ORIGINAL ARTICLE

Title: Drift, selection, and adaptive variation in small populations of a threatened rattlesnake

Running title: Adaptive variation in small populations

Authors: Alexander Ochoa[1] | Michael Broe[1] | Emily Moriarty Lemmon[2] | Alan R. Lemmon[3] | Darin R. Rokyta[2] | H. Lisle Gibbs[1]

[1]Ohio Biodiversity Conservation Partnership and Department of Evolution, Ecology, and Organismal Biology, Ohio State University, Columbus, OH, USA

[2]Department of Biological Science, Florida State University, Tallahassee, FL, USA

[3]Department of Scientific Computing, Florida State University, Tallahassee, FL, USA

Correspondence

Alexander Ochoa, Ohio Biodiversity Conservation Partnership and Department of Evolution, Ecology, and Organismal Biology, Ohio State University, Columbus, OH, USA. Email: ochoa.43@osu.edu. ORCID ID: 0000-0002-3271-3684

and

H. Lisle Gibbs, Ohio Biodiversity Conservation Partnership and Department of Evolution, Ecology, and Organismal Biology, Ohio State University, Columbus, OH, USA. Email: gibbs.128@osu.edu. ORCID ID: 0000-0001-7461-3393

Abstract

An important goal of conservation genetics is to determine if the viability of small populations is reduced by a loss of adaptive variation due to genetic drift. Here, we assessed the impact of drift and selection on direct measures of adaptive variation (toxin loci encoding venom proteins) in the Eastern Massasauga rattlesnake (*Sistrurus catenatus*), a threatened reptile that exists in small isolated populations. We estimated levels of individual polymorphism in 46 toxin loci and 1467 control loci across 12 populations of this species, and compared the results with patterns of selection on the same loci following speciation of *S. catenatus* and its closest relative, the Western Massasauga (*S. tergeminus*). Multiple lines of evidence suggest that both drift and selection have had observable impacts on standing adaptive variation. In support of drift effects, we found little evidence for selection on toxin variation within populations and a significant positive relationship between current levels of adaptive variation and long-term and short-term estimates of effective population size. However, we also observed levels of directional selection on toxin loci among populations that are broadly similar to patterns predicted from interspecific selection analyses that predate the effects of recent drift, and that functional variation in these loci persists despite small short-term effective sizes. This suggests that much of the adaptive variation present in populations may represent an example of "drift debt," a non-equilibrium state where present-day levels of variation overestimate the amount of functional genetic diversity present in future populations.

KEYWORDS

Eastern Massasauga rattlesnake, *Sistrurus catenatus*, toxin loci, adaptive variation, genetic drift, selection, drift debt

1 | INTRODUCTION

Theory predicts that small populations can suffer loss of adaptive variation that will limit future population growth and viability (Frankham et al., 2017). However, identifying the extent to which threatened species are impacted by processes such as drift and inbreeding has proved difficult, in part because of methodological limitations for identifying and assaying the genetic variants that underlie adaptations and inbreeding depression in natural populations (Rockman, 2012). Recent advances in our ability to collect and analyze genome scale data (Barrett & Hoekstra, 2011) have dramatically improved our ability to identify such variants and to assess the impact of drift and inbreeding on adaptive variation. For instance, multiple studies have documented genetic costs related to these processes in some endangered species (Blomqvist, Pauliny, Larsson, & Flodin, 2010; Norén, Godoy, Dalén, Meijer, & Angerbjörn, 2016). But other species with similar demographic histories appear to experience limited impacts due to the presence of genomic hotspots where balancing selection has remained strong or through the purging of highly deleterious mutations following severe bottleneck events (Aguilar et al., 2004; Robinson et al., 2016; Bennazzo et al., 2017; Grossen, Guillaume, Keller, & Croll, 2020). These contrasting results emphasize that understanding the evolutionary forces that affect levels of adaptive

variation in small populations remains an important question for conservation genetic research.

A reason why certain adaptations could persist despite strong drift lies in the increasing appreciation that the genetic architecture of fitness-related variation can lead to different evolutionary responses to drift by different adaptive variants (Barrett & Hoekstra, 2011). For example, in one widely promoted model for the evolution of adaptations, the effect size distribution of causal variants that underlie an adaptive trait has a negative exponential distribution with few large-effect and many small-effect loci to fitness (Orr, 1998; but see discussion in Rockman, 2012). This pattern is similar to the distribution of selection intensities on molecular variants in natural populations (Thurman & Barrett, 2016) and suggests some adaptive variants may be more resistant than others to a given level of drift.

More specifically, conservation biologists have long focused on inferring the impact of drift on adaptive variation from measures of effective population size, N_e, alone (e.g., the 50/500 rule for assessing short-term and long-term viability of populations; Jamieson & Allendorf, 2012). Yet, in general, the impact of drift on adaptive variants is given by the product $N_e s$, where s is the selection differential acting on that variant (Wright, 1931; Barton & Partridge, 2000). This means that the impact of a given level of drift on fitness-related variation in a small population is not due to N_e alone, but is also dependent on the fitness-effect size across loci, with large-effect variants being more resistant to drift than small-effect variants. As such, taking into account the genetic architecture of fitness variation may be important for understanding differences in the levels of adaptive variation

present in these populations (Funk, Forester, Converse, Darst, & Morey, 2019; Mable, 2019).

The Eastern Massasauga (*Sistrurus catenatus*) is a small rattlesnake found in eastern North America. Population declines throughout its range due to habitat fragmentation and destruction have led to the listing of this species as threatened under the Endangered Species Act in the U.S. (U.S. Fish and Wildlife Service, 2016) and as a species at risk in Canada (Government of Canada, 2009). This species exhibits little phylogeographic structure across its range (Sovic, Fries, & Gibbs, 2016), and so the relevant management units within this species are individual populations. Recent work by Sovic, Fries, Martin, and Gibbs (2019) has shown that the contemporary N_e values of almost all populations of this species are < 50, suggesting that if these sizes remain the same or become smaller, drift will lead to a substantial loss of genetic variation in these populations over the next 100 years.

The influence of drift on fitness-related variation in *S. catenatus* remains unclear for several reasons. First, long-term N_e estimates for many of these populations are an order of magnitude higher than that for short-term estimates, suggesting weaker effects of drift over longer time scales (Gibbs & Chiucchi, 2012; Sovic et al., 2019). Second, heterozygosity-fitness correlations based on neutral genetic markers and body condition show few positive relationships consistent with the negative effects of drift and inbreeding due to small population sizes (Gibbs & Chiucchi, 2012; Sovic et al., 2019), although the interpretation of this result depends on the assumption that body condition is a reliable index of fitness. Both results suggest that the impacts of drift on current levels of adaptive variation may be

limited, but this question can only be resolved through direct measure of genetic variants plausibly linked to fitness and by examining the evolutionary forces that have shaped this variation.

To date, only Jaeger et al. (2016) have used this approach in *S. catenatus* by evaluating allelic variation at a single exon in one major histocompatibility complex (MHC) locus and multiple microsatellite loci in three populations in Illinois. They found evidence for both balancing selection and drift jointly determining levels of MHC diversity. However, their inability to precisely determine the genetic basis of the MHC variation and the small number of populations examined limit the usefulness of these results. A valuable next step would be to analyze functional variation at the genomic level and then estimate the relative impacts of selection versus drift as drivers of levels of adaptive variation within and among populations (Benazzo et al., 2017).

One method to identify adaptive genetic variation is to use a "top-down" approach that starts with defining phenotypic traits of known functional importance and ends with discovering the genes underlying those traits (Barrett & Hoekstra, 2011). In rattlesnakes, the proteins that make up whole venom are an exceptionally important functional trait because they play a key role in the capture and digestion of prey by individual snakes (Casewell, Wüster, Vonk, Harrison, & Fry, 2013).

Snake venom shows population-level patterns of variation in whole venom protein composition, which suggests that the loci encoding such proteins could be under strong selection at the intraspecific level in Eastern Massasauaga rattlesnakes. For example, as in many species of venomous snakes (Chippaux, Williams, & White, 1991), there is significant

variation in individual venom proteins both within and among populations of *S. catenatus* (Gibbs, Sanz, & Calvete, 2009; Gibbs & Chiucchi, 2011). Further, population-level differentiation in venom proteins is not correlated with levels of neutral genetic differentiation or N_e, which argues that patterns of venom variation are not simply a consequence of population structure (Gibbs & Chiucchi, 2011). Finally, functional analyses of venom in a closely related species (*S. miliarius*) demonstrate that differences in venom phenotype translate into differences in the ability to kill specific prey, suggesting that this variation is the product of natural selection. Specifically, Smiley-Walters, Farrell, and Gibbs (2017) provided experimental evidence that population differences in whole venom represent local adaptations, while Smiley-Walters, Farrell, and Gibbs (2019) showed differences in the ability of venom from individual snakes from the same population to kill lizard prey. Together these results provide strong evidence for a possible role of selection on venom protein coding genes in generating these patterns of variation. In particular, they suggest directional selection for population differences in venom genes combined with balancing selection to maintain levels of protein polymorphism within populations. Such patterns should be evident when examining functional substitutions in loci encoding venom proteins within and among populations (but see Margres, Bigelow, Lemmon, Lemmon, & Rokyta, 2017a; Rautsaw et al., 2019).

A challenge in assessing variation in venom genes is that many of these loci belong to gene families consisting of large numbers of functional paralogs (Lynch, 2007; Casewell, Wagstaff, Harrison, Renjifo, & Wüster, 2011; Rokyta, Wray, & Margres, 2013). This makes the assessment of variation at specific loci challenging due to the difficulty of distinguishing

polymorphisms from the same locus from variation between distinct, paralogous loci (McKinney, Waples, Seeb, & Seeb, 2017) as is the case for variation in MHC in non-model species (Babik, 2010). Margres et al. (2017a) recently described a set of toxin gene-based capture probes that allows a comprehensive assay of sequence variation at the exon level for most venom genes found in pit vipers. They also described probes for > 1000 nontoxin, control loci that can be used to estimate demographic effects alone on variation (see below). Data from both probe sets open the door to using the genetic variation underlying venom genes as a direct measure of adaptive variation in *S. catenatus* and to assessing the evolutionary forces that influence functional variation in small populations of this threatened snake.

