSMC[66]) to infer population demographic histories and to inform forward-time simulations (described below). PSMC estimates indicate that each species experienced multiple expansion and contraction events within the last 10 million years (Fig. 1d), including contractions coincident with Pleistocene glacial cycles. Inferred population sizes coalesce roughly 3 million years ago, the estimated divergence time for the two species[62] (Fig. 1a).

Diversity and differentiation in major venom gene regions.

To test the predictions of directional versus balancing selection on major functional axes of venom variation, we dissected patterns of genetic diversity and differentiation in genomic regions housing SVMP, SVSP and PLA2 gene families (Fig. 2). The SVMP and SVSP venom gene clusters on chromosomes 9 and 10, respectively, exhibit local π peaks in each population (Fig. 2a,b and Extended Data Fig. 2). Estimates in these regions exceed chromosome-specific and genome-wide mean π values by >2.5-fold on average, and SVMP and SVSP π distributions are significantly higher than chromosomal backgrounds (Supplementary Table 2; Mann–Whitney U, $P < 0.05$), with multiple venom genes in each family showing elevated point estimates (Extended Data Fig. 3 and Supplementary Tables 2 and 3). SVMP and SVSP clusters are also associated with d_{xy} peaks between *C. viridis* and *C. oreganus* and intraspecies population pairs (Fig. 2a,b, Extended Data Fig. 2 and Supplementary Table 2), indicating the persistence of ancestral polymorphism in these regions. Accordingly, diversity peaks in SVMP and SVSP clusters correspond with local depressions in Fst (Fig. 2a,b and Extended Data Fig. 2). Collectively, the SVMP and SVSP regions have among the highest genetic diversity, highest sequence divergence and lowest relative differentiation in the genome, including a concentration of genetic

diversity in coding regions of these venom gene clusters (Supplementary Information and Supplementary Table 4).

We observe diversity estimates similar to the chromosomal background in the PLA2 region of chromosome 15 (Fig. 2c and Extended Data Fig. 2). However, detailed examinations across PLA2 genes in *C. viridis* populations show fine-scale variation in π, with low values across PLA2 B1 and PLA2 K, intermediate values across PLA2 C1 and high diversity peaks associated with PLA2 A1 (Fig. 2c, Extended Data Figs. 2 and 3 and Supplementary Table 3). There is also a PLA2 A1 π peak in the CO1 population, corresponding with high d_{xy} and low Fst between *C. viridis* and *C. oreganus* (Supplementary Table 2). In contrast, the CO2 population lacks a pronounced diversity peak across PLA2 A1 and Fst is elevated between *C. oreganus* populations. Similarly, intraspecific Fst is high for *C. viridis* across PLA2 B1, corresponding with low π in both CV1 and CV2 populations.

Signatures of selection in major venom gene regions.

Population genomic patterns in major venom gene regions indicate that ancestral variation has been maintained during the evolution of *C. viridis* and *C. oreganus* populations, suggesting a role for balancing selection. We quantified additional population genetic statistics across the three major venom gene regions to test expectations for patterns at sites linked to targets of directional versus balancing selection and to evaluate whether venom regions have evolved under alternative forms of selection versus the null hypothesis of neutrality.

The SVMP region exhibits multiple signatures of balancing selection, including positive Tajima's *D* indicating excess intermediate-frequency alleles in CV1 and CO1 populations (Fig. 3a). We focus on Tajima's *D* interpretations in CV1 and CO1, as northern populations show evidence of more extreme recent population size contractions (Fig. 1d), producing positively skewed genome-wide *D* distributions (Extended Data Fig. 4). SVMP *D* peaks in CV1 and CO1 stand out against negatively skewed genome-wide values and are significantly higher than chromosome 9 background distributions (>19-fold on average; Fig. 3a and Supplementary Tables 5 and 6; Welch's two-sample *t*, $P < 0.05$). The SVMP region also has a lower proportion of fixed differences, d_f, between *C. viridis* and *C. oreganus* than does the chromosome background (Fig. 3b and Supplementary Table 5; Mann–Whitney *U*, $P = 1 \times 10^{-6}$).

We find similar patterns in the SVSP region in which Tajima's *D* is on average 21-fold higher than background distributions in CV1 and CO1 (Fig. 3c and Supplementary Table 5; Welch's two-sample *t*, $P < 0.05$). This region also has extremely low d_f between species, significantly lower than background distributions (Fig. 3d; Mann–Whitney *U*, $P < 0.05$). The paucity of fixed differences between *C. viridis* and *C. oreganus*, together with elevated Tajima's *D* and observed patterns of genetic diversity and sequence divergence, is consistent with trans-species polymorphism contributing to intermediate allele frequencies in the SVMP and SVSP venom gene regions.

Additional statistics further reinforce the role of selection in the SVMP and SVSP regions (Fig. 3e–h). We measured the integrated haplotype statistic, |iHS| (ref.[67]) (Extended Data

Fig. 5), to detect longer-than-expected haplotypes due to selection across venom genes. Haplotype lengths depend on the background recombination rate, such that haplotypes will tend to be longer in regions of low recombination and shorter in regions of high recombination[68]. Importantly, |iHS| compares haplotypes within the same region of the genome, controlling for background recombination rate[67], and has high power when selected haplotypes are at intermediate frequencies[69]. The SVMP and SVSP regions are punctuated by outstanding |iHS| peaks in *C. viridis* and *C. oreganus* that are on average 2.5-fold higher in magnitude than chromosome-specific and non-venom homologue backgrounds (Fig. 3e,g, Supplementary Table 5 and Extended Data Figs. 5–7; Welch's two-sample t, $P < 0.001$).

Balancing selection produces clusters of linked sites with correlated allele frequencies surrounding balanced polymorphisms[51,70,71]. We quantified allele frequency correlation, $ß$ (refs.[71,72]), across the genome under a range of parameter settings (Extended Data Fig. 5 and Supplementary Table 7). Our results reveal high $ß$ in SVMP and SVSP regions of both species (Fig. 3f,h and Extended Data Figs. 6 and 7). Clustered $ß$ peaks in these regions are on average threefold higher than background distributions (Supplementary Table 5; Mann–Whitney U, $P < 0.001$, except between the SVMP region and non-venom homologues in *C. viridis*, $P = 0.12$). The $ß$ peaks highlight signatures of balancing selection across SVMP and SVSP regions and are consistent with additional evidence from the combination of elevated Tajima's D and |iHS| in both species. By contrast, we would expect low Tajima's D combined with an absence of $ß$ peaks under directional positive selection.

To explicitly test for evidence of balancing selection we performed model-based genome scans of a composite-likelihood ratio test statistic[73,74] to distinguish departures from neutral evolution. Higher ratios, measured as $B_{0,\mathrm{MAF}}$ scores, more strongly reject neutrality in favour of the alternative hypothesis of balancing selection, and we considered regions with scores above the genome-wide 95th quantile to be significant. The SVMP and SVSP regions each include multiple genes with outlier $B_{0,\mathrm{MAF}}$ scores that reject neutrality in one or both species (Fig. 3i–l, upper panels). Strong candidates under balancing selection in both species include SVMPs 1, 2, 3, 4, 5, 8 and 10 and SVSPs 7 and 9. Genes showing evidence of balancing selection in one species include SVSPs 5 and 10 in *C. viridis* and SVMPs 6, 7 and 9 in *C. oreganus*. Additional indicators of balancing selection are low $\log_{10}(\widehat{A})$ and high $\hat{x}$ parameters associated with high $B_{0,\mathrm{MAF}}$ scores, which measure balancing selection footprint size and equilibrium minor allele frequency, respectively (Fig. 3i–l, middle panels). Large footprints in these regions, indicated by low $\log_{10}(\widehat{A})$ values, provide very strong evidence of multilocus balancing selection[73,75]. Finally, the sign and magnitude of the dispersion parameter, $\hat{a}$, in high $B_{0,\mathrm{MAF}}$ regions indicate balancing selection as opposed to positive selection (Extended Data Fig. 8). It is notable that multiple venom gene regions with among the strongest evidence for balancing selection show little or no CNV presence (Fig. 3i–l), reinforcing that bioinformatic artifacts due to CNVs do not explain these findings (Supplementary Information).

The PLA2 region exhibits varied signals of selection (Extended Data Figs. 6–9). Tajima's D is higher than background distributions in *C. oreganus* (Welch's two-sample t, $P < 0.001$), but we find no significant differences in *C. viridis* (Supplementary Table 5). However,

fine-scale variation in D across *C. viridis* PLA2 genes aligns with variation in π and Fst. Negative D estimates for PLA2 B1 and PLA2 K contrast with positive D for PLA2 C1 and PLA2 A1 (Extended Data Fig. 9a and Supplementary Table 6). PLA2 A1 has significantly fewer fixed differences between *C. viridis* and *C. oreganus* than do non-venom homologues (Extended Data Fig. 9b and Supplementary Table 5; Mann–Whitney U, $P = 0.0064$). We observe, on average, 1.5-fold higher |iHS| across the PLA2 region than in background distributions (Extended Data Fig. 9c and Supplementary Table 5; Welch's two-sample t, $P < 0.05$). Finally, $ß$ varies across PLA2 genes; lower values for PLA2 B1, PLA2 K and PLA2 C1 contrast with a $ß$ peak over PLA2 A1 in *C. viridis* (Extended Data Fig. 9d and Supplementary Table 6). A similar PLA2 A1 $ß$ peak is present in *C. oreganus*, and overall PLA2 region $ß$ is >1.5-fold higher than background distributions in both species (Mann–Whitney U, $P < 0.05$). Explicit tests of balancing selection are equivocal, however, and do not reject neutrality across the PLA2 region in both species (Extended Data Figs. 8 and 9e,f). Small localized peaks of higher $B_{0,\mathrm{MAF}}$ scores are present across PLA2 C1 in *C. oreganus* and PLA2 A1 in both species, yet these scores fall below the genome-wide 95th quantile. It therefore remains an open question whether the PLA2 region has indeed experienced a mixture of positive and balancing selection, or perhaps there is reduced power to discriminate among processes, including neutrality, due to its comparatively small size.

To examine evidence for the persistence of trans-species polymorphism due to long-term balancing selection in venom gene regions, we quantified the relative proportions of fixed differences, private polymorphisms and trans-species polymorphisms among phased *C. viridis* and *C. oreganus* variants. We predict that long-term balancing selection will result in a higher frequency of trans-species polymorphisms relative to fixed differences between species. Indeed, we find that the SVMP and PLA2 venom gene regions are enriched for trans-species polymorphisms relative to chromosomal backgrounds (Fig. 4a; Fisher's exact, $P < 0.05$), and each major venom region has comparatively low proportions of fixed differences between *C. viridis* and *C. oreganus*, as also indicated by scans of d_f above. We then scanned venom exon alignments for trans-species amino acid polymorphisms, limiting analyses to exons without CNV evidence within or between species. We find 19, 24 and 6 trans-species amino acid polymorphisms among SVMP, SVSP and PLA2 genes, respectively, indicating that balanced polymorphisms are indeed relevant to venom protein variation (Fig. 4b).

Multiple specific forms of balancing selection could drive long-term maintenance of polymorphism, including negative frequency-dependent selection or heterozygote advantage (over-dominance). To discriminate among these forms, we trained a neural network classification model (Supplementary Fig. 2) with summary statistics calculated from simulations generated under neutrality and alternative balancing selection mechanisms based on inferred demographic histories (Fig. 1d). We used this model to predict the probabilities of neutrality, negative frequency-dependent selection and heterozygote advantage in genomic windows from summary statistics computed on empirical data. These analyses provide strong support for balancing selection in venom gene regions of both *C. viridis* and *C. oreganus* (Fig. 5a,b), with few windows showing higher probabilities of neutrality (Fig. 5c,d). Fine-scale variation in the probability of negative frequency dependence versus heterozygote advantage across these venom regions (Fig. 5e,f) further suggests that both

mechanisms have contributed to the maintenance of polymorphism in major venom gene regions.

Recombination rate variation in major venom gene regions.

Given multiple signatures of selection across major venom gene clusters, we investigated if the efficiency of selection and maintenance of variation have been facilitated by elevated levels of recombination. We first examined estimates of population-scaled recombination rate (ρ) in the SVMP, SVSP and PLA2 regions (Extended Data Fig. 10) and compared these to respective non-venom homologues and chromosomal backgrounds after dividing windowed estimates by nucleotide diversity (ρ/π) as a proxy for local N_e. The SVMP and SVSP regions show higher ρ/π than background distributions in both species (Fig. 6a–d and Supplementary Table 8; Mann–Whitney U, $P < 0.001$). PLA2 ρ/π is comparatively modest, although *C. viridis* PLA2 ρ/π is higher than in the chromosome 15 background (Fig. 6e,f; $P = 0.005$).

We hypothesized that high recombination rates have reduced selective interference between venom clusters and surrounding regions and potentially between venom paralogues in the larger SVMP and SVSP regions. To test this we compared the decay of linkage disequilibrium (LD) measured using r^2 as a function of physical distance between SNPs in venom regions to their immediate flanking regions. Indeed, LD decays more rapidly in the SVMP region in both species, decaying on average to 0.2 within 1.9 kilo-bases (kb) in *C. viridis* and 0.5 kb in *C. oreganus*, compared to 2.6 kb and 2.1 kb in flanking regions, respectively (Fig. 6a,b). Similarly, SVSP LD decays to 0.2 within 1.5 kb in *C. viridis* and 0.7 kb in *C. oreganus* compared to 6.1 kb and 4.3 kb in flanking regions (Fig. 6c, d). By contrast, we find longer blocks of LD on average in the PLA2 region of each species than in the immediate flanking regions (Fig. 6e,f). This finding makes sense considering the much tighter physical linkage of PLA2 genes compared to the larger SVMP and SVSP clusters (the PLA2 region spans 26 kb versus 440 kb and 605 kb SVSP and SVMP regions). Interestingly, evidence for rapid LD decay in flanking regions suggests that there may be less selective interference between venom PLA2s and neighbouring genes, which serve 'housekeeping' functions[76]. The existence of a recombination hotspot in *C. viridis* between PLA2gIIE (non-venom) and PLA2 B1 and the remaining venom cluster further supports this conclusion (Extended Data Fig. 10).

Finally, 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 LD[52]. We find no associations between variables in any of the major venom gene regions that suggest high recombination rate estimates are driven by balancing selection (Supplementary Information).

Discussion

Analyses of snake venom evolution that lack a broad genomic context with a bias toward deep-time comparisons have left substantial gaps in our understanding of population-level processes contributing to venom variation and adaptation. By analysing whole genomes

and comparing genotypes of entire venom multigene families in closely related populations (Fig. 1), this study provides new perspectives on venom evolution at shallow timescales, which we show may be shaped by fundamentally distinct evolutionary forces from those that shaped ancient venom gene origins. Our findings support a role of balancing selection in maintaining adaptive genetic diversity in major venom families and indicate remarkably little evidence for the widely accepted view that contemporary venom evolution is dominated by directional positive selection. Further, we find evidence that long-term balancing selection on venom gene regions has driven elevated levels of trans-species polymorphism, which is thought to be rare in nature[43], and that high regional recombination rates facilitate snake venom adaptation.

A role of balancing selection in venom evolution.