In this study, we used capture probe-based methods (Lemmon, Emme, & Lemmon, 2012; Ruane, Raxworthy, Lemmon, Lemmon, & Burbrink, 2015) to assess variation in the loci encoding venom proteins and in a large number of control loci in individuals from 12 *S. catenatus* populations across its range. We used variation in toxin loci as a measure of adaptive genetic variation to address two questions relevant to evaluating the impact of drift and selection on adaptive variation in small populations. First, do levels of variation in toxin and control loci show a positive association with long-term and/or short-term measures of N_e as predicted if drift is a significant evolutionary force in these snake populations? Second, what are patterns of selection within and among populations on functional variation (defined as nonsynonymous substitutions) at toxin loci? To provide a needed historical perspective to help interpret intraspecific patterns of selection on toxin

loci, we also analyzed selection on toxin loci following speciation between *S. catenatus* and its sister taxon, the Western Massasauaga rattlesnake (*S. tergeminus*).

2 | MATERIALS AND METHODS

2.1 | Samples and DNA extraction

We collected 100 µl whole blood from 93 *S. catenatus* individuals from 12 populations spanning the range of this species across the U.S. and Canada—see Figure 1 for geographical locations, population names, and sample sizes. We also collected 100 µl of whole blood from one *S. tergeminus* individual from Cheyenne Bottoms, Kansas, for comparative purposes. We stored blood samples in 90% ethanol and isolated genomic DNA from them using a phenol-chloroform extraction protocol as described in Sovic et al. (2016).

2.2 | Capture probes, library preparation, and sequencing

For the *S. catenatus* samples, we used capture probe-based methods (Lemmon et al., 2012; Margres et al., 2017a; Ruane et al., 2015) to assay single nucleotide polymorphism (SNP) variation in two types of loci: i) toxin loci, derived from exons of genes encoding snake venom proteins, which we assume represent putative adaptive variation; and ii) control loci, derived from a small number of conserved coding loci and a much larger number of noncoding sequences from throughout the genome (Margres et al., 2017a; see below for specific numbers of loci in each class). We refer to the latter as “control” loci because we assume they reflect the demographic history of populations and, hence, provide an

198 evolutionary context for interpreting patterns of variation in the toxin loci (Margres et al.,
199 2017a). For most analyses, we pooled information from the different classes of control loci
200 after confirming that they yielded similar results (see below). For a small number of
201 analyses it was necessary to compare nonsynonymous and synonymous substitution
202 patterns between toxin and control loci. In these cases, we analyzed the conserved coding
203 loci separately.

204 To generate these data, we used sets of 120-bp tiled probes (Margres et al., 2017a)
205 to capture DNA sequences from 16 venom gene families (206 exons) and 1617 control loci
206 (200 unlinked exons, 348 anchored regions, and 1069 anonymous sequences) from each
207 sample at the Center for Anchored Phylogenomics at Florida State University
208 (www.anchoredphylogeny.com). Briefly, genomic DNA was sonicated to ~340-bp
209 fragments and individual DNA libraries were indexed, amplified, purified, and pooled at
210 equal quantities (24 samples per pool), as described in Lemmon et al. (2012). We then used
211 an Agilent Custom SureSelect kit (Agilent Technologies) to hybridize the probes against
212 each multi-sample pool and to collect the fraction of bound DNA fragments. Enriched pools
213 were sequenced on independent PE200 Illumina HiSeq2500 lanes at the Translational
214 Science Laboratory of the College of Medicine at Florida State University.

215

216 **2.3 | Quality control and data processing**

217 We used Trimmomatic v.0.38 (Bolger, Lohse, & Usadel, 2014) to remove the adapter
218 sequences from the raw PE reads, eliminate leading and trailing low quality (< 20) or "N"
219 bases, scan reads with a 4-base wide sliding window and clip when the average quality per

base dropped to < 20, and remove reads with either an average quality < 30 or with a sequence length < 50 bases. We then used Musket v.1.1 (Liu, Schröder, & Schmidt, 2013) to improve data quality by detecting and correcting sequencing errors. As such, we computed the number of occurrences of any given 21-mer, estimated the coverage cut-off value to separate spurious from true 21-mers, and corrected for individual erroneous bases in the spurious 21-mers.

2.4 | Mapping and variant calling

In the absence of a reference genome, capture-based data is typically assembled *de novo* (Margres et al., 2017a, b) or quasi-*de novo*, where reads are mapped directly to the probe sequences and then extended into their flanking regions (Lemmon et al., 2012; Prum et al., 2015). Although these methods are suitable for assembling single-copy loci, they have limitations for assembling multi-copy, paralogous loci, including most venom genes (Casewell, Harrison, Wüster, & Wagstaff, 2009). First, for any given multi-copy locus, the number of copies in the genome and thus, the number of copies targeted by the probes, are unknown. Second, reads derived from homologous conserved regions across copy variants are likely to be included in chimeric alignments. Third, gene discontinuities, caused by inherent non-overlapping probe flanking sequences or by probe binding inefficiencies to hypervariable regions, could further complicate exon correspondence to individual paralogs.

Therefore, instead of creating *de novo* assemblies from capture-based data, we used BWA v. 0.7.17 (Li & Durbin, 2009) to map the processed PE reads against a whole-genome

assembly for *S. catenatus* (Broe et al., in prep.). This reference genome was generated from combining short (PE150, Illumina HiSeq X) and long (read N50 = 18 Kb, PacBio) reads with MaSuRCA v.3.2.4 (Zimin et al., 2013). We inferred toxin and control locus boundaries with Geneious v.10.2.2 (Kearse et al., 2012) after performing BLAST searches (Altschul, Gish, Miller, Myers, & Lipman, 1990) for each probe sequence. Visual inspections of these annotations confirmed contiguity for most toxin loci.

We used SAMtools v.1.4 (Li et al., 2009) to remove potential PCR duplicates, eliminate reads with a mapping quality < 20, and filter out reads with unmapped mates. To reduce possible biases introduced by chimeric alignments, we retrieved the mapped reads from the targeted genomic regions and used Geneious to remap them against the reference genome after applying stricter parameters. Specifically, we retained only remapped reads with no indels, with mates clustered within a window size of 170–510 bases, and with ≤ 1, 2, and 5% mismatches from three independent runs (one per mismatch threshold).

Comparisons among runs for toxin gene families with different numbers of copy variants (e.g., *PDE4*: 1 locus; *PLA2*: 6 loci; and *SVMP*: 12 loci [Table S1]) revealed a trade-off between mapping sensitivities and sequence coverages as the number of copy variants per locus progressed (Figure S1). For instance, coverage distributions for the *PDE4* locus remained unaltered when using either 1, 2, or 5% mismatch thresholds, whereas coverage distributions for the *PLA2* and *SVMP* loci were positively correlated with these thresholds (Figure S1). In all cases, but especially for multi-copy loci, we also detected heterogeneous coverage distributions among exons (Figure S1).

Since additional observations of the reference-based assemblies confirmed fewer chimeric alignments whenever mismatch thresholds were more restricted, we used the "UnifiedGenotyper" model implemented in GATK v.4.1.0.0 (McKenna et al., 2010) to define SNPs in the data generated with ≤ 1% mapping mismatches. We excluded potentially variable sites if they contained > 2 alleles or had a minor allele frequency < 0.05, quality by depth < 2.0, strand bias > 40.0, root mean square of mapping quality < 20.0, or haplotype score > 12.0 (Campagna et al., 2017).

We used VCFtools v.0.1.13 (Danecek et al., 2011) to mask individual genotypes with coverages < 5× (Jones et al., 2018) and to remove polymorphic and monomorphic sites not represented in > 50% of the sample set. Ultimately, we only included sites contained within the exon boundaries of the toxin loci and sites contained within 1000 bases upstream and downstream of the control loci. We excluded control loci with < 200 validated sites from any posterior analysis.

2.5 | Nucleotide variation in relation to N_e estimates

We used VCFtools to obtain allele frequency estimates across sites for each locus and population. We then estimated nucleotide diversity, π, as the mean expected heterozygosity across sites and, ultimately, weighted loci with respect to the potential number of sequences from each population, $2n/(2n-1)$, to account for sample size variation (Nei, 1987).

To assess the relationship between levels of standing variation and N_e, we regressed population-specific measures of π for three classes of SNPs (toxin nonsynonymous, toxin

synonymous, and control) against long-term and short-term estimates of N_e generated from RADseq data, as reported by Sovic et al. (2019). Long-term estimates of N_e evaluate the parameter over evolutionary timescales and are more relevant to interpreting the historical impact of selection and drift on variation, whereas short-term measures of N_e estimate the parameter over recent timescales (e.g. < 5 generations) and is more relevant as a measure of contemporary impacts of drift (Hare et al., 2011). Long-term N_e values were estimated using fastsimcoal v.2.5.2 (Excoffier, Dupanloup, Huerta-Sanchez, Sousa, & Foll, 2013), which uses maximum-likelihood methods to estimate parameters based on the site frequency spectrum calculated from population genotypes. Short-term N_e values were generated using the LDNe method, which estimates N_e based on patterns of linkage disequilibrium across loci (Waples & Do, 2008). Significant positive associations with either measure of N_e would suggest that drift has been and/or is an important evolutionary force shaping current levels of genetic variation within populations.

2.6 | Selection within and among *S. catenatus* populations

To assess selection on toxin gene variation within populations, we calculated Tajima's *D* statistic (Tajima, 1989) for toxin (nonsynonymous SNPs only) and control loci. Specifically, we weighted absolute π estimates with respect to the number of segregating sites for each locus (Tajima, 1989). We used 95% confidence interval (CI) distributions of the control loci as our neutral expectation. We then characterized toxin loci to be under balancing or directional/purifying selection if they showed values greater or less than the 95% CI, respectively.

We also performed Spearman rank correlations to test for the consistency of the magnitude of *D* values for toxin loci between population pairs. Our logic was that if selection consistently acted on a given set of loci between populations, particularly for those in close geographic proximity, then values of the *D* statistic should be correlated. In contrast, if drift was the primary force shaping functional variation within populations, then it should act on loci independently, and this should be reflected in a lack of correlation between locus-specific *D* values for population pairs. To account for multiple comparisons, we adjusted the resulting *P* values using the false discovery rate method (Benjamini & Hochberg, 1995).

To assay diversifying selection leading to local adaptation in toxin loci, we used Arlequin v.3.5.2.2 (Excoffier & Lischer, 2010) to estimate pooled F_{ST} values among populations for individual SNPs. Following Margres et al. (2017b), we detected selection by comparing toxin and control F_{ST} distributions. We classified toxin nonsynonymous SNPs with F_{ST} values greater than the 95th-percentile F_{ST} value of the control distribution as being under diversifying selection. We averaged SNP-specific toxin nonsynonymous and control F_{ST} values within loci to perform an equivalent comparison at the locus level. Unlike the within-population analyses, no adjustment of *P* values for multiple tests is required because this analysis comprises a single test that involves one comparison of two distributions.

Finally, we functionally characterized outlier toxin nonsynonyous SNPs as radical if the new amino acid resulted in a change for one or more of the following classifications: amino acids grouped by charge alone, polarity alone, and polarity and volume combined

(Zhang, 2000). We classified changes that resulted in no shift between any of these groups as conservative.

2.7 | Evolutionary history of selection on toxin loci

Few previous studies have analyzed selection on the complete set of genes encoding venom proteins within a single species (but see Margres et al., 2017b). As such, we have few results on which to base our expectations of patterns of selection across toxin loci within and among populations of *S. catenatus*. Insights about expected patterns of selection may come from documenting how selection acts on venom genes following speciation between closely related sister taxa. Although such patterns reflect evolutionary processes occurring over longer timescales than adaptation within species, they nonetheless reflect general patterns of how selection operates on a common set of potentially adaptive variants in ecologically similar entities recently.

To characterize the evolutionary history of selection on toxin genes in *S. catenatus*, we compared toxin and control coding sequences from the *S. catenatus* genome assembly with homologous regions in a *S. tergeminus* reference-based genome assembly. This assembly was generated from whole-genome PE150 Illumina reads mapped to the *S. catenatus* genome as previously outlined.

To analyze patterns of selection on toxin and control coding loci, we used SnIPRE (Eilertson, Booth, & Bustamante, 2012), which models nonsynonymous and synonymous polymorphism (*S. catenatus* only) and divergence (*S. catenatus* vs. *S. tergeminus*) counts in a McDonald-Kreitman (McDonald & Kreitman, 1991) test framework. We estimated the

"Bayesian selection effect" statistic ($\gamma = 2N_e s$) across exons for toxin and control loci separately after 250,000 iterations and a 40% burnin. For this analysis, we grouped exons from the same locus together (applicable only to the toxin loci), excluded codons with missing information, eliminated codons not represented in either species, and removed loci lacking both intraspecific and interspecific variation. We considered significant positive γ values as evidence for directional selection and negative values as indicating selection on slightly deleterious nonsynonymous mutations consistent with purifying selection (Fijarczyk, Dudek, & Babik, 2016).

3 | RESULTS

3.1 | Distribution of SNPs in toxin and control loci

We generated SNP data from 93 individuals for 46 toxin loci from 16 gene families and for 1467 control loci consisting of 171 unlinked exons with flanking intronic sequences, 344 anchored regions, and 952 anonymous sequences. In subsequent analyses using control loci, we combined data from coding and noncoding sequences because coding sequences only contributed ~1% SNPs of the total control dataset and excluding data from these loci did not change any results (data not shown). Overall, we assayed variation at 40,854 toxin coding sites (i.e., ~81% of all potential sites in exons for which probes had been designed to assay; Table S1) and 1,446,069 control coding and noncoding sites.

372 Across all individuals, we identified 367 SNPs in the toxin loci (215 nonsynonymous
373 and 152 synonymous substitutions) and 6335 SNPs in the control loci (13 nonsynonymous,
374 48 synonymous, and 6274 noncoding substitutions). Overall, toxin loci were more variable
375 than control loci, with the former containing twice as many SNPs per site and five times as
376 many nonsynonymous-to-synonymous SNPs.
377
378 **3.2 | Nucleotide variation in relation to N_e estimates**
379 Representing nucleotide diversity estimates as π (x 10^{-3}), nonsynonymous variation in toxin
380 loci ranged from π = 0.80 (ROME) to π = 1.24 (BPNP), with a mean π = 1.00 (SE = 0.04)
381 across populations (Table 1). Synonymous variation in these loci was lower, ranging from π
382 = 0.53 (PROF) to π = 0.87 (KPWA), with a mean π = 0.69 (SE = 0.04) across populations
383 (Table 1). Control loci showed similar levels of nucleotide diversity found in
384 nonsynonymous variation in toxin loci (range π = 0.75 [CICE] to π = 1.16 [KPWA and
385 BPNP]; mean π = 0.99, SE = 0.03) (Table 1).
386 All three population-specific π measures were positively correlated with both long-
387 term and short-term estimates of N_e, with R values ≥ 0.32 in all cases (Figure 2a and b).
388 However, these correlations were only significant for associations of toxin nonsynonymous
389 variation with long-term (R = 0.63, P = 0.04; Figure 2a) and short-term (R = 0.67, P = 0.03;
390 Figure 2b) estimates of N_e.
391 Interpreting these results is complicated by the fact that long-term and short-term
392 estimates of N_e for individual populations are significantly correlated with each other (R =
393 0.82, P < 0.01). To assess the independent associations between different measures of N_e

394 and a given measure of polymorphism, we conducted a partial correlation analysis that
395 included both long-term and short-term measures of N_e for a given population and one of
396 each of the three measures of polymorphism. While all *R* values remained positive (range of
397 partial *R* values = 0.03–0.35), none were significant (all $P > 0.05$). Overall, our results
398 suggest that drift has weak but detectable genome-wide impacts on levels of both adaptive
399 and neutral genetic variation in these populations, but the time scale of these effects is
400 unclear.
401

402 **3.3 | Selection within *S. catenatus* populations**

403 We characterized selection on toxin loci within populations by comparing *D* statistics
404 between toxin and control loci after assuming that control loci reflect background levels of
405 variation due to demographic processes. For the toxin loci, we focused on variation defined
406 by potentially functionally significant nonsynonymous SNPs. We used data from 40 toxin
407 loci (six toxin loci lacked nonsynonymous SNPs) and 1227 control loci (240 control loci
408 lacked SNPs). As such, toxin and control loci presented 5.4 and 5.2 SNPs per locus,
409 respectively.

410 Our results suggest that only a small fraction of toxin loci, if any, are under
411 significant selection within populations (Figure 3a). Only five loci (equivalent to ~1% of the
412 entire toxin locus data set across 12 populations) had *D* values that fell outside the 95% CI
413 estimated from variation in the control loci. Each significant *D* value was for a different
414 locus in each population, and all had *D* values less than the null expectation consistent with
415 directional/purifying selection acting on these loci. Finally, after adjusting *P* values for

multiple comparisons, there were no general significant correlations in locus-specific *D* values between populations, even for those in close geographic proximity, suggesting that patterns of selection on specific toxin loci are not consistent across populations (Figure 3b). Overall, these results are consistent with drift having the primary role in shaping patterns of contemporary adaptive variation as represented by toxin loci within populations.

3.4 | Selection among *S. catenatus* populations

To assess diversifying selection among populations in toxin loci, we generated F_{ST} values for 189 nonsynonymous SNPs from 38 toxin loci. We did the same for 6060 SNPs from 1209 control loci to generate a distribution of F_{ST} values attributable to demographic effects only. SNPs weighted within control loci reflected a mean F_{ST} = 0.27 with an upper bound 95th-percentile of 0.52 (Figure 4a), while individual control SNPs reflected a mean F_{ST} = 0.27 with an upper bound 95th-percentile of 0.58 (Figure 4b). If we define toxin loci and toxin nonsynonymous SNPs as under diversifying selection when they have F_{ST} values greater than these percentile values, then we detected diversifying selection in three (i.e., ~8%) of the toxin loci and 17 (i.e., ~9%) of the toxin nonsynonymous SNPs (Figure 4a and b).

The three single toxin loci identified as under selection encode snake venom metaloproteinases (*SVMP-o*), serine proteinases (*SVSP-m*), and vascular endothelial growth factor toxin proteins (*VEGF-a*) (Figure 4a). These loci are single representatives of large multigene families consisting of 12 *SVMP*, 14 *SVSP*, and two *VEGF* paralogous copies, respectively (Table S1), which are abundant in the whole venom of *S. catenatus* (Sanz et al., 2006).

The toxin nonsynonymous SNPs identified as being under selection come from the same multigene families (Figure 4b). We detected 11 SNPs from five *SVMP* loci (*SVMP-c*, *SVMP-d*, *SVMP-i*, *SVMP-n*, and *SVMP-o*), three SNPs from two *SVSP* loci (*SVSP-d and SVSP-m*), and three SNPs from the same *VEGF* locus (*VEGF-a*) (Figure 4b). We classified most of these amino acid changes (~82%) as radical substitutions (Table 2), thereby increasing the chance that they have significant impacts on venom phenotype.

3.5 | Evolutionary history of selection on toxin loci

After mapping both *S. catenatus* capture-based reads and *S. tergeminus* whole-genome reads against the *S. catenatus* reference genome, we identified intraspecific (*S. catenatus* only) and/or interspecific (*S. catenatus vs. S. tergeminus*) variation in 43 toxin loci and 72 of the control unlinked exons. When data across toxin loci were analyzed together, the nonsynonymous-to-synonymous polymorphism ratio, Pn/Ps = 1.35, was significantly smaller than the nonsynonymous-to-synonymous divergence ratio, Dn/Ds = 2.18 (Fisher's exact test, P = 0.03; Figure 5). These results indicate an overall significant effect of directional selection acting on the toxin loci. In contrast, there was no significant difference in the same ratios (Pn/Ps = 0.27 *vs.* Dn/Ds = 0.26) for the control loci (Fisher's exact test, P = 1.00; Figure 5).