Classic signatures of directional selection poorly characterize variation observed in major rattlesnake venom gene regions, especially the large SVMP and SVSP tandem arrays. Instead, these venom loci manifest as readily distinguishable regions of elevated genetic diversity (Fig. 2 and Extended Data Fig. 2). Several genes in the smaller PLA2 region also harbour high genetic diversity (for example, PLA2 A1). In both *C. viridis* and *C. oreganus*, genes with elevated genetic diversity in major venom regions show additional signatures of balancing selection, including local concentrations of intermediate-frequency alleles, extended haplotype lengths and clusters of SNPs with correlated allele frequencies (Fig. 3 and Extended Data Fig. 9). The latter signature, in particular, is expected when balanced polymorphism has been maintained over moderately long evolutionary time periods (long-term balancing selection[42,70,71]). Model-based inferences further reject neutrality for multiple genes in the SVMP and SVSP clusters in both species (Fig. 3 and Extended Data Figs. 8 and 9), supporting that balancing selection mechanisms have shaped allelic diversity in these regions.

Different underlying selection regimes can drive signatures of balancing selection, such as negative frequency-dependent selection and heterozygote advantage, although distinguishing between these can be challenging[43]. Our simulation study to test evidence for distinct balancing selection mechanisms indicates that both negative frequency dependence and heterozygote advantage have probably contributed to venom diversity (Fig. 5), with regional variation in the probability of these alternative mechanisms. These conclusions align with known aspects of venom composition in snakes and the ecological context of predator–prey antagonism. Documented evidence of taxon-specific toxicity of venom protein isoforms in the same snake species lends support to heterozygote advantage at the functional level[77,78]. Various prey types are important in the diets of *C. oreganus* and *C. viridis*[79], with deer mice (*Peromyscus* sp.) and voles (*Microtus* sp.) representing large proportions of prey items[80]. If mouse- and vole-specific venom isoforms exist, for example, then the presence of these prey species in the diets of both species alone could exert overdominant selection and produce the observed signatures. Alternatively, alleles may be adaptive only until they reach certain frequencies (and prey potentially evolve resistance), at which point selection favours an alternative allele. Such frequency dependence is a strong expectation given evidence for coevolution between snake venom and prey resistance and aligns with the rapid turnover in venom composition at fine geographic scales observed in nature[19,81].

Our results do not entirely reject the role of directional selection in venom adaptation, as a subset of venom genes show signatures of directional selection (for example, PLA2 B1 and vespryn in *C. viridis* and CTL in *C. oreganus*; Supplementary Information). We also acknowledge 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 dominant effects of balancing selection at nearby genes. Still other venom genes have population genetic signatures that were indistinguishable from the genomic background and are probably sorting neutrally or evolving under purifying selection. These results suggest 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, which aligns with less dynamic evolution of minor venom gene families in pitvipers, generally[9,16,18,80,82,83].

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 genotype[84]. 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. Our results also provide evidence for the long-term maintenance of venom gene allelic diversity through balancing selection leading to trans-species amino acid polymorphisms (Fig. 4). These findings align with contemporary venom evolution being dominated by predator–prey coevolutionary dynamics that include prey evolving resistance to venom and snake venom evolving to circumvent this resistance. Our data therefore provide new population genomic evidence for mechanisms that explain the otherwise well-known[85] biological phenomenon of snake venom coevolution with prey resistance.

Recombination and selection shape venom adaptation.

It has remained unclear how compact tandem arrays of fitness-relevant venom loci could experience independent selection to target distinct molecules and prey through vastly different biological functions (for example, 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 in these regions (Fig. 6 and Extended Data Fig. 10). Indeed, the rapid decay of LD between venom gene paralogues suggests reduced selective interference among loci, allowing natural selection to operate more efficiently on individual loci and new allele combinations[57]. 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 polymorphisms[51]. Another consequence of

high recombination may be increased exposure of slightly deleterious alleles to selection[48], reducing the build-up of genetic load in gene-dense venom regions.

Methods

Reference genome and venom gene annotation.

We used the Prairie Rattlesnake genome assembly and annotation[12] as the reference for all analyses. Annotations of venom gene families include three of the main multigene families present in Prairie Rattlesnake venom, snake venom metalloproteinases (SVMPs; 11 genes), snake venom serine proteases (SVSPs; 11 genes) and phospholipases A2 (PLA2s; 4 genes), among other genes contributing to the venom phenotype (for example, CTL, CRISPs and LAAOs). The SVMP, SVSP and PLA2 venom gene families are each clustered in tandem arrays and occur on distinct microchromosomes (chromosomes 9, 10 and 15, respectively). For comparisons between genomic regions housing major venom gene families and the genomic background, we defined SVMP and SVSP regions as starting 50 kb upstream of the start site of the first gene and ending 50 kb downstream of the last gene in each family. The PLA2 venom gene region is much smaller than the SVMP and SVSP regions; thus, we only included 10 kb flanking regions to its coordinates. We also used SVMP, SVSP and PLA2 non-venom homologues identified in ref.[12] in our comparisons. Non-venom homologues are widely distributed across the genome, linked to chromosomes 1, 2, 4, 5, 6, 7, 9, 12, 15 and 18, as well as the Z chromosome.

Population sampling, whole-genome resequencing and variant calling.

We generated whole-genome resequencing data for individuals collected from northern populations of *C. viridis* (Montana) and *C. oreganus* (Idaho) to complement previous sampling of populations from refs.[62,86] (Fig. 1a and Supplementary Table 1). We also used resequencing data for an individual *C. atrox* as an outgroup for alignment-based analyses. Our total sampling included 17 California *C. oreganus* (*C. o. helleri* sensu ref.[64]), 17 Idaho *C. oreganus*, 19 Colorado *C. viridis*, 14 Montana *C. viridis* and 1 *C. atrox*. All procedures using animals and tissues were performed according to the University of Northern Colorado Institutional Animal Care and Use Committee protocols 0901C-SM-MLChick-12 and 1302D-SM-S-16.

We extracted DNA from blood tissue stored in DNA lysis buffer using phenol-chloroform-isoamyl extractions and then prepared sequencing libraries using Illumina Nextera Flex kits with sample-specific barcodes. Libraries were sequenced on Illumina NovaSeq 6000 lanes using 150 bp paired-end reads. Reads were quality filtered using Trimmomatic v.0.39 (ref.[87]) using the settings LEADING:20 TRAILING:20 MINLEN:32 AVGQUAL:30 and then mapped to the *C. viridis* reference genome using bwa mem[88] with default settings. A mean of 97% ± 1.7% reads aligned uniquely, corresponding to 28.7× ± 16.9× read depth per sample (Supplementary Table 1). We called genomic variants using the GATK v.4.0.8.1 best-practices workflow[89,90]. Specifically, we called individual variants using GATK HaplotypeCaller, specifying '--ERC GVCF' to generate a genomic variant call file (VCF; ref.[91]) per sample. We then called variant sites among the cohort of samples using GATK GenotypeGVCFs, specifying '--all-sites' to retain information for variant and

invariant genotypes. Following variant calling, we masked sites in the *C. viridis* repeat annotation[12] using GATK VariantFiltration and recoded repeats, indels, low depth and quality bases (depth < 5 and genotype quality < 30) and sites with mean read depths above the 97.5th quantile (depth > 36.24) as missing genotypes using bcftools filter[92]. The latter filtering step was applied to avoid spurious SNP calls based on paralogous mappings. Finally, we filtered to remove sites on scaffolds not assigned to chromosomes. The filtered variant dataset included 27,749,933 biallelic SNPs. We compared the distributions of genotype quality scores in venom gene regions to the genome-wide distribution to verify that there was not a bias in read depths due to potential paralogous mappings.

Phased variants were available for the Colorado *C. viridis* and California *C. oreganus* populations[62]. Briefly, non-singleton variants were phased using SHAPEIT v.2.904 (ref.[93]) after first identifying phase-informative reads using the extractPIRs extension. Using phase-informative reads and input VCFs for each population, we ran SHAPEIT using the settings -states 1000 -burn 200 -prune 210 -main 2000 and then confirmed the results by calculating low switch error rates across independent runs. To avoid spurious results from paralogous mappings in multigene venom gene regions, we filtered phased variants with mean read depths greater than the 97.5th quantile per chromosome. The final phased variant datasets included 9,737,794 SNPs for *C. viridis* and 6,365,456 SNPs for *C. oreganus*.

Removal of bioinformatic artifacts of copy-number variation.

Multigene venom gene families are composed of closely related paralogues and copy number can vary within and between species[14,94]. To avoid the influence of CNVs on population genetic inferences in venom gene regions, we performed pairwise CNV detection analyses between all individuals in all populations and a randomly selected individual from the CV1 population (the same population as the reference genome animal) with the expectation of little structural variation in this population compared to the reference genome (Extended Data Fig. 1a). Indeed, relative read depths across the venom gene regions in the CV1 population, measured as the $\log_2$ of the ratio of read depth divided by the autosomal median read depth, indicate that structural variation is minimal or absent in this population (Extended Data Fig. 1b). We performed pairwise CNV detection for each individual using CNV-seq[95]. This approach detects CNVs based on pairwise differences in read depths when mapped to a common reference genome. We extracted alignments on chromosomes 9, 10 and 15 (microchromosomes housing major venom gene families) for CNV analysis using samtools v.1.10 (ref.[92]). We ran CNV-seq on mapping hits, specifying --log2 0.6, --p 0.001, --bigger-window 1 and --minimum-windows 2 and specifying a window size of 2.5 kb. We then used the coordinates of significant CNVs to mask genotypes per individual such that masked genotypes were not used in downstream population genetic inference. This approach has the advantage of accounting for polymorphism in CNVs within and between populations and species, and we further quantified the number of individuals per population with masked genotypes in detected CNVs (% CNV) to visualize the degree of genotype masking across the venom gene regions (Extended Data Fig. 1a,b).

We used a one-tailed test of Hardy–Weinberg equilibrium (HWE) adapted from ref.[96] to remove potential bioinformatic artifacts that persisted after CNV masking. We used a

one-tailed test across our dataset to specifically remove sites with significant departures from HWE due to excess heterozygosity unlikely to be explained by natural variation. We performed a Benjamini–Hochberg false discovery rate correction with a threshold of 0.001 on HWE *P* values to account for multiple testing and to avoid false positives. This procedure found 12,612 excess heterozygosity SNPs across the genome with significant departures from HWE, which we removed from further analysis. To validate that our CNV and HWE filtering approaches removed potential bioinformatic artifacts of structural variation in venom genes, we visualized minor allele frequency spectra for each gene (Extended Data Fig. 1c).

Population structure.

We inferred population structure in *C. viridis* and *C. oreganus* using the likelihood-model approach in ADMIXTURE[97] for *K* genetic clusters, with *K* ranging from 1 to 16. Before analysis, we used VCFtools[91] to prune our SNP dataset to retain biallelic ingroup SNPs with minor allele frequencies ≥0.05 (--maf 0.05) that were genotyped in at least 60% of samples (--max-missing 0.4). We also thinned the dataset to only keep SNPs separated by at least 1 kb to reduce the effects of tight physical linkage on estimates of assignment probabilities to one or more *K* clusters. These filtering settings yielded 348,264 SNPs for analysis. We converted data in VCF format for ADMIXTURE using Plink v.1.90 (ref.[98]) and then ran iterations of ADMIXTURE for models with $K = 1$ to 16 clusters. We evaluated the fit of *K* models to the data using the cross-validation procedure.

Demographic history.

We used PSMC v.0.6.5 (ref.[66]) to estimate effective population size through time for each of the four populations. We sampled read mapping data for a representative individual from each population at random and then called heterozygous sites using bcftools mpilup and call functions[92], using bcftools filter and view functions to remove indels, SNPs with low (DP < 5) and high (DP > 50) read depths and SNPs overlapping the *C. viridis* repeat annotation. We ran PSMC specifying a generation time of 3 yr and the generalized squamate mutation rate of 2.4×10^{-9} reported in ref.[99]. We specified a time segment pattern of $4 + 30 \times 2 + 4 + 6 + 10$ after performing preliminary analyses using a range of patterns and tested the robustness of effective population size estimates using 100 bootstrap replicates per analysis.

Measurements of genetic diversity, differentiation and selection.

We measured various population genetic summary statistics across the genome to understand the roles of evolutionary processes in shaping the broad genomic landscape of diversity. We further investigated patterns in venom gene regions, specifically, to detect signatures of selection on venom loci and evaluate the effects of selection on linked variation. We first estimated within-population nucleotide diversity (π), between-population sequenced divergence (d_{xy}) and between-population relative differentiation (Fst) using Pixy v.0.95 (ref.[100]). Pixy explicitly accounts for missing genotypes when estimating π and d_{xy}, which are both biased by the presence of missing data among sites and associated sampling variance[101]. Pixy therefore provides more robust estimates of within- and between-population diversity than do methods that do not account for missing data[100]. We estimated π for each *C. viridis* and *C. oreganus* population and also for outgroup species and

calculated d_{xy} and Fst between pairs of *C. viridis* and *C. oreganus* populations, specifying minimum depth filters for the inclusion of sites (--variant_filter_expression 'DP ≥ 5' and --invariant_filter_expression 'DP ≥ 5')'. We performed multiple runs to calculate statistics in 100, 10 and 1 kb sliding windows. Because the PLA2 region is relatively small, we also calculated statistics in 250 bp sliding windows across chromosome 15 to evaluate patterns across this region.

Additional statistics were calculated in sliding windows to evaluate evidence of selection in venom gene regions suggested by patterns of genetic diversity and differentiation. We tested for deviations from neutral expectations using Tajima's *D* statistic[102], which we measured using VCFtools[91]. We measured the frequency of fixed differences, d_f, between populations. We further measured extended haplotype homozygosity on phased variants in the CV1 and CO1 populations using the |iHS| statistic[67] in the R package rehh[103,104]. SNPs with minor allele frequencies <0.05 were filtered before calculating |iHS|, and we specified 'polarized = false' because phased data for outgroups were not available to polarize ancestral from derived alleles. We tested for evidence of balancing selection using BetaScan[71,72], which calculates the allele frequency correlation summary statistic, *ß*, in sliding windows. The *ß* statistic is capable of detecting clusters of intermediate-frequency polymorphisms surrounding a central balanced variant. These clusters reflect regional distortions in the time to most recent common ancestor, an expected result of balancing selection[42,43]. We used glactools[105] to convert phased variants called in CV1 and CO1 populations to folded site frequency spectra. We then ran BetaScan on each chromosome per population with a minimum folded core SNP allele frequency -m 0.15 and under a series of -w (500, 1,000 and 2,000 bp) and -p (2, 5, 10, 20) parameters to determine if the results were sensitive to differences in these settings (Supplementary Table 7). As for Pixy analyses, we calculated summary statistics in 100, 10 and 1 kb non-overlapping sliding windows across the genome and in 250 bp windows on chromosome 15. We used the composite-likelihood ratio test implemented in BalLerMix+ ($B_{0,MAF}$; refs.[73,74]) to explicitly test for evidence of balancing selection versus neutral evolution across the genome. We calculated the genome-wide 95% significance threshold for *C. viridis* and *C. oreganus* to distinguish genomic regions in each species with outstanding $B_{0,MAF}$ scores.