Locus-specific analyses conducted with SnipRE (Figure 5) revealed that four of 43 (~9%) toxin loci in *S. catenatus* (*PLA2-f*, *SVMP-l*, *SVSP-a*, and *SVSP-k*) had significantly positive γ values ($\gamma > 0$), indicating that they have been under directional selection since divergence from *S. tergeminus*. This percentage is similar to the percent of loci identified to

be under diversifying selection among populations. Inspection of the non-significant γ values for the remaining loci shows that they are under a diverse set of selective pressures. Roughly half had negative γ values ($\gamma < 0$), suggesting purifying selection, while the rest had positive γ values ($\gamma > 0$), suggesting the presence of small-effect loci under weak directional selection. In contrast, γ values across all control loci were negative ($\gamma < 0$), suggesting purifying selection on these loci.

4 | DISCUSSION

4.1 | Interpreting evidence for drift and selection in small populations

Support for an impact of drift on adaptive variation in *S. catenatus* populations comes from two results: positive relationships between π and N_e, and evidence for, at best, weak selection on toxin loci within and among populations. Inferring an impact of drift from population-level relationships of genetic variation and effective size is based on the theoretical expectation that, as N_e decreases, inbreeding and drift will lead to declines in genetic diversity (Frankham, 1996; 2005), which may lead to impacts on population viability (Gilpin & Soulé, 1986; Frankham et al., 2017). We found the expected positive relationship between both long-term and short-term estimates of N_e for three classes of variation that encompass both functional and neutral genetic variation, but this relationship was only statistically significant for two of six comparisons. Both, however, involved our surrogate measure of adaptive variation (nonsynonymous substitutions in toxin loci).

The time scale over which these patterns have developed is unclear because long-term and short-term measures of N_e are correlated with each other. Our attempts to find independent associations between either measure of N_e were unsuccessful possibly due to low statistical power given our small sample sizes. We suspect that these correlations more likely reflect the impact of drift over historical rather than contemporary timescales because *S. catenatus* populations are currently not in genetic equilibrium with respect to levels of variation predicted from short-term estimates of N_e (Sovic et al., 2019). This is consistent with the argument that evolutionary processes that affect levels of adaptive variation are more likely to occur over long rather than short timescales (Garrigan & Hedrick, 2003). Overall, we conclude that these patterns suggest a weak but detectable impact of genetic drift on adaptive variation in these populations of snakes although the timescale of this impact is unclear.

Limited selection on potentially adaptive variation can also be interpreted as evidence for a strong impact of drift. In small populations, the impact of selection on patterns of adaptive variation is less efficient due to the increasing effects of drift (Wright, 1931). This leads to the prediction that, in small populations, detectable selection on adaptive variation should be rare, reflecting a dominant role of drift in shaping patterns of adaptive variation in such populations (Lande, 1994; Kohn, Murphy, Ostrander, & Wayne, 2006).

Our analyses of the patterns of selection on toxin genes in populations of *S. catenatus* support this prediction. First, only ~1% of toxin loci showed patterns of variation consistent with directional/purifying selection within populations. Second, only ~8% of

toxin loci had levels of differentiation suggesting diversifying selection among populations leading to local adaption. As in other studies (Funk et al., 2016), the limited selection detected on toxin loci could be taken as evidence that drift is now the primary evolutionary force shaping adaptive variation in these snake populations.

However, this interpretation relies on establishing a benchmark as to what the expected levels of selection will be when the impacts of drift are not present. Ideally, this could come from an intraspecific comparison of patterns of selection between sets of populations in which drift effects are strong in some populations and weak in other large "reference" populations. However, suitable reference populations may be difficult to identify for *S. catenatus* (as for many threatened and endangered species) because range-wide anthropogenic impacts make it likely that drift has influenced variation in most populations (Szymanski et al., 2016).

As an alternative, we analyzed historical selection on venom protein coding loci between *S. catenatus* and *S. tergeminus*, and these results suggest that at least directional selection on toxin loci may be rare and/or hard to detect. Roughly half of all toxin loci surveyed had negative selection coefficients, indicating they were under purifying selection and hence, unlikely to be targets of directional selection leading to local adaptation within populations. Indeed, only four (~9%) loci showed evidence of significant directional selection. We note that the proportion of loci under directional selection is remarkably similar to the proportion of toxin loci (~8%) inferred to be under directional selection from the intraspecific analysis. Therefore, the limited directional selection observed on toxin loci among *S. catenatus* populations is not necessarily due to the impacts of drift but could

reflect a general pattern of limited directional selection across toxin loci as a whole at both intraspecific and interspecific levels. This interpretation is supported by a small number of previous studies that have shown that the type and intensity of selection on potentially functional variants differs across closely related species, and that positive selection leading to local adaptation within species is rare (Aird et al., 2015; Margres et al., 2017a; Rautsaw et al., 2019). This may also reflect increasing evidence that genetic elements that underlie regulation of expression in toxin loci are likely the main "loci of evolution" for intraspecific variation in venom (Rokyta, Lemmon, Margres, & Aronow, 2012).

Our results illustrate the challenges of disentangling the effects of selection and drift on the genetic variation underlying adaptive traits in small populations (Hoelzel, Bruford, & Fleischer, 2019). We conclude that our findings of weak correlations between π and N_e and limited selection on toxin genes provide qualitative evidence for only weak effects of genetic drift on adaptive variation in contemporary populations of *S. catenatus*. However, if populations remain at their present day sizes, the impacts of drift on levels of variation are expected to increase with models, suggesting that most populations will lose 30% or more of their current day variation (measured as levels of heterozygosity in neutral loci) over the next 50 generations (Sovic et al., 2019).

The fact these populations are not at genetic equilibrium raises the possibility that current observed levels of adaptive variation in *S. catenatus* populations may represent an example of what Gilroy, Phillips, Richardson, and van Oosterhout (2017) have termed as populations in a state of "drift debt". Under this scenario, populations that have undergone recent severe population declines are no longer in drift-mutation-selection equilibrium and,

548 therefore, present-day measures overestimate the amount of functional genetic diversity
549 that will be present in these populations in the future. Such populations may be poised to
550 enter an extinction vortex with the true genetic cost of living at their current population
551 size yet to be "paid".
552
553 **4.2 | Genetic architecture of adaptive variation in small populations**
554 Assessing the risks faced by small populations in terms of potential losses of adaptive
555 variation has traditionally focused on the impact of drift mediated by N_e alone, such as the
556 50/500 rule (Jamieson & Allendorf, 2012). Yet, the probability that a given variant will be
557 lost is a function of the strength of drift mediated through both N_e and s (Wright, 1931). As
558 such, the distribution of s for a given set of functional variants will play a key role in
559 determining the number and fitness-effect size of variants that persist in small populations
560 of threatened and endangered species.
561 Thurman and Barrett (2016) recently summarized available information on s values
562 for individual SNPs from experimental studies of selection in natural populations. Their
563 summary is based on data from few studies from a limited set of taxa, but their results offer
564 a relevant perspective on the potential fate of adaptive variation in small populations. In
565 particular, the distributions of s values they document follow an exponential distribution
566 consistent with many SNPs having small effects on fitness and fewer SNPs having large
567 effects in terms of directional selection. This pattern is consistent with the distribution of
568 variant specific fitness effects under the geometric model of adaptation proposed by Orr
569 (1998). Moreover, the absolute values of s for SNPs are sufficiently large, suggesting that

570 many of the SNPs that underlie adaptive variation will persist despite the strong effects of
571 drift in populations with a very small N_e. For example, if we apply the criterion of $s >$
572 $1/(2N_e)$ to identify variants that will persist in the face of drift, then at $N_e = 10$, variants with
573 s values > 0.05 will persist. Based on the summary of distributions of s values for directional
574 selection (see Figure 1 of Thurman & Barrett, 2016), this represents ~ 70% of the variants
575 detected to be under directional selection. Our application of a simplistic metric is
576 unrealistic, but our more general point is that the observed intensity of selection on
577 putatively adaptive variants in natural populations suggests that many would persist
578 despite the level of drift found even in very small populations.
579
580 **4.3 | Conservation implications**
581 These results have several implications for conservation. First, combined with results from
582 Sovic et al. (2019), they suggest that despite having a small contemporary N_e, populations of
583 *S. catenatus* may only be in the early stages of the extinction vortex process (Gilpin & Soulé,
584 1986). This emphasizes the need for genetic monitoring of individual populations to track
585 levels of adaptive genetic variation over time and to assess if genetic costs through
586 inbreeding depression affect population viability (Kardos, Taylor, Ellegren, Luikart, &
587 Allendorf, 2016). If this occurs, then one potential management strategy is assisted
588 migration of individual snakes between populations (Bell et al., 2019), where genetic
589 and/or phenotypic variation related to venom proteins could be used to guide the choice of
590 donor animals for transfers between populations.

Second, our focus on assessing adaptive variation in genes that underlie a single ecologically important trait runs the risk of experiencing the undesirable outcomes of "gene-centered" conservation (Kardos & Shafer, 2018; Pearse, 2016), such as the loss of genome-wide genetic variation in loci other than the targeted ones. We stress the importance of additional genome-wide analysis of putatively adaptive genetic variants in populations of these snakes to determine the generality of the results presented in this study.

Finally, the results of Thurman and Barrett (2016) suggest that even in cases where N_e is small, threatened and endangered species could still retain a significant fraction of large-effect loci to fitness; this is consistent with observations from several species that have undergone severe bottlenecks (Aguilar et al., 2004; Benazzo et al., 2017; Grossen et al., 2020). Thus, rather than being exceptional, these results may be more common than previously suspected. They also imply that the functional variants that persist will be biased towards large-effect variants that, in turn, may limit the ability of species to track future environmental change due to the rate at which adaptation can occur (Kardos & Luikart, 2019). Both findings argue that the genetic architecture of fitness-related variation needs to be incorporated into evaluations of the genetic risks faced by small populations of threatened and endangered species if we are to have an accurate and realistic accounting of the genetic risks facing such populations (Funk et al., 2019; Mable, 2019).