We calculated the relative proportion of fixed differences, private polymorphisms and trans-species polymorphisms among variant sites observed in the phased CV1 and CO1 datasets by first generating chromosome-length haplotype alignments using BCFtools[92], with *C. atrox* as an outgroup. We then identified variant sites for which at least 50% of samples in each species had called genotypes and classified private polymorphisms as sites where one species was invariant while the other was at a minor allele frequency in the range between 0.1 and 0.5. Similarly, we conservatively classified trans-species polymorphisms as sites where both species had minor allele frequencies between 0.1 and 0.5. To evaluate functional variation in venom genes themselves, we extracted and translated the exon alignments to quantify the number of amino acid replacements produced by private and trans-species polymorphisms in both species.

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 3.7.1 (ref.[106]), we generated $L = 10$ kb long sequences under demographic histories inferred by PSMC[66] 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 $\mu = 8.4 \times 10^{-9}$(ref.[99]) 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*, respectively[62]. Details of classification model training, simulations and comparisons with empirical data to determine probabilities of neutrality, negative frequency-dependent selection and heterozygote advantage in venom gene regions can be found in the Supplementary Information.

Analysis of recombination rate and linkage disequilibrium.

Population-scaled recombination rates ($\rho = 4N_e r$) for *C. viridis* and *C. oreganus* were estimated previously[62]. Briefly, ρ was estimated from phased variants using the rjMCMC procedure in LDhelmet[107] for chromosome-assigned scaffolds in the *C. viridis* genome using a block penalty ('bpen') of 10. Recombination hotspots in *C. viridis* and *C. oreganus* were defined as intervals where estimated ρ was greater than tenfold higher than immediate upstream and downstream 40 kb regions and hotspot 'heat' was calculated by dividing ρ within hotspot intervals by the mean rate in flanking regions. Hotspots within 5 kb of one another were filtered to retain the hotspot with the highest relative heat. We investigated ρ variation in the SVMP, SVSP and PLA2 venom gene regions in 1, 10 and 100 kb sliding windows and extracted all intervals in recombination maps within coordinates for the venom gene families to quantify the distribution of ρ in venom gene regions. We also quantified background ρ distributions outside of these regions for each of the microchromosomes housing venom genes and for non-venom homologues distributed across macrochromosomes and microchromosomes. We compared distributions after dividing ρ estimates by π per genomic window to account for effective population size (Results).

We quantified LD in major venom gene regions by calculating r^2 between all pairs of phased SNPs using VCFtools --hap-r2 (ref.[91]), after filtering to retain SNPs with minor allele frequencies above 0.1 (--maf 0.1). We examined LD decay by calculating the mean and interquartile range of r^2 values as a function of physical distance between all pairs of SNPs. We then repeated this process for upstream and downstream flanking regions of equal size to venom gene regions; mean and interquartile ranges were calculated after combining pairwise r^2 for all SNPs in both flanking regions.

Statistical analysis.

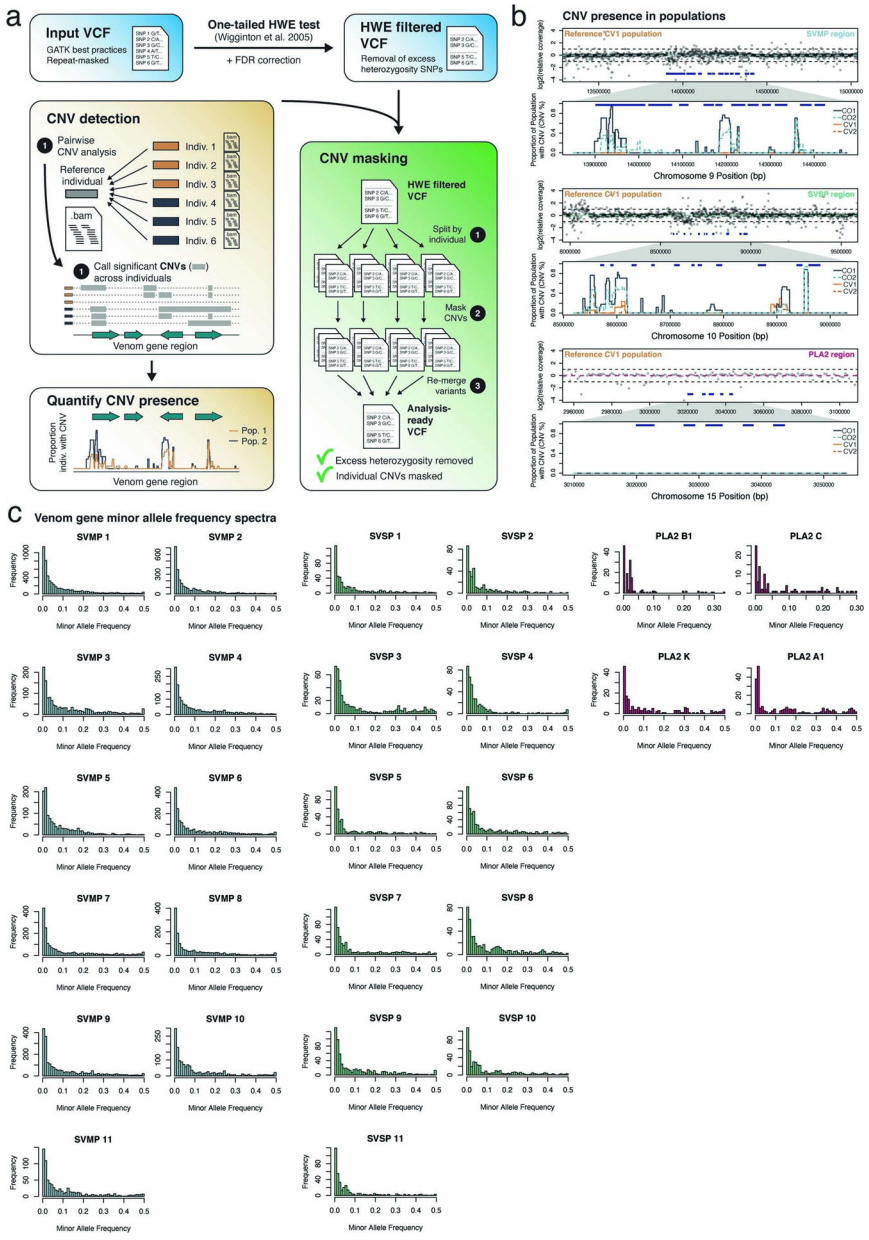
We performed all statistical analyses in R (ref.[108]). We used Mann–Whitney *U* tests and Welch's two-sample *t*-tests to compare distributions of population genetic estimators in venom gene regions to respective chromosomal backgrounds and non-venom homologues.

We calculated Spearman's rank order correlation coefficients to examine associations between parameters across the genome and in venom gene regions, specifically.

Reporting summary.

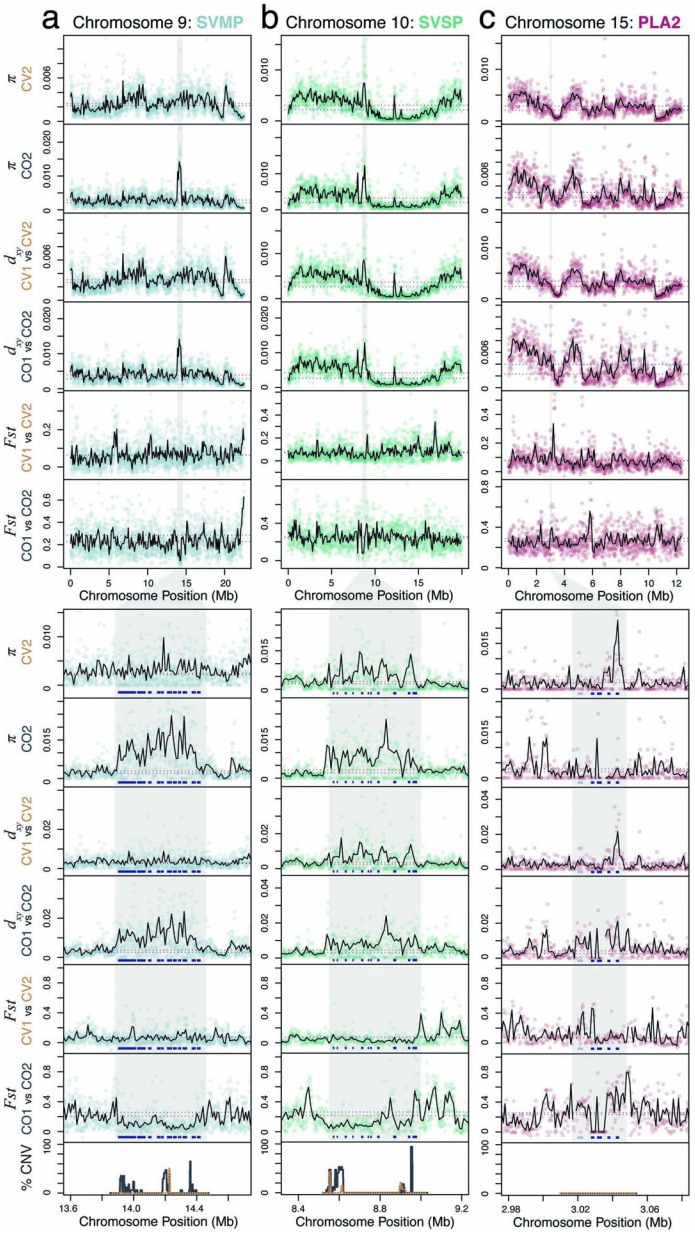
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Extended Data



Extended Data Fig. 1 |. Overview of filtering strategy to remove potential bioinformatic artifacts of copy-number variation (CNv) in major venom gene regions.

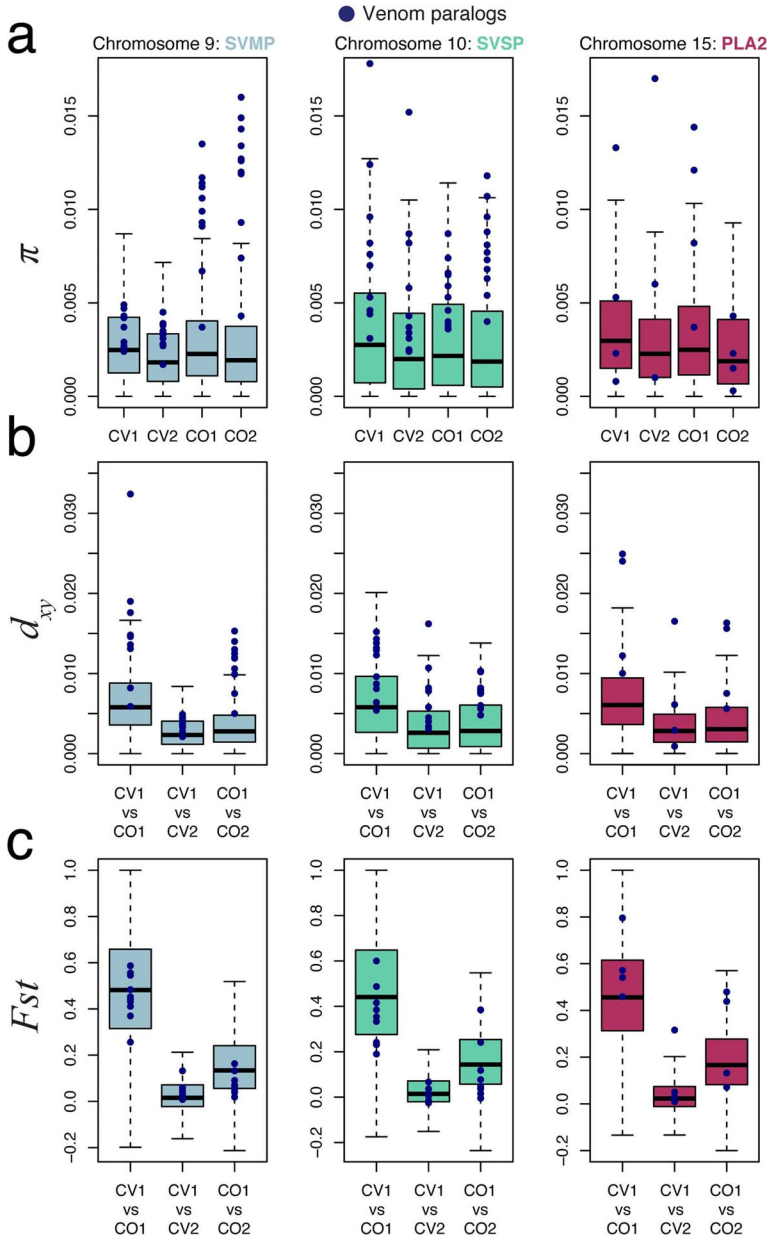
a Schematic representation of the workflow used to filter departures from Hardy–Weinberg equilibrium (HWE) due to excess heterozygosity at SNP positions. Input variant calls were analysed using a one-tailed HWE test (adapted from Wigginton et al. 2005) and SNPs with significant HWE *P*-values after FDR correction were filtered. Individual CNVs were detected, scored, and regional variation in CNV presence, quantified as the proportion of individuals in a population with a detected CNV (% CNV), was measured across each major venom gene region. CNV calls were used to mask genotypes per individual prior to population genetic analysis. **b** Regional variation in CNV presence per population across the major venom gene regions. Top panels for each venom gene region show log2 read depths in sliding windows relative to the autosomal median depth for the CV1 reference population. Here values of zero indicate equal coverage to the autosomal median, values of −1 equal half coverage, and values of 1 equal twice the autosomal median coverage. Dark blue segments indicate the locations of venom genes in each region. Lower panels for each region show variation in the proportion of individuals with masked genotypes in a detected CNV (% CNV) per population. **c** Minor allele frequency spectra across the dataset for each gene in the major venom gene families after using the described HWE and CNV filtering strategy.



Extended Data Fig. 2 |. Genome scans of genetic diversity within populations and differentiation between populations across chromosome housing major venom gene families (SvMPs, SvSPs, and PLA2s) in Cv2 and Co2 populations.