Acknowledgements:

We thank all those individuals who have provided samples or assisted with collections across the range of *S. catenatus* over the past 25 years—this work would not be possible without their help. Many individuals helped, but we especially thank Jeff Davis, Michael Dreslik, Brian Fedorko, Tony Frazier, Kim Frolich, Dan Harvey, Matt Kowalski, Greg Lipps, Chris Parent, Chris Phillips, Paul Pratt, Kent Prior, Kevin Shoemaker, Michelle Villenueve, and Doug Wynn for their enthusiastic and generous assistance with finding snakes and/or providing samples. We also thank members of the Gibbs Lab for discussion, Alyssa Hassinger and Jose Diaz for technical assistance, and Kate Parsons and Carolyn Caldwell for their long-time support of our conservation genetic work on endangered snakes. J. Andrew DeWoody and Robert R. Fitak provided valuable and detailed comments on early drafts of the manuscript that significantly shaped its content. Additional comments and/or discussion was provided by Anna Brüeniche-Olsen, Avril Harder, Andrew Mularo, Samarth Mathur, Rian Bylsma, John Bickham, and Matt Hahn. Computational analyses were performed on the Ohio Biodiversity Conservation Partnership Computing Cluster and using resources provided by the Ohio Supercomputer Center. This work was primarily supported by the State Wildlife Grants Program, administered jointly by the U.S. Fish and Wildlife Service and the Ohio Division of Wildlife, with funds provided to HLG by the Ohio Biodiversity Conservation Partnership between The Ohio State University and the Ohio Division of Wildlife. HLG was also supported by National Science Foundation (USA) Grant DEB 11633872 during the preparation of the manuscript. Finally, DDR was supported by National Science Foundation (USA) Grant DEB 1145987.

Data Availability Statement:

Raw sequencing reads have been deposited in the NCBI SRA as BioSamples SAMN15063710–SAMN15063802 under BioProject PRJNA636095.

Author Contributions:

Designed research: AO and HLG

Performed research: AO, MB, and HLG

Contributed reagents: ARL, EML, and DRR

Analyzed data: AO and MB

Wrote the paper: AO and HLG

References

Aguilar, A., Roemer, G., Debenham, S., Binns, M., Garcelon, D., & Wayne, R. K. (2004). High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences*, *101*(10), 3490–3494. doi:10.1073/pnas.0306582101

Aird, S. D., Aggarwal, S., Villar-Briones, A., Tin, M. M.-Y., Terada, K., & Mikheyev, A. S. (2015). Snake venoms are integrated systems, but abundant venom proteins evolve more rapidly. *BMC Genomics*, *16*, 647. doi:10.1186/s12864-015-1832-6

656 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment
657 search tool. *Journal of Molecular Biology*, *215*(3), 403–410. doi:10.1016/S0022-
658 2836(05)80360-2
659 Babik, W. (2010). Methods for MHC genotyping in non-model vertebrates. *Molecular*
660 *Ecology Resources*, *10*(2), 237–251. doi:10.1111/j.1755-0998.2009.02788.x.
661 Barrett, R. D. H., & Hoekstra, H. E. (2011). Molecular spandrels: Tests of adaptation at the
662 genetic level. *Nature Reviews Genetics*, *12*(11), 767–780. doi:10.1038/nrg3015
663 Barton, N., & Partridge, L. (2000). Limits to natural selection. *Bioessays*, *22*(12), 1075–1084.
664 doi:10.1002/1521-1878(200012)22:12<1075::AID-BIES5>3.0.CO;2-M
665 Bell, D. A., Robinson, Z. L., Funk, W. C., Fitzpatrick, S. W., Allendorf, F. W., Tallmon, D. A., &
666 Whiteley, A. R. (2019). The exciting potential and remaining uncertainties of genetic
667 rescue. *Trends in Ecology & Evolution*, *S0169–5347*(19), 30173–30179.
668 doi:10.1016/j.tree.2019.06.006
669 Benazzo, A., Trucchi, E., Cahill, J. A., Delser, P. M., Mona, S., Fumagalli, M., … Bertorelle, G.
670 (2017). Survival and divergence in a small group: The extraordinary genomic history
671 of the endangered Apennine brown bear stragglers. *Proceedings of the National*
672 *Academy of Sciences*, *114*(45), E9589–E9597. doi:10.1073/pnas.1707279114
673 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and
674 powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*
675 *(Methodological)*, *57*(1): 289–300. doi:10.2307/2346101

676 Blomqvist, D., Pauliny, A., Larsson, M., & Flodin, L.-Å. (2010). Trapped in the extinction
677 vortex? Strong genetic effects in a declining vertebrate population. *BMC Evolutionary*
678 *Biology*, *10*, 33. doi:10.1186/1471-2148-10-33
679 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina
680 sequence data. *Bioinformatics*, *30*(15), 2114–2120.
681 doi:10.1093/bioinformatics/btu170
682 Campagna, L., Repenning, M., Silveira, L. F., Fontana, C. S., Tubaro, P. L., & Lovette, I. J.
683 (2017). Repeated divergent selection on pigmentation genes in a rapid finch
684 radiation. *Science Advances*, *3*(5), e1602404. doi:10.1126/sciadv.1602404
685 Casewell, N. R., Harrison, R. A., Wüster, W., & Wagstaff, S. C. (2009). Comparative venom
686 gland transcriptome surveys of the saw-scaled vipers (Viperidae: *Echis*) reveal
687 substantial intra-family gene diversity and novel venom transcripts. *BMC Genomics*,
688 *10*, 564. doi:10.1186/1471-2164-10-564
689 Casewell, N. R., Wagstaff, S. C., Harrison, R. A., Renjifo, C., & Wüster, W. (2011). Domain loss
690 facilitates accelerated evolution and neofunctionalization of duplicate snake venom
691 metalloproteinase toxin genes. *Molecular Biology and Evolution*, *28*(9): 2637–2649.
692 doi:10.1093/molbev/msr091
693 Casewell, N. R., Wüster, W., Vonk, F. J., Harrison, R. A., & Fry, B. G. (2013). Complex cocktails:
694 The evolutionary novelty of venoms. *Trends in Ecology & Evolution*, *28*(4): 219–229.
695 doi:10.1016/j.tree.2012.10.020

696 Chippaux, J. P., Williams, V., & White, J. (1991). Snake venom variability: Methods of study,
697 results and interpretation. *Toxicon*, *29*(11), 1279–1303. doi:10.1016/0041-
698 0101(91)90116-9
699 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes
700 Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*,
701 *27*(15), 2156–2158. doi:10.1093/bioinformatics/btr330
702 Eilertson, K. E., Booth, J. G., & Bustamante, C. D. (2012). SnIPRE: Selection inference using a
703 Poisson random effects model. *PLoS Computational Biology*, *8*(12), e1002806.
704 doi:10.1371/journal.pcbi.1002806
705 Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to
706 perform population genetics analyses under Linux and Windows. *Molecular Ecology*
707 *Resources*, *10*(3), 564–567. doi:10.1111/j.1755-0998.2010.02847.x
708 Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa V. C., & Foll M. (2013). Robust
709 demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905.
710 doi:10.1371/journal.pgen.1003905
711 Fijarczyk, A., Dudek, K., & Babik, W. (2016). Selective landscapes in newt immune genes
712 inferred from patterns of nucleotide variation. *Genome Biology and Evolution*, *8*(11),
713 3417–3432. doi:10.1093/gbe/evw236
714 Frankham, R. (1996). Relationship of genetic variation to population size in wildlife.
715 *Conservation Biology*, *10*(6), 1500–1508. doi:10.1046/j.1523-1739.1996.10061500.x
716 Frankham, R. (2005). Genetics and extinction. Biological Conservation, *126*(2), 131–140.
717 doi:10.1016/j.biocon.2005.05.002

718 Frankham, R., Ballou, J. D., Ralls, K., Eldridge, M., Dudash, M. R., Fenster, C. B., … Sunnucks, P.
719 (2017). *Genetic management of fragmented animal and plant populations*. Oxford, UK:
720 Oxford University Press.
721 Funk, W. C., Forester, B. R., Converse, S. J., Darst, C., & Morey, S. (2019). Improving
722 conservation policy with genomics: A guide to integrating adaptive potential into
723 U.S. Endangered Species Act decisions for conservation practitioners and geneticists.
724 *Conservation Genetics*, *20*(1), 115–134. doi:10.1007/s10592-018-1096-1
725 Funk, W. C., Lovich, R. E., Hohenlohe, P. A., Hofman, C. A., Morrison, S. A., Sillett, T. S., …
726 Andelt, W. F. (2016). Adaptive divergence despite strong genetic drift: Genomic
727 analysis of the evolutionary mechanisms causing genetic differentiation in the island
728 fox (*Urocyon littoralis*). *Molecular Ecology*, *25*(10), 2176–2194.
729 doi:10.1111/mec.13605
730 Garrigan, D., & Hedrick, P. W. (2003). Detecting adaptive molecular polymorphism: lessons
731 from the MHC. *Evolution*, *57*(8), 1707–1722. doi:10.1111/j.0014-
732 3820.2003.tb00580.x
733 Gibbs, H. L., & Chiucchi, J. E. (2011). Deconstructing a complex molecular phenotype:
734 population-level variation in individual venom proteins in Eastern Massasauga
735 Rattlesnakes (*Sistrurus c. catenatus*). *Journal of Molecular Evolution*, *72*(4), 383–397.
736 doi:10.1007/s00239-011-9437-4. Epub 2011 Mar 11.
737 Gibbs, H. L., & Chiucchi, J. E. (2012). Inbreeding, body condition, and heterozygosity-fitness
738 correlations in isolated populations of the endangered eastern massasauga

739 rattlesnake (*Sistrurus c. catenatus*). *Conservation Genetics*, *13*(4), 1133–1143.
740 doi:10.1007/s10592-012-0360-z
741 Gibbs, H. L., Sanz, L., & Calvete, J. J. (2009). Snake population venomics: Proteomics-based
742 analyses of individual variation reveals significant gene regulation effects on venom
743 protein expression in *Sistrurus* rattlesnakes. *Journal of Molecular Evolution*, *68*(2),
744 113–125. doi:10.1007/s00239-008-9186-1
745 Gilpin, M. E., & Soulé, M. E. (1986). Minimum viable populations: Processes of species
746 extinction. In M. E. Soulé (Ed.), *Conservation biology: The science of scarcity and*
747 *diversity* (pp. 19–34). Sunderland, MA: Sinauer Associates, Inc.
748 Gilroy, D. L., Phillips, K. P., Richardson, D. S., & van Oosterhout, C. (2017). Toll-like receptor
749 variation in the bottlenecked population of the Seychelles warbler: Computer
750 simulations see the 'ghost of selection past' and quantify the 'drift debt'. *Journal of*
751 *Evolutionary Biology*, *30*, 1276–1287. doi:10.1111/jeb.13077
752 Government of Canada (2009). Species at risk public registry. Retrieved from
753 https://www.sararegistry.gc.ca
754 Grossen, C., Guillaume, F., Keller, L. K., & Croll, D. (2020). Purging of highly deleterious
755 mutations through severe bottlenecks in Alpine ibex. *Nature Communications*, *11*(1),
756 1001. doi:10.1038/s41467-020-14803-1
757 Hare, M. P., Nunney, L., Schwartz, M. K., Ruzzante, D. E., Burford, M., Waples, R. S., … Palstra,
758 F. (2011). Understanding and estimating effective population size for practical
759 application in marine species management. *Conservation Biology*, *25*(3), 438–449.
760 doi:10.1111/j.1523-1739.2010.01637.x.