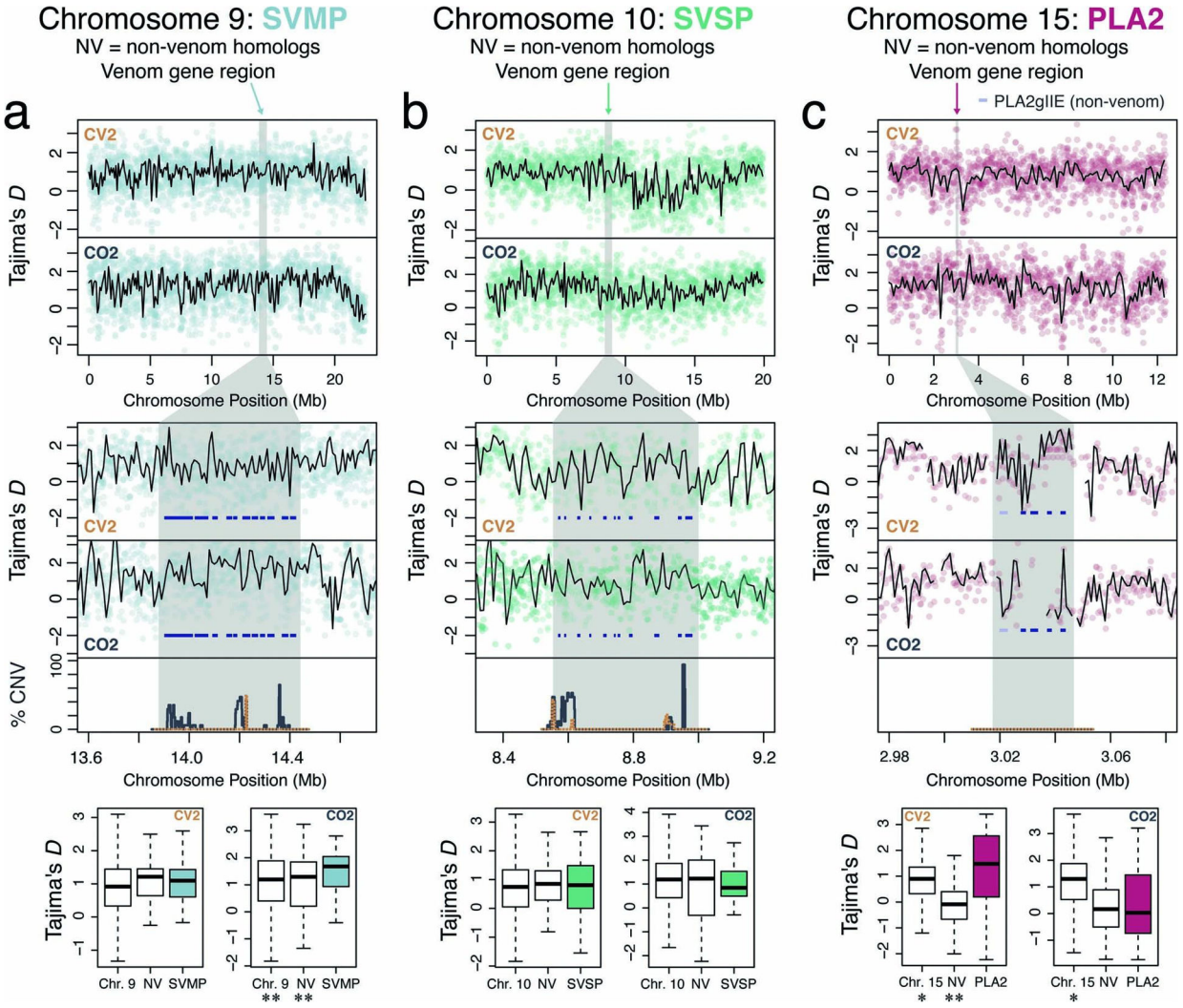
a Sliding windows of nucleotide diversity (π) in *C. viridis* (CV2) and *C. oreganus* (CO2) populations and sequence divergence (d_{XY}) and relative differentiation (*Fst*) between CV1 and CV2 and between CO1 and CO2 across Chromosome 9 (top panels), and in the SVMP region (bottom panels). **b** π, d_{XY}, and *Fst* across Chromosome 10 (top), and in the SVSP region (bottom). **c** π, d_{XY}, and *Fst* across Chromosome 15 (top), and in the PLA2 region (bottom). Shaded points in top panels show estimates in 10 kb windows and lines show estimates in 100 kb windows. In bottom panels in **a** and **b**, shaded points show estimates in 1 kb windows, and lines show estimates in 10 kb windows. In bottom panels in **c**, shaded points show estimates in 250 bp windows and lines are estimates in 1 kb

windows. The regions housing venom genes are shaded in grey in all panels. Chromosome-specific and genome-wide mean values for each statistic are represented by blue and red dashed horizontal lines. The locations of individual venom genes are shown as blue boxes (bottom panels). The non-venom homologue PLA2gIIE is shown in light purple. Gaps in measurements are locations that were masked due to significant evidence of copy-number variation between *C. viridis* and *C. oreganus*.



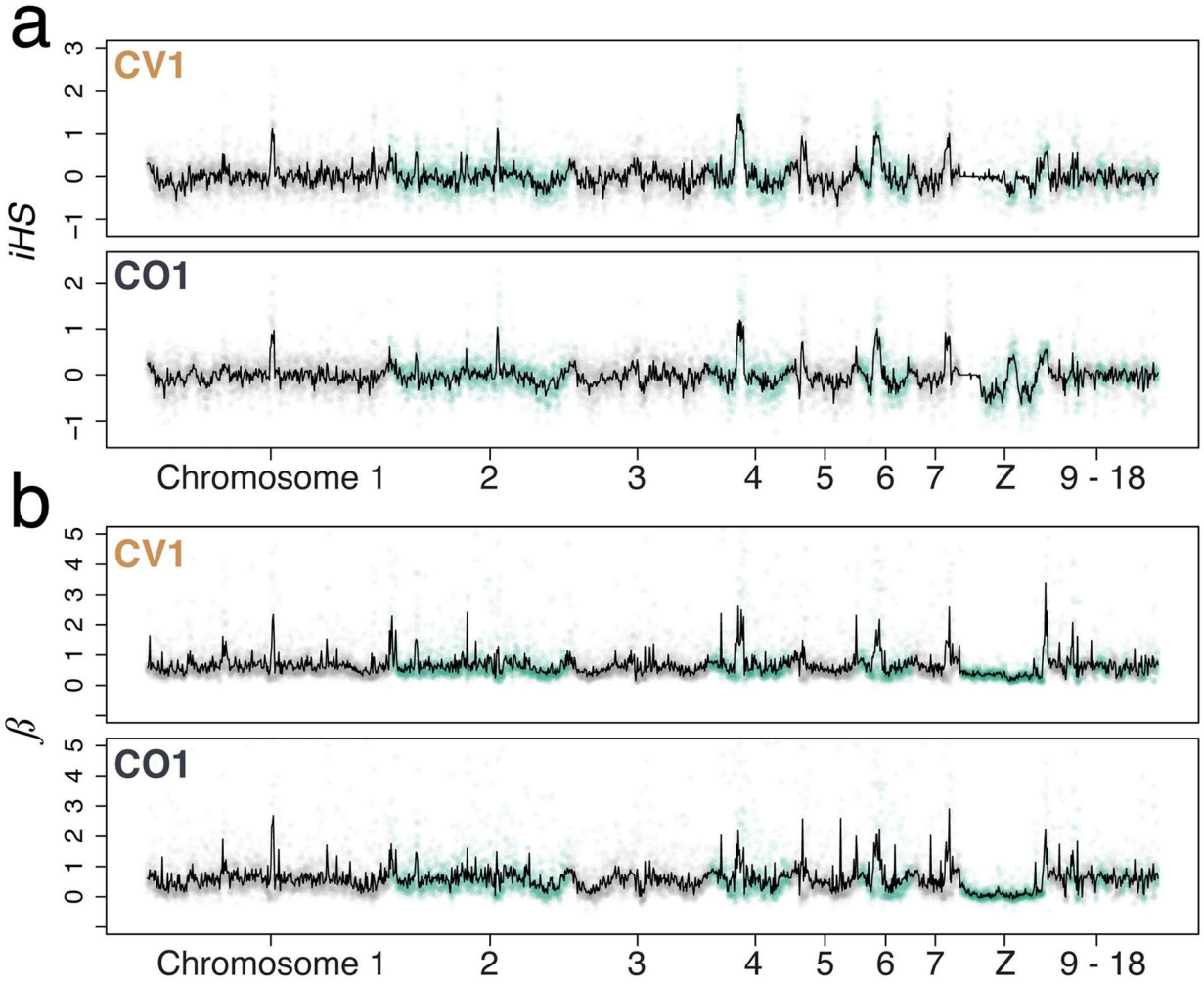
Extended Data Fig. 3 |. Genetic diversity point estimates for major venom paralogues.
Point estimates (dark blue circles) of π (**a**), d_{xy} (**b**), and *Fst* (**c**) for paralogues in the three major venom gene families, compared to chromosome-specific background distributions outside of venom gene regions. Boxplots show the median (horizontal lines), interquartile

(box limits), and range (whiskers) based on n = 2,253, 1,998, and 1,239 10 kb sliding windows for chromosomes 9, 10, and 15, respectively.



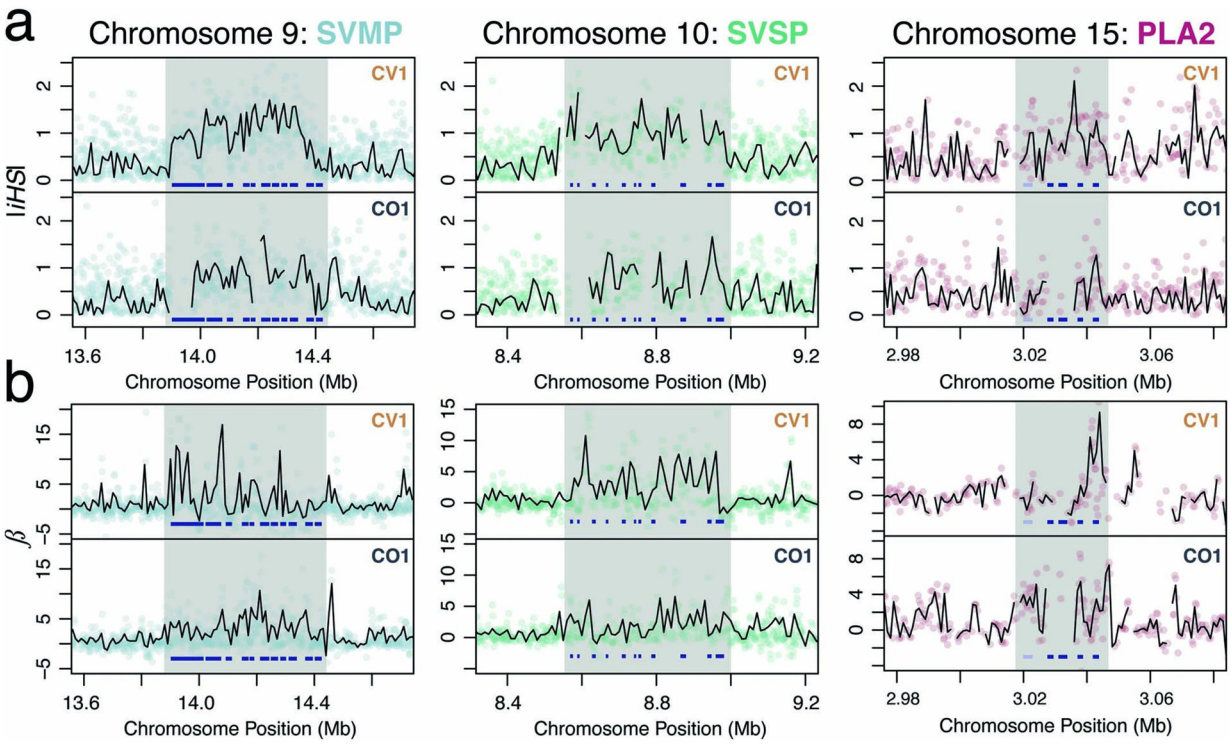
Extended Data Fig. 4 |. Tajima's *D* across major venom regions in Cv2 and Co2 populations. Tajima's *D* across SVMP (**a**), SVSP (**b**), and PLA2 (**c**) regions in CV2 and CO2 populations. Top panels show sliding window Tajima's *D* estimates in 100 kb (lines) and 10 kb (points points) windows. Venom gene regions are shaded in grey. Middle panels in **a** and **b** show zoomed in venom gene regions with sliding window estimates in 10 kb (lines) and 1 kb (points) windows. Middle panels in **c** show estimates in 1 kb (lines) and 250 bp (points) windows. Gaps in lines represent windows where there was insufficient data to calculate a mean estimate. Dark blue segments show the locations of venom genes in each region, and the non-venom homologue PLA2gIIE is shown in light purple in **c**. Regional variation in the presence of CNVs (% CNV) in CV2 (orange dashed line) and CO2 (dark blue line) is shown below venom region scans. Individual genotypes in detected CNVs were masked. Bottom panels show distributions of chromosome-specific and non-venom homologue (NV) backgrounds compared to values in each venom gene region, with boxplots showing the

median (horizontal lines), interquartile (box limits), and range (whiskers). Asterisks indicate significant differences between venom gene regions and chromosome backgrounds and non-venom homologues based on two-tailed Welch's two-sample *t*-tests and $n = 2{,}293$, 2,068, and 1,272 10 kb sliding windows for SVMP, SVSP, and PLA2 comparisons, respectively (*$P < 0.05$; **$P < 0.001$). Exact *P*-values for comparisons can be found in Supplementary Table 5.

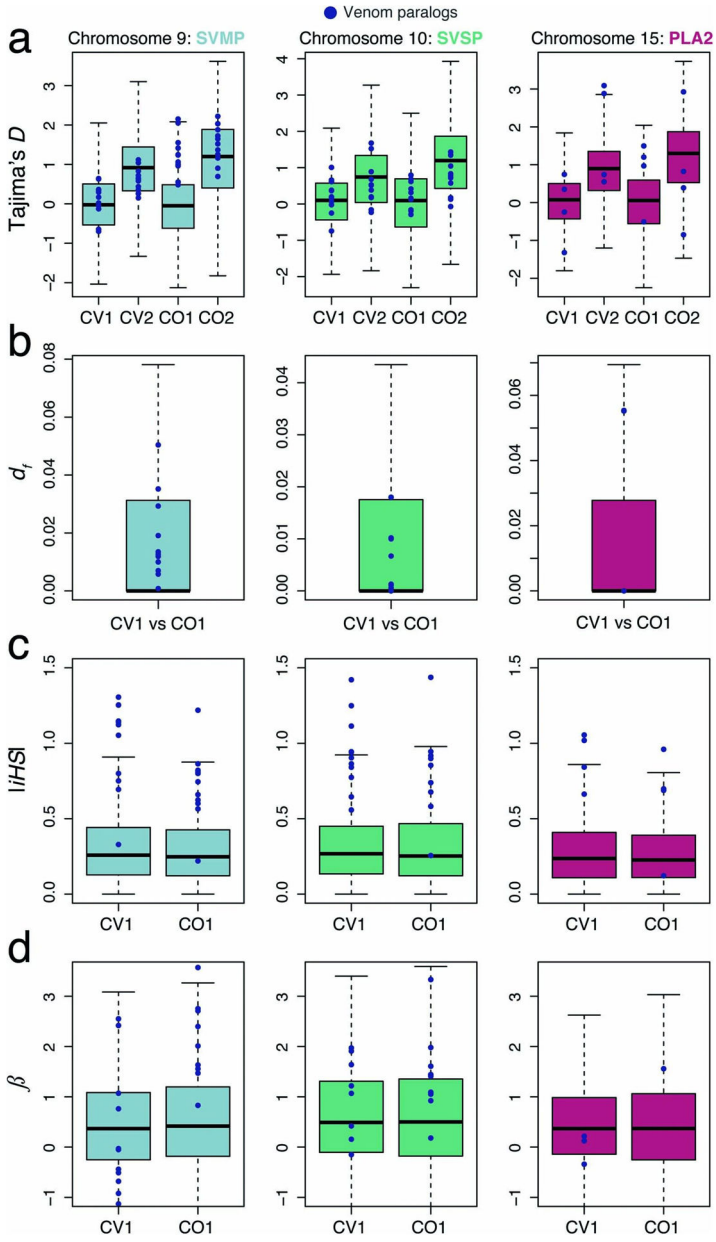


Extended Data Fig. 5 |. Genomic scans of *iHS* and *ß* selection statistics in Cv1 and Co1 populations.

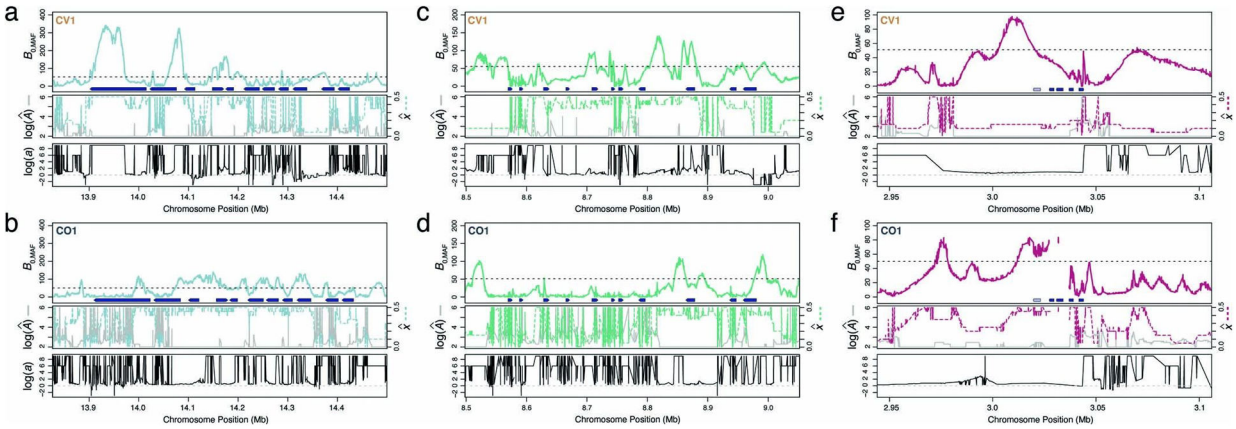
Lines show mean estimates in 1 Mb sliding windows. Shaded points show mean estimates in 100 kb sliding windows.



Extended Data Fig. 6 |. Scans of *iHS* and ß selection statistics across major venom regions.
Scans of **a**|*iHS*|and **b** *ß* selection statistics in CV1 and CO1 populations across SVMP, SVSP, and PLA2 venom gene regions. Lines in SVMP and SVSP panels show mean estimates in 10 kb sliding windows. Shaded points in SVMP and SVSP panels show mean estimates in 1 kb sliding windows. Lines in PLA2 panels show mean estimates in 1 kb sliding windows. Shaded points in PLA2 panels show mean estimates in 250 bp sliding windows. Gaps in lines represent windows where there was insufficient data to calculate a mean estimate. Venom gene locations are shown with dark blue segments. The non-venom PLA2gIIE homologue is shown as a light purple segment.

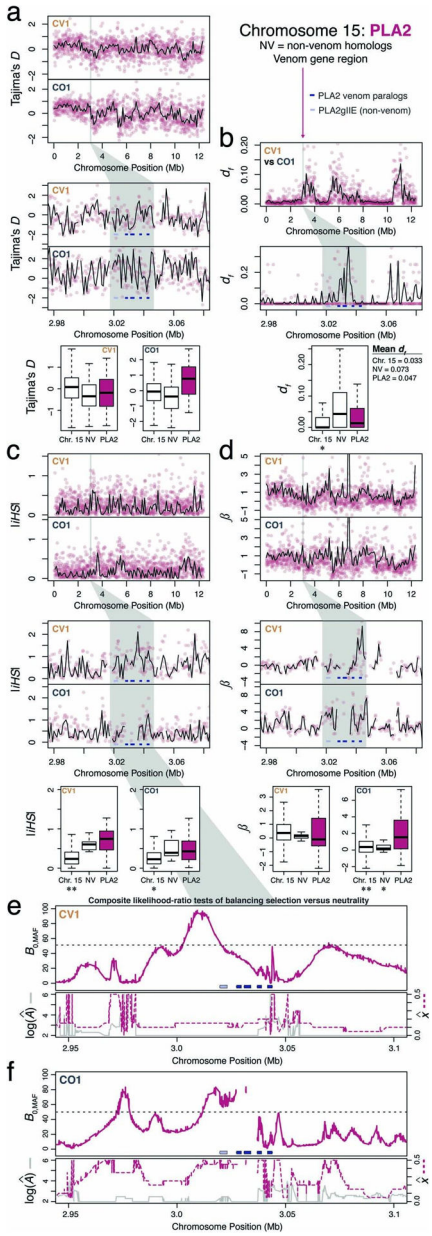


Extended Data Fig. 7 |. Selection statistic point estimates for major venom paralogues.
Point estimates (dark blue circles) of Tajima's *D* (**a**), d_f (**b**), |*iHS*|(**c**), and ß (**d**) for paralogues in the three major venom gene families, compared to chromosome-specific background distributions outside of venom gene regions. Boxplots show the median (horizontal lines), interquartile (box limits), and range (whiskers) based on n = 2,253, 1,998, and 1,239 10 kb sliding windows for chromosomes 9, 10, and 15, respectively.



Extended Data Fig. 8 |. Diagnostic parameters in composite-likelihood ratio tests of selection in major venom regions.

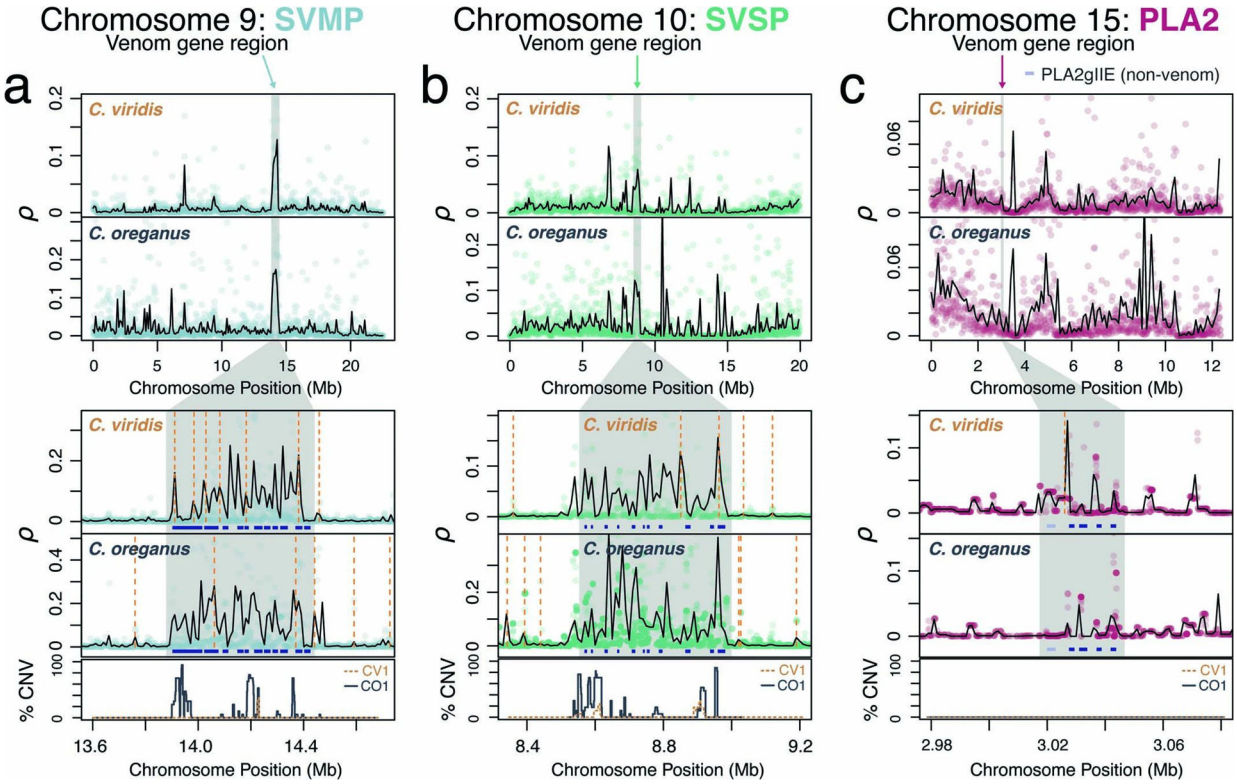
Results of composite-likelihood ratio tests of balancing selection and diagnostic parameters in the SVMP (**a-b**), SVSP (**c-d**), and PLA2 (**e-f**) venom gene regions. Top and middle panels show $B_{0,\mathrm{MAF}}$ scores, $\log(\widehat{A})$, and $\hat{x}$ parameters as shown in Fig. 3 and Extended Data Fig. 9. Dark blue arrows show locations of venom genes and dashed lines show the genome-wide 95$^{\mathrm{th}}$ quantile. Lower panels show the dispersion parameter, $\log(\widehat{a})$, where positive values indicate balancing selection in regions with high $B_{0,\mathrm{MAF}}$ scores and negative values are indicative of positive selection. The dashed horizontal line indicates 0 on the y-axis.



Extended Data Fig. 9 |. Signatures of selection in the PLA2 venom gene region.

Signatures of selection in the PLA2 venom gene region, with comparisons to chromosomal backgrounds and non-venom homologues (NV). **a** Tajima's D across chromosome 15 in *C. viridis* (CV1) and *C. oreganus* (CO1) populations (top). Middle panels show variation zoomed into the PLA2 region, shaded in grey. Boxplots show distributions of Tajima's *D* for chromosome 15, non-venom homologues of the PLA2 family, and PLA2s. **b** Proportion of fixed differences (d_f) between CV1 and CO1. **c** Integrated haplotype statistics (|*iHS*|) in CV1 and CO1. **d** ß statistic measuring allele frequency correlation in CV1 and CO1. Points in chromosome scan panels represent mean estimates in 10 kb sliding windows, and lines represent 100 kb windowed estimates. Points in zoomed PLA2 region scans represent mean estimates in 250 bp windows and lines show 1 kb windowed estimates.

Locations of individual PLA2 genes are shown as dark blue segments (bottom panels). The PLA2gIIE non-venom homologue is shown in light purple. Gaps in lines represent windows where there was insufficient data to calculate a mean estimate. Boxplots in **a-d** show the median (horizontal lines), interquartile (box limits), and range (whiskers). Asterisks indicate significant differences between venom gene regions and chromosome backgrounds and non-venom homologues based on two-tailed Welch's two-sample *t*-tests (Tajima's D and $|iHS|$) and Mann–Whitney U tests (d_f and ß) and $n = 1{,}272$ 10 kb sliding windows for comparisons ($*P < 0.05$; $**P < 0.001$; $***P < 2.2 \times 10^{-16}$). Exact *P*-values for comparisons can be found in Supplementary Table 5. **e-f** Results of composite-likelihood ratio tests of balancing selection versus neutrality using $B_{0,\mathrm{MAF}}$ scores in the PLA2 region in CV1 and CO1. Upper panels show $B_{0,\mathrm{MAF}}$ scores, with higher values indicating greater evidence for balancing selection. Arrows show locations of PLA2 genes. Lower panels show inferred footprint size, $\log(\widehat{A})$, as solid grey lines and equilibrium allele frequency, $\hat{x}$, as dashed lines. Dashed lines in top panels of e-f show the genome-wide 95$^{\mathrm{th}}$ quantile.



Extended Data Fig. 10 |. Population-scaled recombination rates in major venom regions. Population-scaled recombination rate ($\rho = 4N_e r$) across SVMP (**a**), SVSP (**b**), and PLA2 (**c**) venom gene regions in *C. viridis* and *C. oreganus*. Upper panels show chromosome-wide variation and lower panels show variation within the venom regions, specifically, highlighted by the grey shading in all panels. Dark blue segments in lower panels show the locations of venom paralogues. The light purple segment in **c** is the non-venom homologue PLA2gIIE. Shaded points in the upper panels of **a-c** represent mean ρ in 10 kb windows and black lines represent 100 kb windowed means. In lower panels of **a** and **b**, points and lines represent 1

kb and 10 kb windowed ρ. In lower panels of **c**, lines represent 1 kb windowed ρ and points are estimates from all SNP intervals. Vertical dashed lines show the locations of inferred recombination hotspots from Schield et al. (2020). Panels at the bottom show regional variation in the proportion of *C. viridis* (orange dashed line) and *C. oreganus* (dark blue line) individuals per population with evidence of copy-number variation (% CNV) across the venom gene regions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank R. Orton and N. Balchan for assistance in the field. We thank J. Vindum and the California Academy of Sciences for tissue loans. A. Ludington kindly provided advice on demographic analysis. We thank S. Flaxman and R. Safran for helpful discussion on the study and feedback on the manuscript. This work was supported by National Science Foundation (NSF) postdoctoral research fellowship grant DBI-1906188 to D.R.S., NSF grant DEB-1501886 to D.R.S. and T.A.C. and NSF grant DEB-1655571 to T.A.C., S.P.M. and J.M.M., NSF grants DEB-1949268, BCS-2001063 and DBI-2130666 to M.D. and National Institutes of Health grant R35GM128590 to M.D.

Data availability

The genomic data that support the findings of this study are available at NCBI SRA under accession PRJNA593834.