761 Hoelzel, A. R., Bruford, M. W., & Fleischer, R. C. (2019). Conservation of adaptive potential
762 and functional diversity. *Conservation Genetics*, *20*(1), 1–5. doi:10.1007/s1059
763 Jaeger, C. P., Duvall, M. R., Swanson, B. J., Phillips, C. A., Dreslik, M. J., Baker, S. J., & King, R. B.
764 (2016). Microsatellite and major histocompatibility complex variation in an
765 endangered rattlesnake, the Eastern Massasauga (*Sistrurus catenatus*). *Ecology and*
766 *Evolution*, *6*(12), 3991–4003. doi:10.1002/ece3.2159
767 Jamieson, I. G., & Allendorf, F. W. (2012). How does the 50/500 rule apply to MVPs? *Trends*
768 *in Ecology & Evolution*, *27*(10), 578–584. doi:10.1016/j.tree.2012.07.001
769 Jones, M. R., Mills, L. S., Alves, P. C., Callahan, C. M., Alves, J. M., Lafferty, D. J. R. … Good, J. M.
770 (2018). Adaptive introgression underlies polymorphic seasonal camouflage in
771 snowshoe hares. *Science*, *360*(6395), 1355–1358. doi:10.1126/science.aar5273
772 Kardos, M., & Luikart, G. (2019). The genetic architecture of fitness drives population
773 viability during rapid environmental change. *bioRxiv*. doi:10.1101/660803
774 Kardos, M., & Shafer, A. B. A. (2018). The peril of gene-targeted conservation. *Trends in*
775 *Ecology & Evolution*, *33*(11), 827–839. doi:10.1016/j.tree.2018.08.011
776 Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., & Allendorf, F. W. (2016). Genomics
777 advances the study of inbreeding depression in the wild. *Evolutionary Applications*,
778 *9*(10), 1205–1218. doi:10.1111/eva.12414
779 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A.
780 (2012). Geneious Basic: An integrated and extendable desktop software platform for
781 the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649.
782 doi:10.1093/bioinformatics/bts199

783 Kohn, M. H., Murphy, W. J., Ostrander, E. A., & Wayne, R. K. (2006). Genomics and
784 conservation genetics. *Trends in Ecology & Evolution*, *21*(11), 629–637.
785 doi:10.1016/j.tree.2006.08.001
786 Lande, R. (1994). Risk of population extinction from fixation of new deleterious mutations.
787 *Evolution*, *48*(5), 1460. doi:10.2307/2410240
788 Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for
789 massively high-throughput phylogenomics. *Systematic Biology*, *61*(5), 727–744.
790 doi:10.1093/sysbio/sys049
791 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
792 transform. *Bioinformatics*, *25*(14), 1754–1760. doi:10.1093/bioinformatics/btp324
793 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The
794 Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.
795 doi:10.1093/bioinformatics/btp352
796 Liu, Y., Schröder, J., & Schmidt, B. (2013). Musket: A multistage *k*-mer spectrum-based error
797 corrector for Illumina sequence data. *Bioinformatics*, *29*(3), 308–315.
798 doi:10.1093/bioinformatics/bts690
799 Lynch, V. J. (2007). Inventing an arsenal: Adaptive evolution and neofunctionalization of
800 snake venom phospholipase A_2 genes. *BMC Evolutionary Biology*, *7*, 2.
801 doi:10.1186/1471-2148-7-2
802 Mable, B. K. (2019). Conservation of adaptive potential and functional diversity: Integrating
803 old and new approaches. *Conservation Genetics*, *20*(1), 89–100. doi:10.1007/s10592-
804 018-1129-9

Margres, M. J., Bigelow, A. T., Lemmon, E. M., Lemmon, A. R., & Rokyta, D. R. (2017a). Selection to increase expression, not sequence diversity, precedes gene family origin and expansion in rattlesnake venom. *Genetics*, *206*(3), 1569–1580. doi:10.1534/genetics.117.202655

Margres, M. J., Wray, K. P., Hassinger, A. T. B., Ward, M. J., McGivern, J. J., Lemmon, E. M., … Rokyta, D. R. (2017b). Quantity, not quality: Rapid adaptation in a polygenic trait proceeded exclusively through expression differentiation. *Molecular Biology and Evolution*, *34*(12), 3099–3110. doi:10.1093/molbev/msx231

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, *351*(6328), 652–654. doi:10.1038/351652a0

McKenna, A., Hanna M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 254–260. doi:10.1101/gr.107524.110.20

McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources*, *17*(4), 656–669. doi:10.1111/1755-0998

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York, NY: Columbia University Press.

Norén, K., Godoy, E., Dalén, L., Meijer, T., & Angerbjörn, A. (2016). Inbreeding depression in a critically endangered carnivore. *Molecular Ecology*, *25*(14), 3309–3318. doi:10.1111/mec.13674

827 Orr, H. A. (1998). The population genetics of adaptation: The distribution of factors fixed
828 during adaptive evolution. *Evolution*, *52*(4), 935–949. doi:10.1111/j.1558-
829 5646.1998.tb01823.x
830 Pearse, D. (2016). Saving the spandrels? Adaptive genomic variation in conservation and
831 fisheries management. *Journal of Fish Biology*, *89*(6), 2697–2716.
832 doi:10.1111/jfb.13168
833 Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., & Lemmon,
834 A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-
835 generation DNA sequencing. *Nature*, *526*(7574), 569–573. doi:10.1038/nature15697
836 Rautsaw, R. M., Hofmann, E. P., Margres, M. J., Holding, M. L., Strickland, J. L., Mason, A. J., …
837 Parkinson, C. L. (2019). Intraspecific sequence and gene expression variation
838 contribute little to venom diversity in sidewinder rattlesnakes (*Crotalus cerastes*).
839 *Proceedings of the Royal Society B*, *286*(1906), 20190810.
840 doi:10.1098/rspb.2019.0810
841 Robinson, J. A., Ortega-Del Vecchyo, D., Fan, Z., Kim, B. Y., vonHoldt, B. M., Marsden, C. D., …
842 Wayne, R. K. (2016). Genomic flatlining in the endangered island fox. *Current Biology*,
843 *26*(9), 1183–1189. doi:10.1016/j.cub.2016.02.062
844 Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: All
845 that's gold does not glitter. *Evolution*, *66*(1), 1–17. doi:10.1111/j.1558-
846 5646.2011.01486.x

847 Rokyta, D. R., Lemmon, A. R., Margres, M. J., & Aronow, K. (2012). The venom-gland
848 transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC*
849 *Genomics*, *13*, 312. doi:10.1186/1471-2164-13-312
850 Rokyta, D. R., Wray, K. P., & Margres, M. J. (2013). The genesis of an exceptionally lethal
851 venom in the timber rattlesnake (*Crotalus horridus*) revealed through comparative
852 venom-gland transcriptomics. *BMC Genomics*, *14*, 394. doi:10.1186/1471-2164-14-
853 394
854 Ruane, S., Raxworthy, C. J., Lemmon, A. R., Lemmon, E. M., & Burbrink, F. T. (2015).
855 Comparing species tree estimation with large anchored phylogenomic and small
856 Sanger-sequenced molecular datasets: An empirical study on Malagasy
857 pseudoxyrhophiine snakes. *BMC Evolutionary Biology*, *15*, 221. doi:10.1186/s12862-
858 015-0503-1
859 Smiley-Walters, S. A., Farrell, T. M., & Gibbs, H. L. (2017). Evaluating local adaptation of a
860 complex phenotype: Reciprocal tests of pigmy rattlesnake venoms on treefrog prey.
861 *Oecologia*, *184*(4), 739–748. doi:10.1007/s00442-017-3882-8
862 Smiley-Walters, S. A., Farrell, T. M., & Gibbs, H. L. (2019). High levels of functional
863 divergence in toxicity towards prey among the venoms of individual pigmy
864 rattlesnakes. *Biology Letters*, *15*(2), 20180876. doi:10.1098/rsbl.2018.0876
865 Sovic, M. G., Fries, A. C., & Gibbs, H. L. (2016). Origin of a cryptic lineage in a threatened
866 reptile through isolation and historical hybridization. *Heredity*, *117*(5), 358–366.
867 doi:10.1038/hdy.2016.56