References

1. Zancolli G & Casewell NR Venom systems as models for studying the origin and regulation of evolutionary novelties. Mol. Biol. Evol 37, 2777–2790 (2020). [PubMed: 32462210]
2. Arbuckle K From molecules to macroevolution: venom as a model system for evolutionary biology across levels of life. Toxicon X 6, 100034 (2020). [PubMed: 32550589]
3. Mackessy SP Handbook of Venoms and Toxins of Reptiles (CRC Press, 2021).
4. Hargreaves AD, Swain MT, Hegarty MJ, Logan DW & Mulley JF Restriction and recruitment—gene duplication and the origin and evolution of snake venom toxins. Genome Biol. Evol 6, 2088–2095 (2014). [PubMed: 25079342]
5. Casewell NR, Huttley GA & Wuster W Dynamic evolution of venom proteins in squamate reptiles. Nat. Commun 3, 1066 (2012). [PubMed: 22990862]
6. Casewell NR, Wuster W, Vonk FJ, Harrison RA & Fry BG Complex cocktails: the evolutionary novelty of venoms. Trends Ecol. Evol 28, 219–229 (2013). [PubMed: 23219381]
7. Fry BG & Wuster W Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. Mol. Biol. Evol 21, 870–883 (2004). [PubMed: 15014162]
8. 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 32, 173–183 (2015). [PubMed: 25338510]
9. Mackessy SP in The Biology of Rattlesnakes (eds Hayes WK et al.) 495–510 (Loma Linda Univ. Press, 2008).
10. Ikeda N et al. Unique structural characteristics and evolution of a cluster of venom phospholipase A 2 isozyme genes of *Protobothrops flavoviridis* snake. Gene 461, 15–25 (2010). [PubMed: 20406671]
11. Dowell NL et al. The deep origin and recent loss of venom toxin genes in rattlesnakes. Curr. Biol 26, 2434–2445 (2016). [PubMed: 27641771]

12. Schield DR et al. The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. Genome Res. 29, 590–601 (2019). [PubMed: 30898880]
13. Lynch VJ Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. BMC Evol. Biol 7, 2 (2007). [PubMed: 17233905]
14. Casewell NR, Wagstaff SC, Harrison RA, 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). [PubMed: 21478373]
15. Mayr E Cause and effect in biology. Science 134, 1501–1506 (1961). [PubMed: 14471768]
16. Aird SD et al. Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. Genome Biol. Evol 9, 2640–2649 (2017). [PubMed: 29048530]
17. Margres MJ et al. Tipping the scales: the migration–selection balance leans toward selection in snake venoms. Mol. Biol. Evol 36, 271–282 (2019). [PubMed: 30395254]
18. Rautsaw RM 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).
19. Holding ML, Biardi JE & Gibbs HL Coevolution of venom function and venom resistance in a rattlesnake predator and its squirrel prey. Proc. R. Soc. B 283, 20152841 (2016).
20. Davies E-L & Arbuckle K Coevolution of snake venom toxic activities and diet: evidence that ecological generalism favours toxicological diversity. Toxins 11, 711 (2019). [PubMed: 31817769]
21. Smiley-Walters SA, Farrell TM & Gibbs HL Evaluating local adaptation of a complex phenotype: reciprocal tests of pigmy rattlesnake venoms on treefrog prey. Oecologia 184, 739–748 (2017). [PubMed: 28516321]
22. Holding ML, Drabeck DH, Jansa SA & Gibbs HL Venom resistance as a model for understanding the molecular basis of complex coevolutionary adaptations. Integr. Comp. Biol 56, 1032–1043 (2016). [PubMed: 27444525]
23. Poran NS, Coss RG & Benjamini ELI Resistance of California ground squirrels (*Spermophilus beecheyi*) to the venom of the northern Pacific rattlesnake (*Crotalus viridis oreganus*): a study of adaptive variation. Toxicon 25, 767–777 (1987). [PubMed: 3672545]
24. Heatwole H & Poran NS Resistances of sympatric and allopatric eels to sea snake venoms. Copeia 1995, 136–147 (1995).
25. Pomento AM, Perry BW, Denton RD, Gibbs HL & Holding ML No safety in the trees: local and species-level adaptation of an arboreal squirrel to the venom of sympatric rattlesnakes. Toxicon 118, 149–155 (2016). [PubMed: 27158112]
26. Biardi JE, Chien DC & Coss RG California ground squirrel (*Spermophilus beecheyi*) defenses against rattlesnake venom digestive and hemostatic toxins. J. Chem. Ecol 32, 137–154 (2006). [PubMed: 16525875]
27. Jansa SA & Voss RS Adaptive evolution of the venom-targeted vWF protein in opossums that eat pitvipers. PLoS ONE 6, e20997 (2011). [PubMed: 21731638]
28. Voss RS & Jansa SA Snake-venom resistance as a mammalian trophic adaptation: lessons from didelphid marsupials. Biol. Rev 87, 822–837 (2012). [PubMed: 22404916]
29. Drabeck DH, Dean AM & Jansa SA Why the honey badger don't care: convergent evolution of venom-targeted nicotinic acetylcholine receptors in mammals that survive venomous snake bites. Toxicon 99, 68–72 (2015). [PubMed: 25796346]
30. Gibbs HL et al. The molecular basis of venom resistance in a rattlesnake–squirrel predator–prey system. Mol. Ecol 29, 2871–2888 (2020). [PubMed: 32593182]
31. Kordiš D, Bdolah A & Gubenšek F Positive Darwinian selection in *Vipera palaestinae* phospholipase A2 genes is unexpectedly limited to the third exon. Biochem. Biophys. Res. Commun 251, 613–619 (1998). [PubMed: 9792822]
32. Kordiš D & Gubenšek F Adaptive evolution of animal toxin multigene families. Gene 261, 43–52 (2000). [PubMed: 11164036]
33. Juárez P, Comas I, González-Candelas F & Calvete JJ Evolution of snake venom disintegrins by positive Darwinian selection. Mol. Biol. Evol 25, 2391–2407 (2008). [PubMed: 18701431]

34. McDonald JH & Kreitman M Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351, 652–654 (1991). [PubMed: 1904993]
35. Strickland JL et al. Evidence for divergent patterns of local selection driving venom variation in Mojave Rattlesnakes (*Crotalus scutulatus*). Sci. Rep 8, 17622 (2018). [PubMed: 30514908]
36. Przeworski M, Coop G & Wall JD The signature of positive selection on standing genetic variation. Evolution 59, 2312–2323 (2005). [PubMed: 16396172]
37. Cutter AD & Payseur BA Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet 14, 262 (2013). [PubMed: 23478346]
38. Aquadro CF, Begun DJ & Kindahl EC in Non-Neutral Evolution (ed. Golding B) 46–56 (Springer, 1994).
39. Begun DJ & Aquadro CF Molecular population genetics of the distal portion of the X chromosome in *Drosophila*: evidence for genetic hitchhiking of the yellow-achaete region. Genetics 129, 1147–1158 (1991). [PubMed: 1664405]
40. Maynard Smith J & Haigh J The hitch-hiking effect of a favourable gene. Genet. Res 23, 23–35 (1974). [PubMed: 4407212]
41. Barton NH Genetic hitchhiking. Philos. Trans. R. Soc. Lond. B 355, 1553–1562 (2000). [PubMed: 11127900]
42. Charlesworth D Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet. 2, e64 (2006). [PubMed: 16683038]
43. Fijarczyk A & Babik W Detecting balancing selection in genomes: limits and prospects. Mol. Ecol 24, 3529–3545 (2015). [PubMed: 25943689]
44. Piertney SB & Oliver MK The evolutionary ecology of the major histocompatibility complex. Heredity 96, 7–21 (2006). [PubMed: 16094301]
45. Kelley J, Walter L & Trowsdale J Comparative genomics of major histocompatibility complexes. Immunogenetics 56, 683–695 (2005). [PubMed: 15605248]
46. Bakker EG, Toomajian C, Kreitman M & Bergelson J A genome-wide survey of R gene polymorphisms in *Arabidopsis*. Plant Cell 18, 1803–1818 (2006). [PubMed: 16798885]
47. Goldberg EE et al. Species selection maintains self-incompatibility. Science 330, 493–495 (2010). [PubMed: 20966249]
48. Llaurens V, Whibley A & Joron M Genetic architecture and balancing selection: the life and death of differentiated variants. Mol. Ecol 26, 2430–2448 (2017). [PubMed: 28173627]
49. Thompson JN The Geographic Mosaic of Coevolution (Univ. Chicago Press, 2005).
50. Yoder JB & Nuismer SL When does coevolution promote diversification? Am. Nat 176, 802–817 (2010). [PubMed: 20950142]
51. Leffler EM et al. Multiple instances of ancient balancing selection shared between humans and chimpanzees. Science 339, 1578–1582 (2013). [PubMed: 23413192]
52. DeGiorgio M, Lohmueller KE & Nielsen R A model-based approach for identifying signatures of ancient balancing selection in genetic data. PLoS Genet. 10, e1004561 (2014). [PubMed: 25144706]
53. Takahata N A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc. Natl Acad. Sci. USA 87, 2419–2423 (1990). [PubMed: 2320564]
54. Clark AG Neutral behavior of shared polymorphism. Proc. Natl Acad. Sci. USA 94, 7730–7734 (1997). [PubMed: 9223256]
55. Wiuf C, Zhao K, Innan H & Nordborg M The probability and chromosomal extent of trans-specific polymorphism. Genetics 168, 2363–2372 (2004). [PubMed: 15371365]
56. Teixeira JC et al. Long-term balancing selection in LAD1 maintains a missense trans-species polymorphism in humans, chimpanzees, and bonobos. Mol. Biol. Evol 32, 1186–1196 (2015). [PubMed: 25605789]
57. Hill WG & Robertson A The effect of linkage on limits to artificial selection. Genet. Res 8, 269–294 (1966). [PubMed: 5980116]
58. Barton NH & Charlesworth B Why sex and recombination? Science 281, 1986–1990 (1998). [PubMed: 9748151]

59. Webster MT & Hurst LD Direct and indirect consequences of meiotic recombination: implications for genome evolution. Trends Genet. 28, 101–109 (2012). [PubMed: 22154475]
60. McGaugh SE et al. Recombination modulates how selection affects linked sites in *Drosophila*. PLoS Biol. 10, e1001422 (2012). [PubMed: 23152720]
61. Begun DJ & Aquadro CF Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356, 519–520 (1992). [PubMed: 1560824]
62. Schield DR et al. Snake recombination landscapes are directed by PRDM9 but concentrated in functional regions. Mol. Biol. Evol 37, 1272–1294 (2020). [PubMed: 31926008]
63. Mackessy SP Evolutionary trends in venom composition in the Western Rattlesnakes (*Crotalus viridis sensu lato*): toxicity vs. tenderizers. Toxicon 55, 1463–1474 (2010). [PubMed: 20227433]
64. Holding ML, Sovic MG, Colston TJ & Gibbs HL 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).
65. Schield DR et al. Allopatric divergence and secondary contact with gene flow: a recurring theme in rattlesnake speciation. Biol. J. Linn. Soc 128, 149–169 (2019).
66. Li H & Durbin R Inference of human population history from individual whole-genome sequences. Nature 475, 493–496 (2011). [PubMed: 21753753]
67. Voight BF, Kudaravalli S, Wen X & Pritchard JK A map of recent positive selection in the human genome. PLoS Biol. 4, e72 (2006). [PubMed: 16494531]
68. Sabeti PC et al. Detecting recent positive selection in the human genome from haplotype structure. Nature 419, 832–837 (2002). [PubMed: 12397357]
69. Ferrer-Admetlla A, Liang M, Korneliussen T & Nielsen R On detecting incomplete soft or hard selective sweeps using haplotype structure. Mol. Biol. Evol 31, 1275–1291 (2014). [PubMed: 24554778]
70. Gao Z, Przeworski M & Sella G Footprints of ancient-balanced polymorphisms in genetic variation data from closely related species. Evolution 69, 431–446 (2015). [PubMed: 25403856]
71. Siewert KM & Voight BF Detecting long-term balancing selection using allele frequency correlation. Mol. Biol. Evol 34, 2996–3005 (2017). [PubMed: 28981714]
72. Siewert KM & Voight BF BetaScan2: standardized statistics to detect balancing selection utilizing substitution data. Genome Biol. Evol 12, 3873–3877 (2020). [PubMed: 32011695]
73. Cheng X & DeGiorgio M Flexible mixture model approaches that accommodate footprint size variability for robust detection of balancing selection. Mol. Biol. Evol 37, 3267–3291 (2020). [PubMed: 32462188]
74. Cheng X & DeGiorgio M BalLeRMix+: mixture model approaches for robust joint identification of both positive selection and long-term balancing selection. Bioinformatics 38, 861–863 (2022).
75. Navarro A & Barton NH The effects of multilocus balancing selection on neutral variability. Genetics 161, 849–863 (2002). [PubMed: 12072479]
76. Fry BG From genome to ‘venome’: molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res. 15, 403–420 (2005). [PubMed: 15741511]
77. Bernardoni JL et al. Functional variability of snake venom metalloproteinases: adaptive advantages in targeting different prey and implications for human envenomation. PLoS ONE 9, e109651 (2014). [PubMed: 25313513]
78. Modahl CM, Mrinalini, Frietze S & Mackessy SP Adaptive evolution of distinct prey-specific toxin genes in rear-fanged snake venom. Proc. R. Soc. B 285, 20181003 (2018).
79. Klauber LM Rattlesnakes: Their Habits, Life Histories, and Influence on Mankind (Univ. California Press, 1956).
80. Holding ML et al. Phylogenetically diverse diets favor more complex venoms in North American pitvipers. Proc. Natl Acad. Sci. USA 118, e2015579118 (2021). [PubMed: 33875585]
81. Axel B, Pook CE, Harrison RA & Wolfgang W Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. Proc. R. Soc. B 276, 2443–2449 (2009).

82. Margres MJ et al. The Tiger Rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. Proc. Natl Acad. Sci. USA 118, e2014634118 (2021).
83. Mason AJ et al. Trait differentiation and modular toxin expression in palm-pitvipers. BMC Genomics 21, 147 (2020). [PubMed: 32046632]
84. Clarke BC The evolution of genetic diversity. Proc. R. Soc. B 205, 453–474 (1979). [PubMed: 42055]
85. Arbuckle K, de la Vega RCR & Casewell NR Coevolution takes the sting out of it: evolutionary biology and mechanisms of toxin resistance in animals. Toxicon 140, 118–131 (2017). [PubMed: 29111116]
86. Schield DR, Perry BW, Nikolakis ZL, Mackessy SP & Castoe TA Population genomic analyses confirm male-biased mutation rates in snakes. J. Hered 112, 221–227 (2021). [PubMed: 33502475]
87. Bolger AM, Lohse M & Usadel B Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014). [PubMed: 24695404]
88. Li H & Durbin R Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760 (2009). [PubMed: 19451168]
89. McKenna A et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010). [PubMed: 20644199]
90. Van der Auwera GA et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43, 10–11 (2013).
91. Danecek P et al. The variant call format and VCFtools. Bioinformatics 27, 2156–2158 (2011). [PubMed: 21653522]
92. Li H et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009). [PubMed: 19505943]
93. Delaneau O, Howie B, Cox AJ, Zagury J-F & Marchini J Haplotype estimation using sequencing reads. Am. J. Hum. Genet 93, 687–696 (2013). [PubMed: 24094745]
94. Dowell NL et al. Extremely divergent haplotypes in two toxin gene complexes encode alternative venom types within rattlesnake species. Curr. Biol 28, 1016–1026 (2018). [PubMed: 29576471]
95. Xie C & Tammi MT CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinf. 10, 80 (2009).
96. Wigginton JE, Cutler DJ & Abecasis GR A note on exact tests of Hardy–Weinberg equilibrium. Am. J. Hum. Genet 76, 887–893 (2005). [PubMed: 15789306]
97. Alexander DH, Novembre J & Lange K Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664 (2009). [PubMed: 19648217]
98. Purcell S et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet 81, 559–575 (2007). [PubMed: 17701901]
99. Green RE et al. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. Science 346, 1254449 (2014). [PubMed: 25504731]
100. Korunes KL & Samuk K pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Mol. Ecol. Resour 21, 1359–1368 (2021). [PubMed: 33453139]
101. Nei M & Roychoudhury AK Sampling variances of heterozygosity and genetic distance. Genetics 76, 379–390 (1974). [PubMed: 4822472]
102. Tajima F Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595 (1989). [PubMed: 2513255]
103. Gautier M & Vitalis R rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28, 1176–1177 (2012). [PubMed: 22402612]
104. Gautier M, Klassmann A & Vitalis R rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. Mol. Ecol. Resour 17, 78–90 (2017). [PubMed: 27863062]
105. Renaud G glactools: a command-line toolset for the management of genotype likelihoods and allele counts. Bioinformatics 34, 1398–1400 (2018). [PubMed: 29186325]
106. Haller BC & Messer PW SLiM 3: forward genetic simulations beyond the Wright–Fisher model. Mol. Biol. Evol 36, 632–637 (2019). [PubMed: 30517680]

107. Chan AH, Jenkins PA & Song YS Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. PLoS Genet. 8, e1003090 (2012). [PubMed: 23284288]

108. R Core Team. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2017).

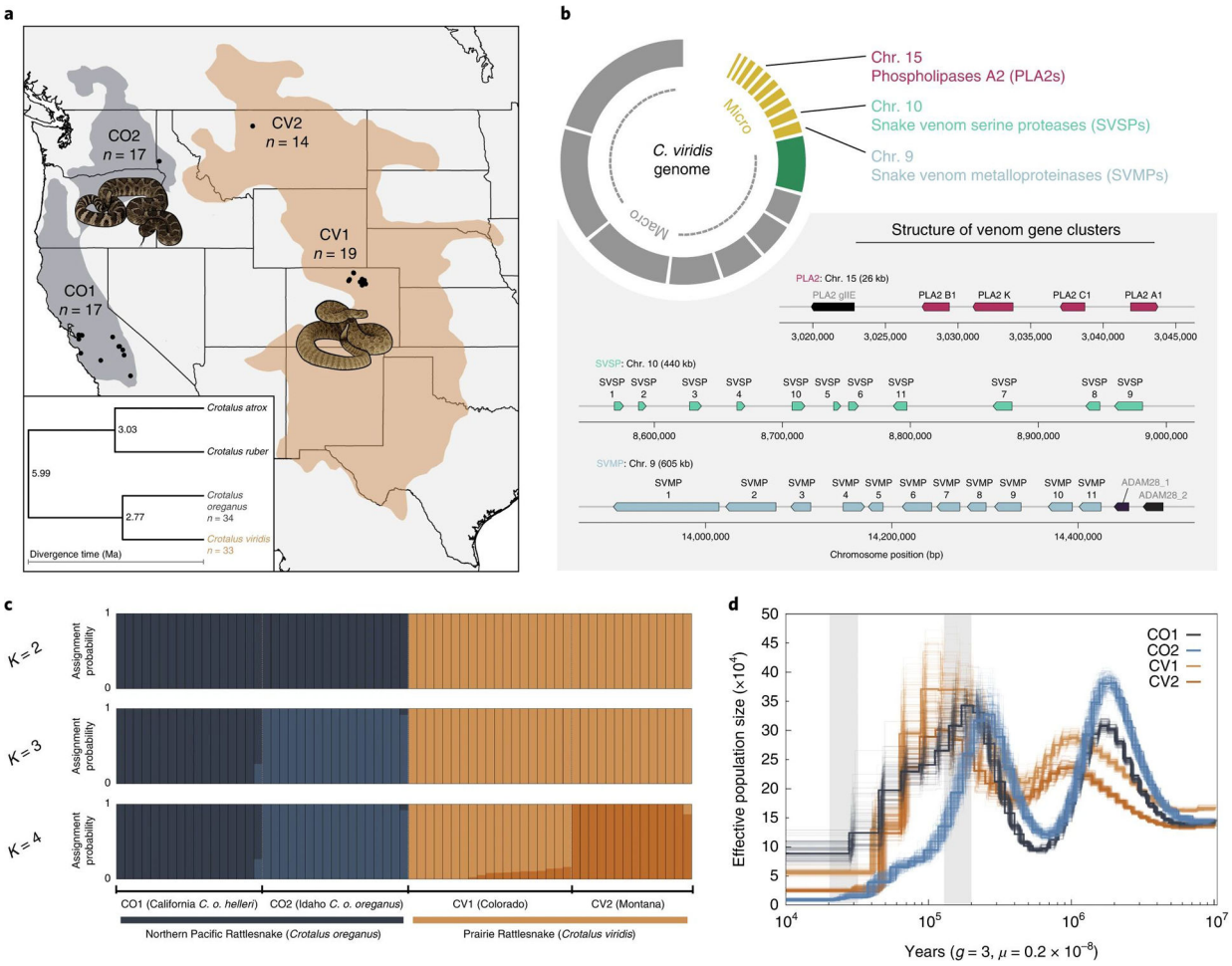


Fig. 1 |. Overview of the study system and estimates of population structure and historical demography.

a, Range map of the Prairie Rattlesnake (*C. viridis*, abbreviated 'CV') and the Northern Pacific Rattlesnake (*C. oreganus*, 'CO') in North America. Sampled localities are shown as black dots. The inset shows the phylogenetic tree with divergence time estimates as million years ago (Ma) for *C. viridis*, *C. oreganus* and outgroups, redrawn from ref.[62]. **b**, The genomic location and structure of major venom gene families. The SVMP, SVSP and PLA2 families are each linked to a separate microchromosome in the *C. viridis* genome assembly. Venom gene annotation tracks for each gene family region are shown below, with arrows denoting gene orientation. In the SVMP and PLA2 regions, non-venom paralogues are shown as black arrows. **c**, Population genetic structure within *C. viridis* and *C. oreganus* estimated using ADMIXTURE under $K = 2$–4 genetic cluster models. Vertical bars depict the assignment probability per individual to one or more K clusters. The best-supported model based on the cross-validation method was $K = 3$, but there was similar support for the $K = 4$ model. **d**, PSMC estimates of effective population size ($N_e \times 10^4$) through time for the four rattlesnake populations, scaled by generation time (g = 3). Estimates from full datasets are shown with bold lines, with faint lines representing individual bootstrap replicates. Grey shaded regions show the approximate timing of recent Pleistocene glacial periods.

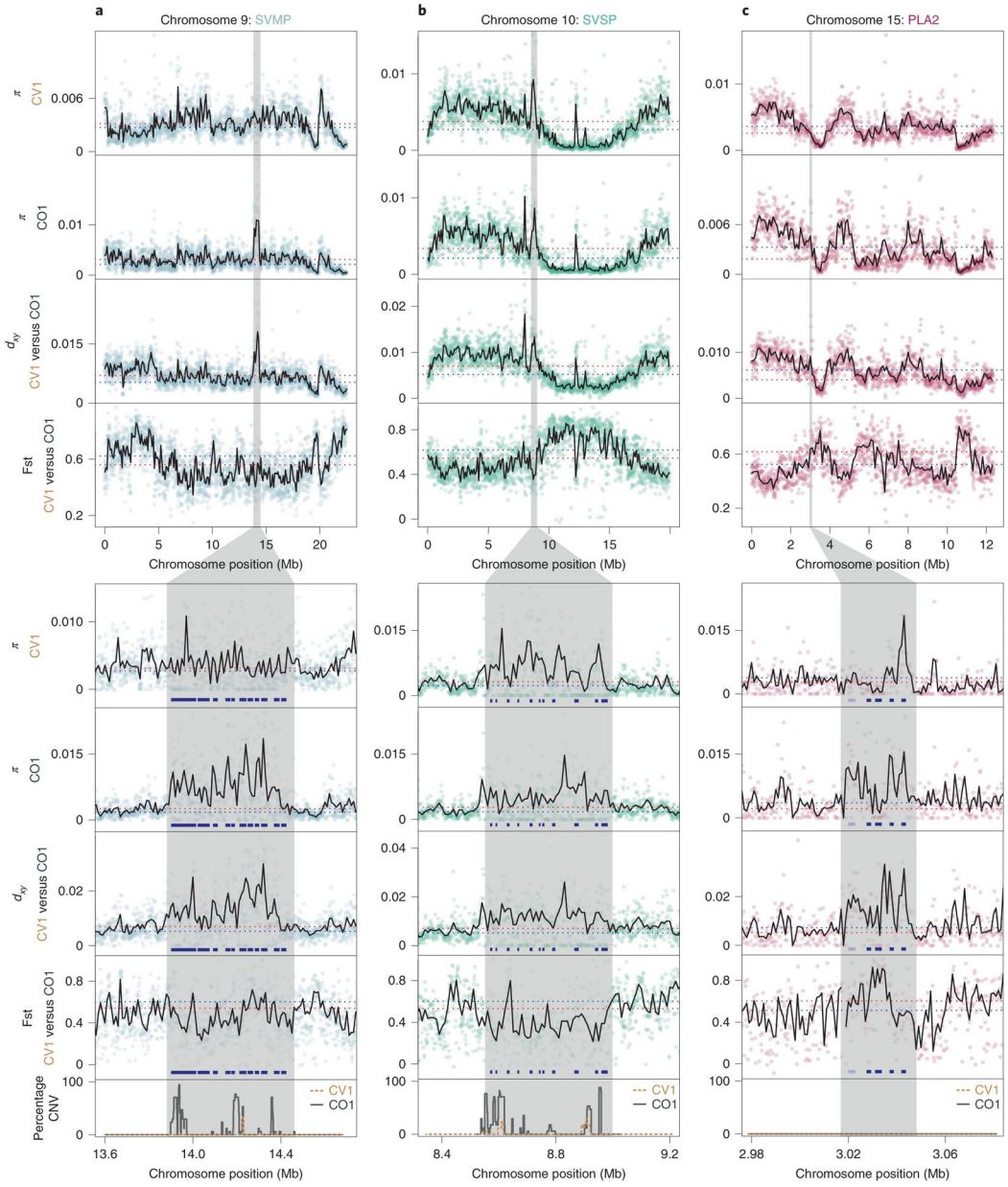


Fig. 2 |. Genome scans of genetic diversity within populations and differentiation between populations across chromosome housing major venom gene families (SvMPs, SvSPs and PLA2s). **a**, Sliding windows of nucleotide diversity (π) in *C. viridis* (CV1) and *C. oreganus* (CO1) populations and sequence divergence (d_{XY}) and relative differentiation (Fst) between CV1 and CO1 across chromosome 9 (upper panels) and in the SVMP region (lower panels). **b**, π, d_{XY} and Fst for CV1 and CO1 across chromosome 10 (upper) and in the SVSP region (lower). **c**, π, d_{XY} and Fst for CV1 and CO1 across chromosome 15 (upper) and in the PLA2 region (lower). Shaded points in upper panels show estimates in 10 kb windows and lines show estimates in 100 kb windows. In lower panels in **a** and **b**, shaded points show estimates in 1 kb windows and lines show estimates in 10 kb windows. In lower panels in **c**, shaded points show estimates in 250 bp windows and lines are estimates in 1 kb windows. The regions housing venom genes are shaded in grey in all panels. Chromosome-specific

and genome-wide mean values for each statistic are represented by blue and red dashed horizontal lines. The locations of individual venom genes are shown as blue boxes (lower panels). The non-venom homologue PLA2gIIE is shown in light purple. Panels at the very bottom show regional variation in the proportion of individuals per population with evidence of copy-number variation (percentage CNV) across the venom gene regions. Individual genotypes in detected CNV regions were masked.

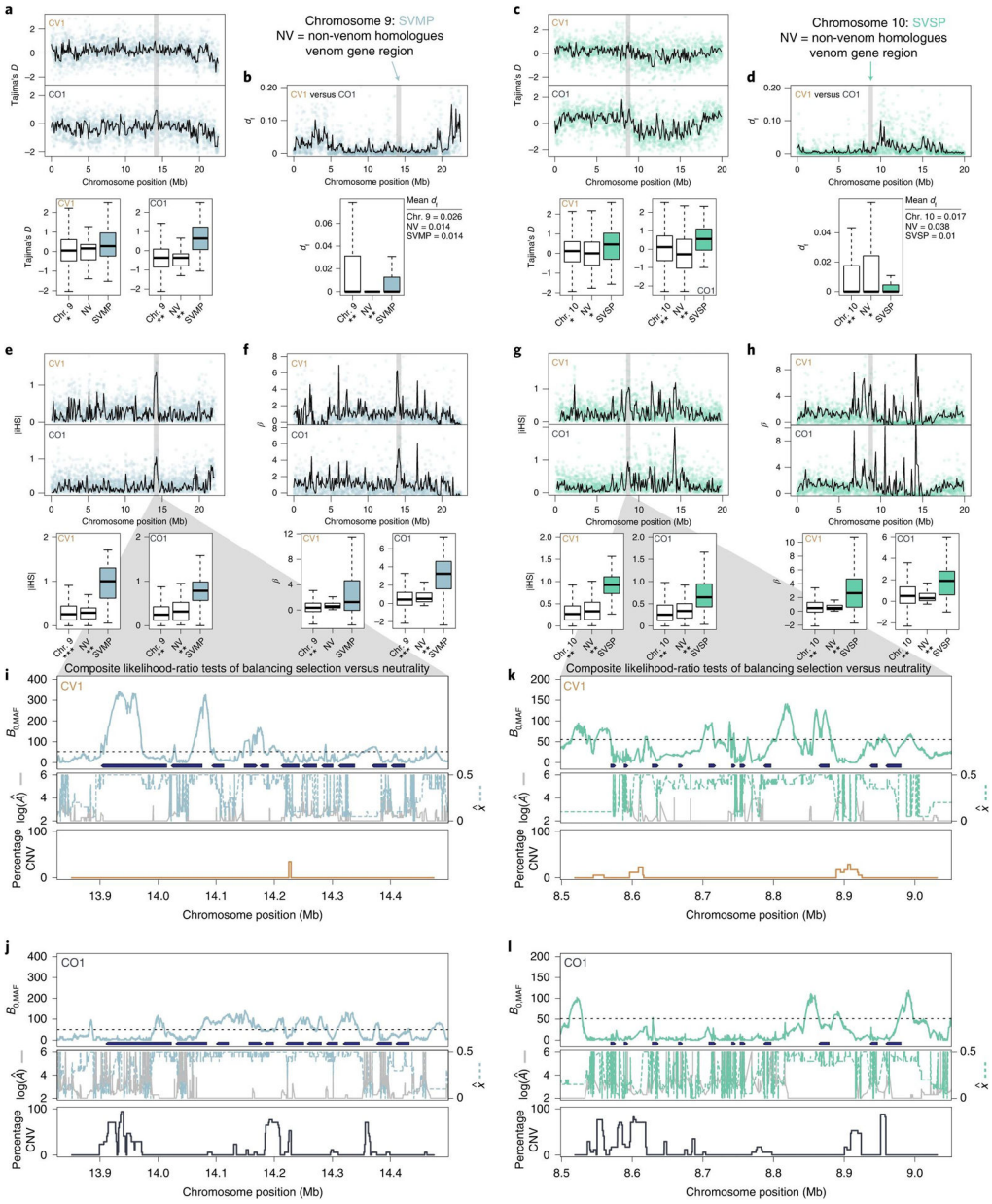


Fig. 3 |. Signatures of selection in venom gene regions, with comparisons to chromosomal backgrounds and non-venom homologues.