868 Sovic, M., Fries, A., Martin, S. A., & Gibbs, H. L. (2019). Genetic signatures of small effective
869 population sizes and demographic declines in an endangered rattlesnake, *Sistrurus*
870 *catenatus*. *Evolutionary Applications*, *12*(4), 664–678. doi:10.1111/eva.12731
871 Szymanski, J., Pollack, C., Ragan, L., Redmer, M., Clemency, L., Voorhies, K., & JaKa, J. (2016).
872 *Species status assessment for the eastern massasauga rattlesnake (Sistrurus*
873 *catenatus)*. U.S. Fish and Wildlife Service, Washington D.C.
874 Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA
875 polymorphism. *Genetics*, *123*(3), 585–595.
876 Thurman, T. J., & Barrett, R. D. (2016). The genetic consequences of selection in natural
877 populations. *Molecular Ecology*, *25*(7), 1429–1448. doi:10.1111/mec.13559
878 U.S. Fish and Wildlife Service (2016). Endangered and threatened wildlife and plants;
879 threatened species status for the eastern massasauga rattlesnake. *Federal Register*,
880 *81*(190), 67193–67214.
881 Waples, R. S., & Do, C. (2008). LDNE: A program for estimating effective population size
882 from data on linkage disequilibrium. *Molecular Ecology Resources*, *8*(4), 753–756.
883 doi:10.1111/j.1755-0998.2007.02061.x
884 Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97–159.
885 Zhang, J. (2000). Rates of conservative and radical nonsynonymous nucleotide substitutions
886 in mammalian nuclear genes. *Journal of Molecular Evolution*, *50*(1), 56–68.
887 doi:10.1007/s002399910007

888 Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The
889 MaSuRCA genome assembler. *Bioinformatics*, *29*(21), 2669–2677.
890 doi:10.1093/bioinformatics/btt476

891

892

893

894

895

896

897

898

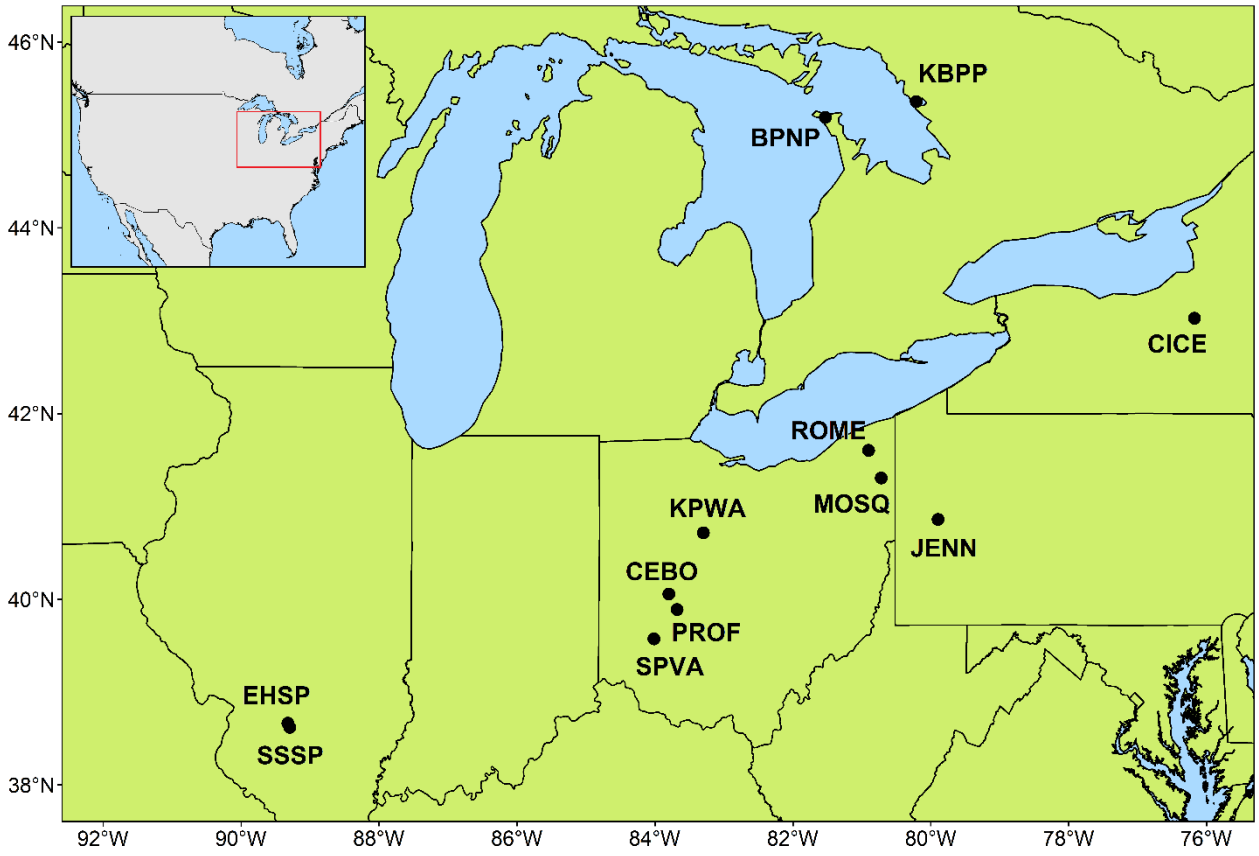


FIGURE 1 Distribution of sampled *S. catenatus* populations in the U.S. and Canada. Names of sampled populations and numbers of individuals sequenced are as follows: Eldon Hazlet State Park (EHSP, *n* = 8), South Shore State Park (SSSP, *n* = 4), Spring Valley (SPVA, *n* = 7), Cedar Bog Nature Preserve (CEBO, *n* = 4), Prairie Road Fen (PROF, *n* = 9), Killdeer Plains Wildlife Area (KPWA, *n* = 10), Mosquito Creek (MOSQ, *n* = 8), Rome (ROME, *n* = 7), Jennings (JENN, *n* = 7), Bruce Peninsula National Park (BPNP, *n* = 10), Killbear Provincial Park (KBPP, *n* = 10), and Cicero (CICE, *n* = 9).

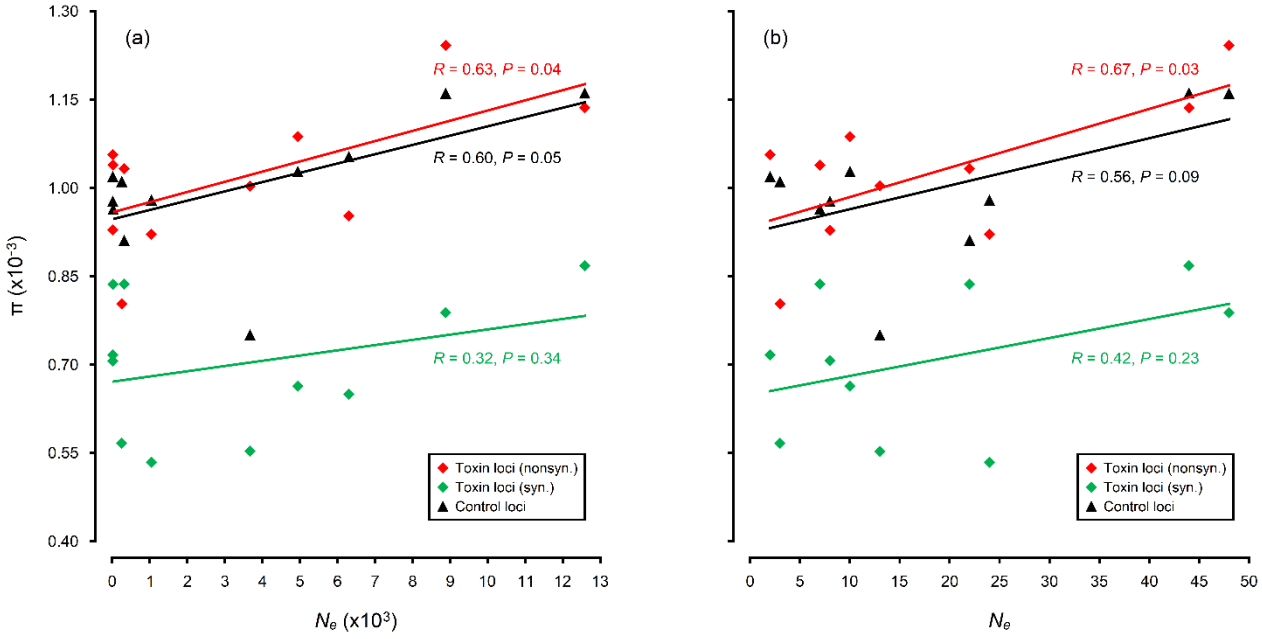
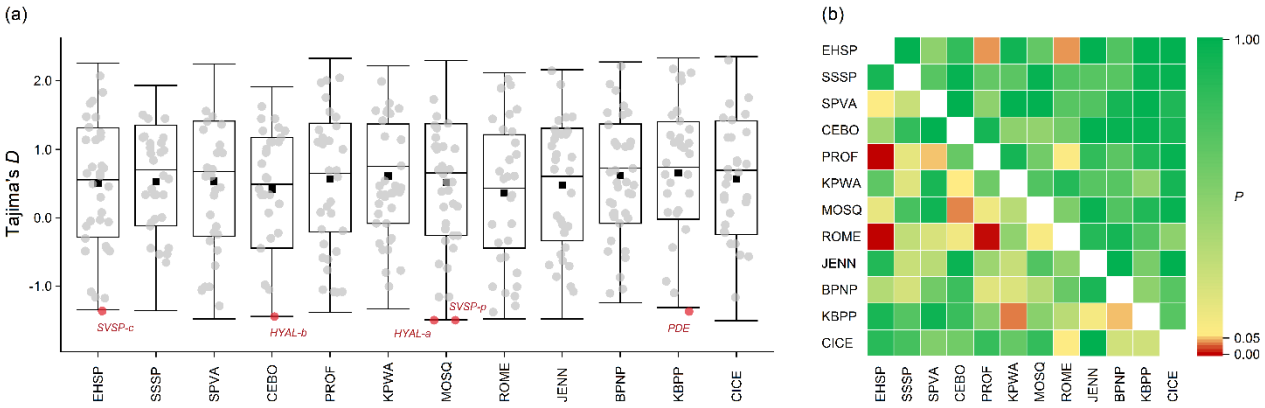


FIGURE 2 Correlations between three classes (toxin nonsynonymous SNPs, in *red*; toxin synonymous SNPs, in *green*; and control SNPs, in *black*) of nucleotide diversity, expressed as π (x10^{-3}), and (a) long-term and (b) short-term effective size (N_e) estimates for *S. catenatus* populations. Each point on the plot reflects the mean value of π across loci for that class of variation. Coefficient of correlation, *R*, and significance, *P*, values are indicated for each π class.