Signatures of selection in SVMP (**a**,**b**,**e**,**f**,**i**,**j**) and SVSP (**c**,**d**,**g**,**h**,**k**,**l**) venom gene regions. **a**, Tajima's *D* across chromosome 9 in *C. viridis* (CV1) and *C. oreganus* (CO1) populations (top). The SVMP region is shaded in grey. Boxplots below show distributions of Tajima's *D* for chromosome 9, non-venom homologues (NV) of the SVMP family and SVMPs. **b**, Proportion of fixed differences (d_f) between CV1 and CO1. **c**, Integrated haplotype statistics (|iHS|) in CV1 and CO1. **d**, ß statistic measuring allele frequency correlation in CV1 and CO1. **e**–**h**, Chromosome scans and distributions of each statistic for the SVSP region, the chromosome 10 background and SVSP non-venom homologues. Points in genome scan panels represent mean estimates in 10 kb sliding windows, and lines represent 100 kb windowed estimates. Boxplots in **a**–**h** show the median (horizontal lines), interquartile (box

limits) and range (whiskers). Asterisks indicate significant differences between venom gene regions and chromosome backgrounds and non-venom homologues based on two-tailed Welch's two-sample *t*-tests (Tajima's *D* and |iHS|) and Mann–Whitney *U* tests (d_f and ß) and n = 2,293 and 2,068 10 kb sliding windows for SVMP and SVSP comparisons, respectively (*$P < 0.05$, **$P < 0.001$, ***$P < 2.2 \times 10^{-16}$). Exact *P* values for comparisons can be found in Supplementary Table 5. **i**,**j**, Results of composite-likelihood ratio tests of balancing selection versus neutrality using $B_{0,MAF}$ scores in the SVMP region in CV1 and CO1. Top panels show $B_{0,MAF}$ scores, with higher values indicating greater evidence for balancing selection. Dark blue arrows show locations of venom genes. Middle panels show inferred footprint size, $\log(\widehat{A})$, as solid grey lines and equilibrium allele frequency, $\hat{x}$, as dashed lines. Bottom panels show variation in the proportion of individuals per population with masked genotypes in detected CNVs (% CNV). **k**,**l**, $B_{0,MAF}$ scores, $\log(\widehat{A})$, $\hat{x}$ and % CNV in the SVSP region in CV1 and CO1. Dashed lines in top panels of **i**–**l** show the genome-wide 95th quantile.

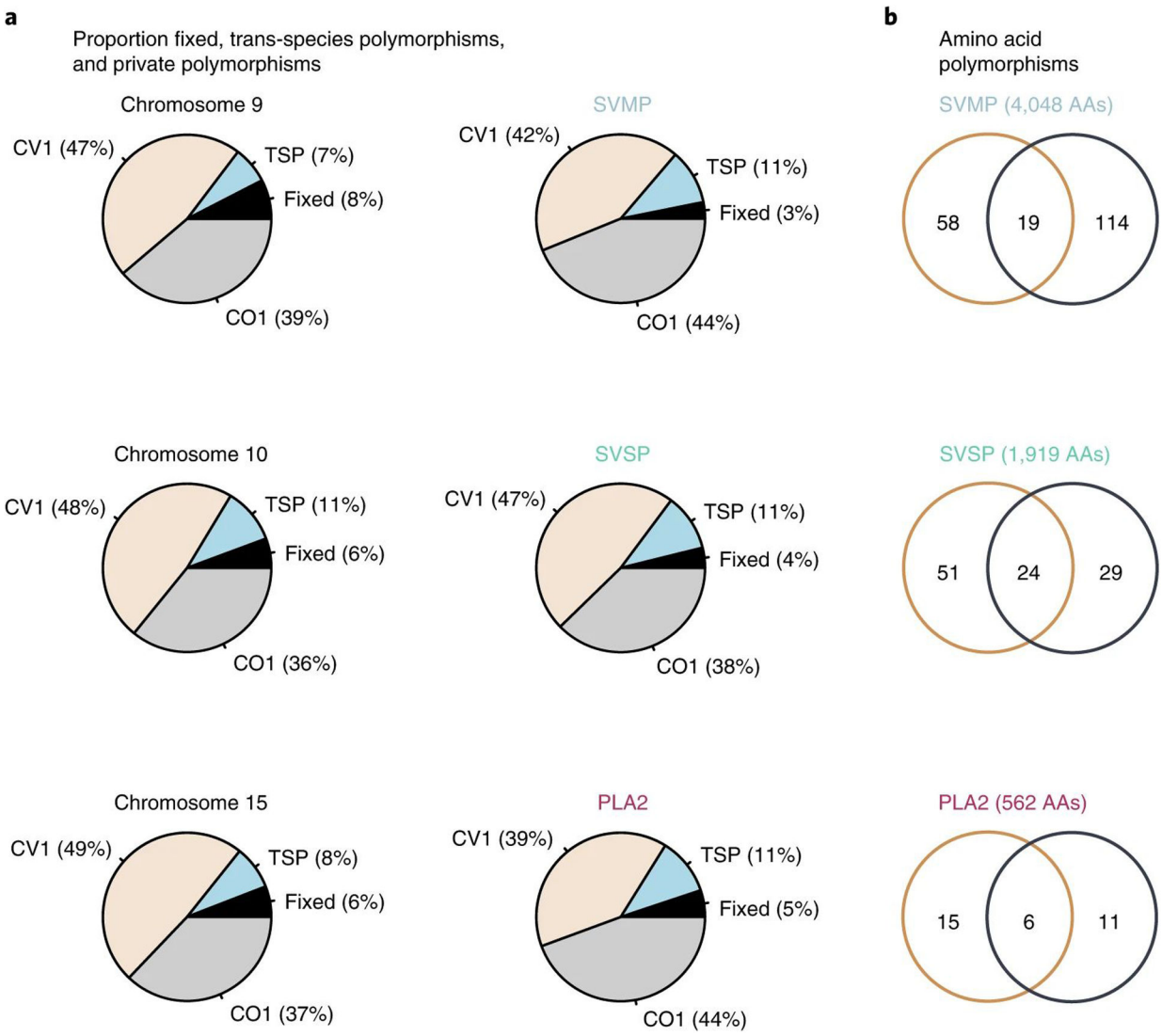


Fig. 4 |. Trans-species amino acid polymorphisms among SvMP, SvSP and PLA2 genes.

a, Relative proportions of fixed differences, trans-species polymorphisms (TSP) and private (CV1 and CO1) polymorphisms among variant sites shared between *C. viridis* and *C. oreganus* in chromosome backgrounds and venom gene regions. **b**, Numbers of observed trans-species (centre) and private amino acid (AA) polymorphisms in each venom gene family.

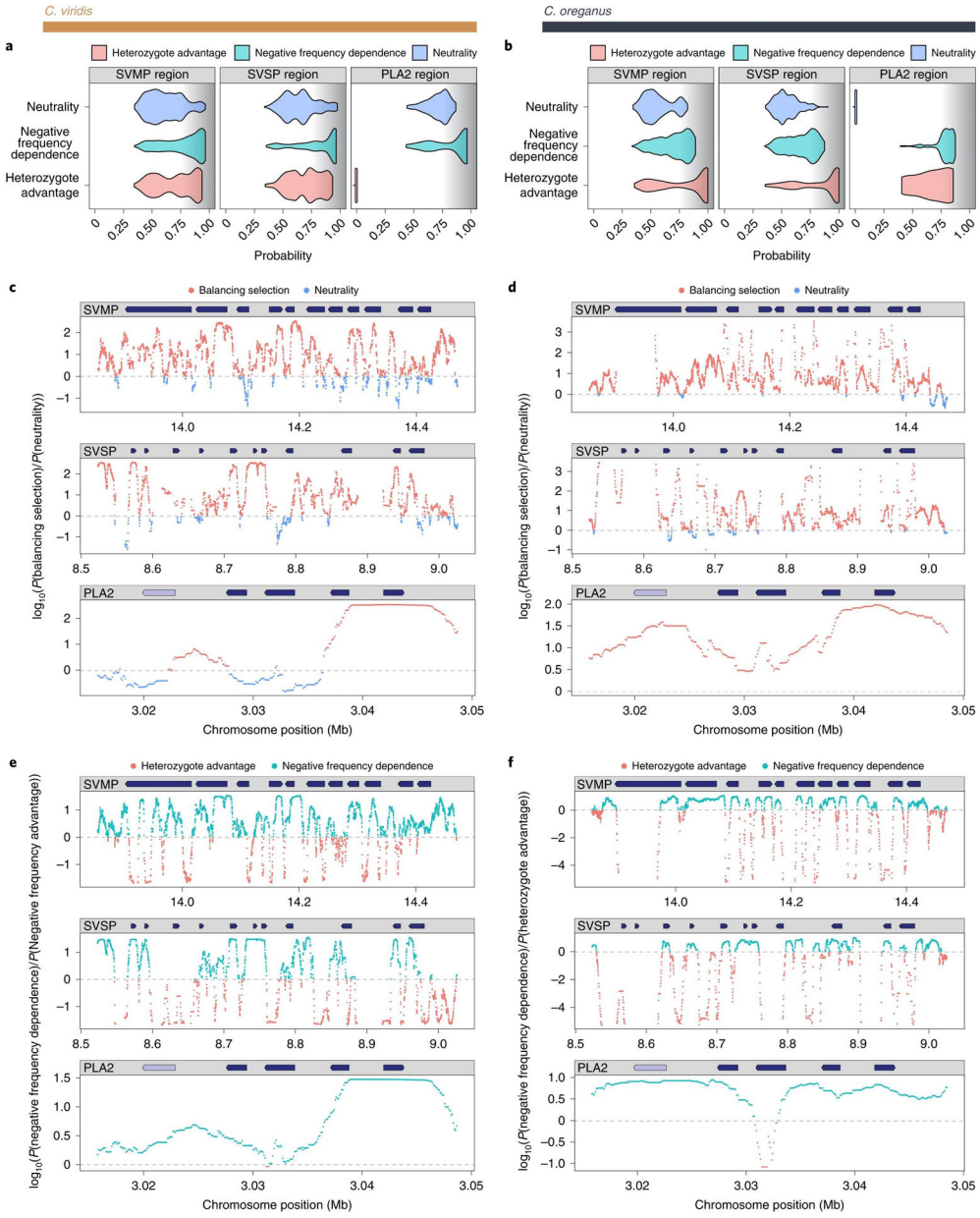


Fig. 5 |. Predicted probabilities of alternative evolutionary mechanisms (neutrality, negative frequency-dependent selection and heterozygote advantage) across SvMP, SvSP and PLA2 venom gene regions.

Predictions are based on training a neural network classification model with forward-time simulations and application of the trained model to empirical data (Supplementary Fig. 2). **a**,**b**, Violin plots show distributions of predicted probabilities of alternative mechanisms in windows of each venom region in *C. viridis* (**a**) and *C. oreganus* (**b**), respectively. **c**,**d**, Regional variation in the predicted probability of balancing selection versus neutrality across venom gene regions in *C. viridis* (**c**) and *C. oreganus* (**d**), respectively, measured as the $\log_{10}$ odds ratio. Points below the dashed lines indicate greater probability of neutrality, and points above the dashed lines indicate greater probability of balancing selection. **e**,**f**, Regional variation in the predicted probability of negative frequency dependence versus

heterozygote advantage in *C. viridis* (**e**) and *C. oreganus* (**f**), respectively, measured as the $\log_{10}$ odds ratio. Points below the dashed lines indicate greater probability of heterozygote advantage, and points above the dashed line indicate greater probability for negative frequency-dependent selection. Regions with missing points in **c**–**f** correspond to windows that were skipped due to too few SNPs (Methods). Dark blue arrows above scans in **c**–**f** show the location and directionality of venom genes.

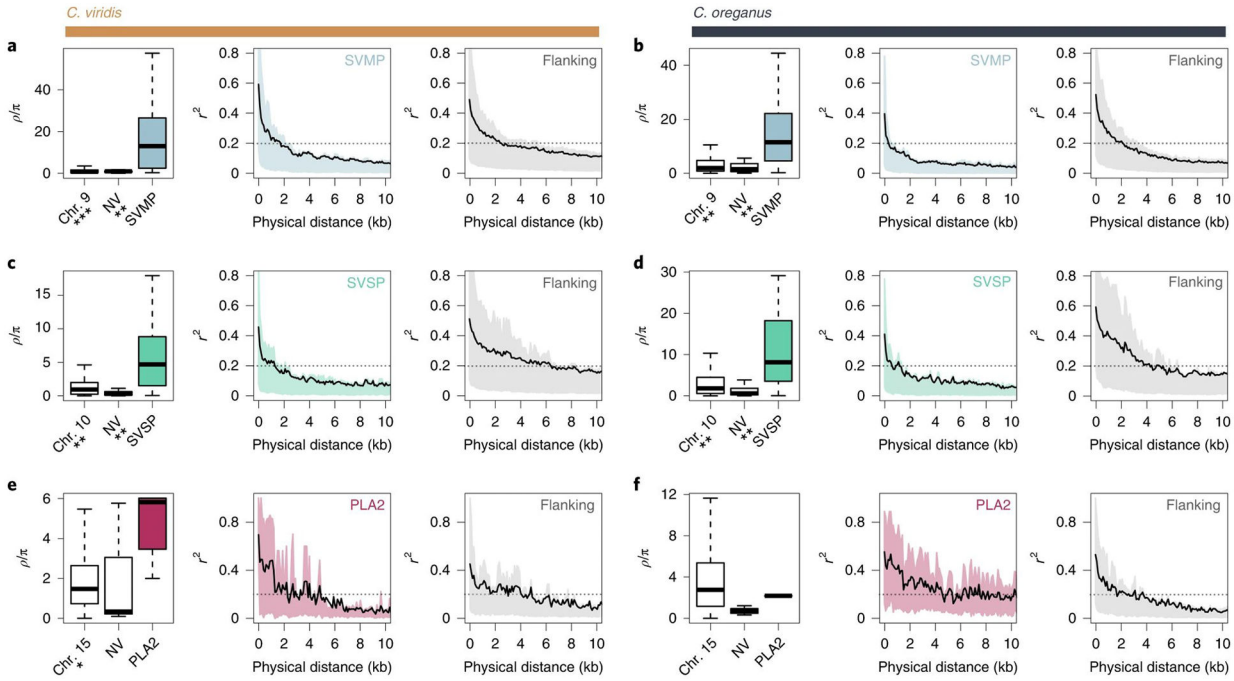


Fig. 6 |. Recombination rate variation in SvMP, SvSP and PLA2 venom gene regions.

a–**f**, Population-scaled recombination rate (π-corrected; ρ/π) and LD decay in venom regions compared background distributions and immediate flanking regions in *C. viridis* (**a**,**c**,**e**) and *C. oreganus* (**b**,**d**,**f**). Boxplots to the left in each panel show distributions of ρ/π in each major venom gene region compared with non-venom homologue (NV) and chromosomal backgrounds, with the median (horizontal lines), interquartile (box limits) and range (whiskers). Asterisks indicate significant differences between venom gene regions and chromosome backgrounds and non-venom homologues based on two-tailed Mann–Whitney *U* tests and n = 2,293, 2,068 and 1,272 10 kb sliding windows for SVMP, SVSP and PLA2 comparisons, respectively (*$P < 0.05$, **$P < 0.001$, ***$P < 2.2 \times 10^{-16}$). Exact *P* values for comparisons can be found in Supplementary Table 8. LD decay, measured as pairwise r^2 between SNPs with increasing physical distance, in venom and immediate flanking regions is shown in the centre and right plots in each panel. Black lines show average r^2 as a function of distance and shaded regions show the interquartile range of observed values. Horizontal dashed lines are fixed at $r^2 = 0.2$ for comparison between venom and flanking regions.