924

925 **FIGURE 3** Tajima's *D* values across *S. catenatus* populations for toxin and control loci. (a)
926 Box plots represent median, 1st quartile, 3rd quartile, and 95% confidence interval
927 distributions for the control loci; black squares represent mean control locus *D* estimates.
928 Circles represent *D* values for each toxin locus; red circles indicate toxin loci under
929 significant directional/purifying selection. (b) *P* values derived from Spearman rank
930 correlation tests for *D* values associated to toxin loci for each population pair. The lower
931 triangle shows "raw" *P* values; the upper triangle indicates *P* values adjusted for multiple
932 tests using a false discovery rate correction.

933

934

935

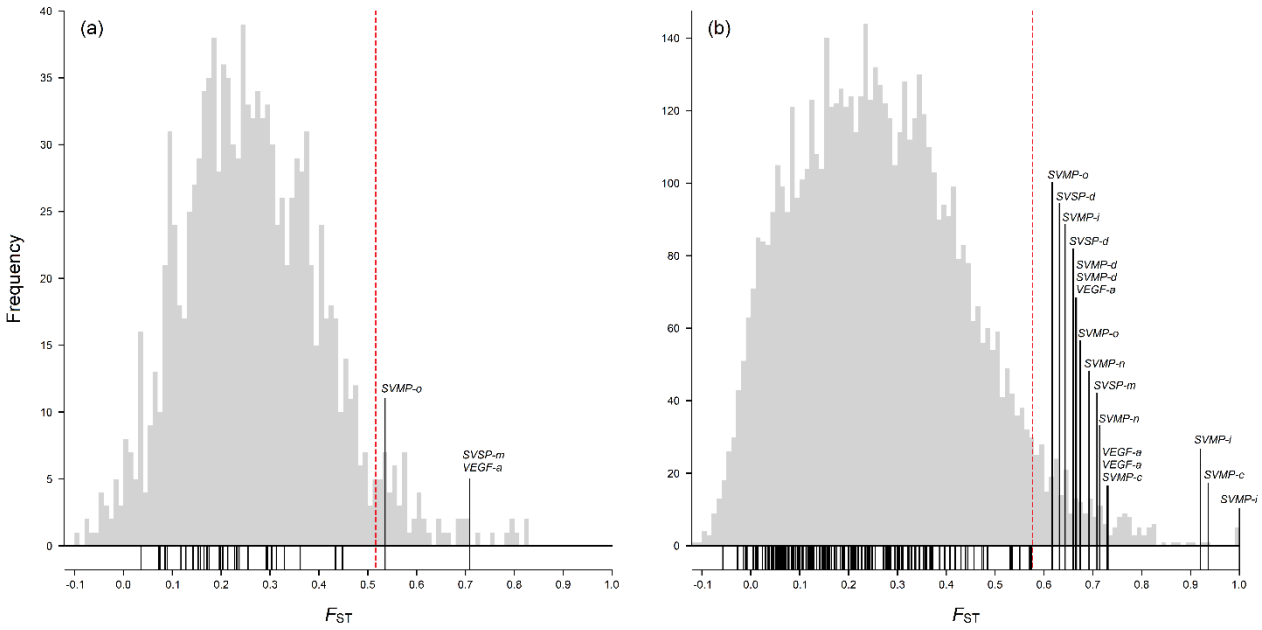
936

937

938

939

940



941

942 **FIGURE 4** F_{ST} frequency histograms among *S. catenatus* populations for (a) loci and (b)
943 SNPs. Light gray bars represent control locus F_{ST} distributions; red dashed lines indicate the
944 95th-percentile F_{ST} value for each control locus distribution. Black vertical lines at the
945 bottom of each plot denote individual toxin F_{ST} estimates (loci or individual SNPs). Names
946 of specific outlier toxin loci and SNPs are labeled (also see Table 2).

947

948

949

950

951

952

953

954

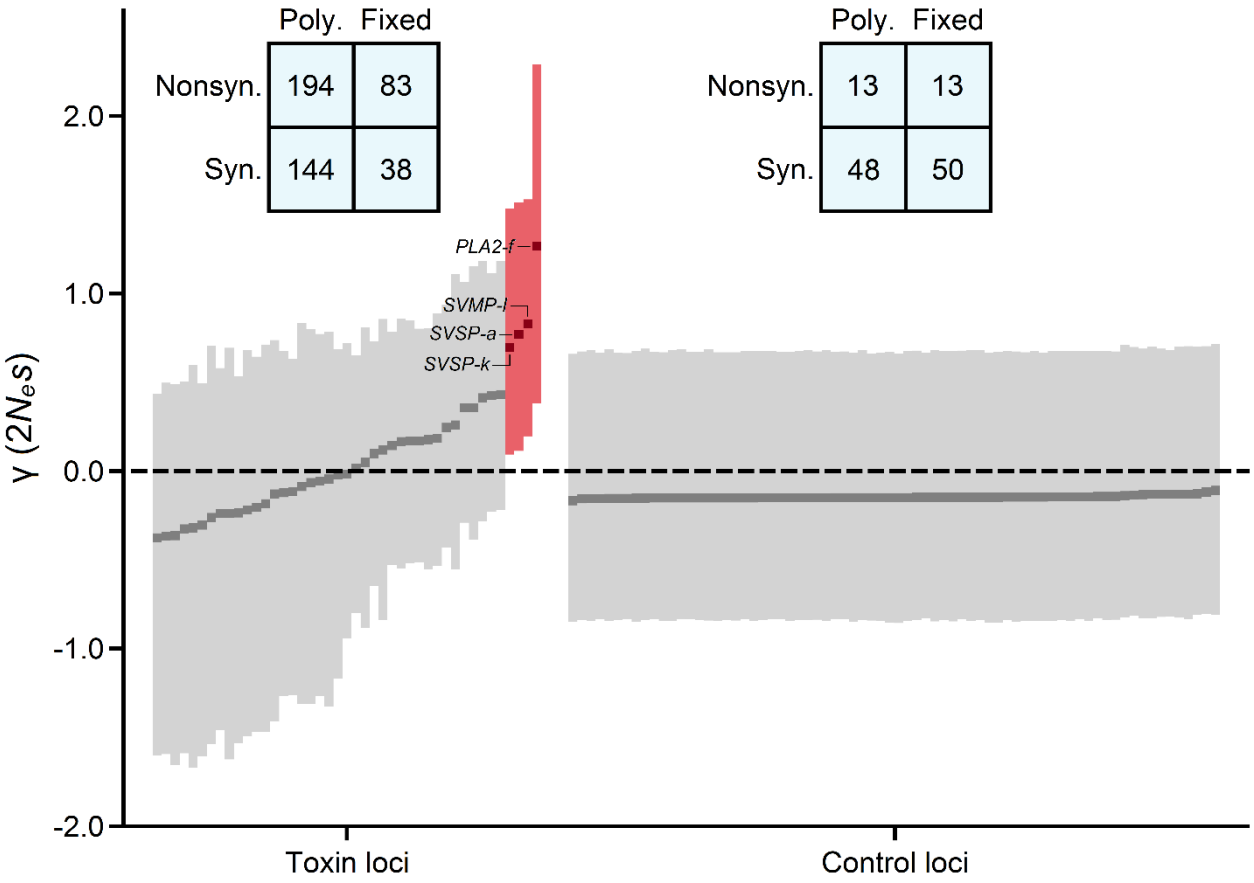


FIGURE 5 Results of SnIPRE analyses for selection on toxin and control loci in *S. catenatus* after using *S. tergeminus* as an outgroup. Overall counts of nonsynonymous and synonymous polymorphisms (*S. catenatus* only) and fixed differences (*S. catenatus vs. S. tergeminus*) for each locus set are shown as contingency tables. Bayesian selection effect ($\gamma = 2N_es$) values for individual toxin and control loci are indicated as squares within bars representing 95% confidence intervals. The black dashed line reflects values expected under neutral selection. Names of loci under significant directional selection (dark red squares within red bars) are also shown.

966 **TABLE 1** Mean nucleotide diversity estimates, expressed as π (x10^{-3}), across loci for each *S.*
967 *catenatus* population for three classes of variation: toxin nonsynonymous SNPs, toxin
968 synonymous SNPs, and control SNPs.

Population	Toxin loci (nonsyn.)	Toxin loci (syn.)	Control loci
EHSP	0.95	0.65	1.05
SSSP	1.03	0.84	0.91
SPVA	0.93	0.71	0.98
CEBO	0.86	0.56	0.92
PROF	0.92	0.53	0.98
KPWA	1.14	0.87	1.16
MOSQ	1.04	0.84	0.96
ROME	0.80	0.57	1.01
JENN	1.06	0.72	1.02
BPNP	1.24	0.79	1.16
KBPP	1.09	0.66	1.03
CICE	1.00	0.55	0.75
Mean	1.00	0.69	0.99
SE	0.04	0.04	0.03

969
970
971
972
973
974
975
976
977
978
979

TABLE 2 Properties of nonsynonymous substitutions in toxin loci identified to be under diversifying selection (see Figure 4b). Amino acid substitutions in the central column are coded as the ancestral amino acid, position in the locus, and derived amino acid. Amino acid changes were judged as radical (R) or conservative (C) based on one or more of the following features: charge alone, polarity alone, and polarity and volume combined.

Locus	Amino acid change Radical (R) or Conservative (C)	F_{ST}
SVMP-i	G272A (C)	1.00
SVMP-c	T40I (R)	0.94
SVMP-i	S270A (R)	0.92
VEGF-a	K103N (R)	0.73
VEGF-a	E114K (R)	0.73
SVMP-c	A12S (R)	0.73
SVMP-n	E427Q (R)	0.71
SVSP-m	K114R (R)	0.71
SVMP-n	P238S (R)	0.69
SVMP-o	Q37R (R)	0.67
SVMP-d	R412S (R)	0.67
SVMP-d	Y417D (R)	0.67
VEGF-a	A56S (R)	0.66
SVSP-d	V140D (R)	0.66
SVMP-I	L516F (R)	0.64
SVSP-d	R37H (C)	0.63
SVMP-o	I4V (C)	0.62