

Genomic signatures of inbreeding and genetic load in a threatened rattlesnake

Alexander Ochoa¹ and H. Lisle Gibbs²

¹University of Central Florida

²Ohio State University

May 15, 2021

Abstract

Theory predicts that threatened species living in small populations will experience high levels of inbreeding that will increase their negative genetic load but recent work suggests that the impact of load may be minimized by purging resulting from long term population bottlenecks. Empirical studies that examine this idea using genome-wide estimates of inbreeding and genetic load in threatened species are limited. Here we use genome resequencing data to compare levels of inbreeding, levels of genetic load and population history in threatened Eastern massasauga rattlesnakes (*Sistrurus catenatus*) which exist in small isolated populations and closely-related yet outbred Western massasauga rattlesnakes (*S. tergeminus*). In terms of inbreeding, *S. catenatus* genomes had a greater number of ROHs of varying sizes indicating sustained inbreeding through repeated bottlenecks when compared to *S. tergeminus*. At the species level, outbred *S. tergeminus* had higher genome-wide levels of genetic load in the form of greater numbers of derived deleterious mutations compared to *S. catenatus* presumably due to long-term purging of deleterious mutations in *S. catenatus*. In contrast, mutations that escaped the “drift sieve” and were polymorphic within *S. catenatus* populations were more abundant and more often found in homozygote genotypes than in *S. tergeminus* suggesting a reduced efficiency of purifying selection in smaller *S. catenatus* populations. Our results support an emerging idea that the historical demography of a threatened species has a significant impact on the type of genetic load present which impacts implementation of conservation actions such as genetic rescue.

ORIGINAL ARTICLE

Title: Genomic signatures of inbreeding and genetic load in a threatened rattlesnake

Running head: Genomic estimates of inbreeding and load

Authors: Alexander Ochoa^{1,2} | H. Lisle Gibbs¹

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

²Present address: Department of Biology, University of Central Florida, Orlando, FL, USA

Correspondence

Alexander Ochoa, Department of Biology, University of Central Florida, Orlando, FL, USA.

Email: alexander.ochoa@ucf.edu

ORCID ID: 0000-0002-3271-3684

and

H. Lisle Gibbs, 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

Theory predicts that threatened species living in small populations will experience high levels of inbreeding that will increase their negative genetic load but recent work suggests that the impact of load may be minimized by purging resulting from long term population bottlenecks. Empirical studies that examine this idea using genome-wide estimates of inbreeding and genetic load in threatened species are limited. Here we use genome resequencing data to compare levels of inbreeding, levels of genetic load and population history in threatened Eastern massasauga rattlesnakes (*Sistrurus catenatus*) which exist in small isolated populations and closely-related yet outbred Western massasauga rattlesnakes (*S. tergeminus*). In terms of inbreeding, *S. catenatus* genomes had a greater number of ROHs of varying sizes indicating sustained inbreeding through repeated bottlenecks when compared to *S. tergeminus*. At the species level, outbred *S. tergeminus* had higher genome-wide levels of genetic load in the form of greater numbers of derived deleterious mutations compared to *S. catenatus* presumably due to long-term purging of deleterious mutations in *S. catenatus*. In contrast, mutations that escaped the “drift sieve” and were polymorphic within *S. catenatus* populations were more abundant and more often found in homozygote genotypes than in *S. tergeminus* suggesting a reduced efficiency of purifying selection in smaller *S. catenatus* populations. Our results support an emerging idea that the historical demography of a threatened species has a significant impact on the type of genetic load present which impacts implementation of conservation actions such as genetic rescue.

Keywords : ROHs; inbreeding; demographic bottlenecks; genetic load; *Sistrurus catenatus*

1 | Introduction

Threatened and endangered species often exist in small isolated populations that theory predicts should suffer elevated risks of extinction due to impacts on their genetic makeup (Frankham et al., 2017; van Oosterhout, 2020). Populations with these characteristics experience increased genetic drift and inbreeding, which lead to a loss of genetic variation, reducing the efficacy of natural selection resulting in increased frequency and expression of deleterious recessive mutations termed genetic load (Charlesworth, 2009; Keller & Waller, 2002). However, the precise genetic endpoint of this process in terms of observed levels of genetic load in natural populations depends on complex interactions between the levels of inbreeding and effectiveness of purifying selection that are present (Grossen, Guillaume, Keller, & Croll, 2020; Hedrick & Garcia-Dorado, 2016).

Specifically, inbreeding results in the expression of recessive deleterious mutations creating the potential for selection to reduce the frequency of these mutations depending on the degree of dominance and the magnitude of the deleterious effects (Glémin, 2003). This process, termed genetic purging is most effective at high levels of inbreeding and so bottlenecks tend to purge highly deleterious, recessive mutations unless population sizes are extremely small (Garcia-Dorado, 2012; Kirkpatrick & Jarne, 2000). However, reduced population size also increases genetic drift and reduces the efficacy of selection which allows deleterious mutations to drift to higher frequency (Renaut & Rieseberg, 2015; Robinson et al., 2016). Thus, the type and level of genetic load present in a given population reflects a balance of the relative magnitudes of these distinct processes (inbreeding versus reduced selection due to drift) with opposite effects on mutation frequencies. As a result, assessing levels of inbreeding, effective population size, and levels of genetic load are an important goal of empirical studies that seek to assess the genetic risks impacting long term viability of threatened species (Mathur & DeWoody, 2021) and for assessing the appropriateness of specific management activities such as genetic rescue for mitigating these risks (Kyriazis, Wayne, & Lohmueller, 2020; but see Ralls, Sunnucks, Lacy, & Frankham, 2020).

In addition, van der Valk, de Manuel, Marques-Bonet, and Guschanski (2019) have recently emphasized the

previously underappreciated role of historical demography on the observed level of genetic load of deleterious mutations in contemporary populations. They show that estimates of genetic load vary widely across mammal species with small effective sizes and are only weakly correlated with genome-wide levels of inbreeding. They argue that this is because long term reductions in effective size can result in sustained purging of deleterious mutations leading to reduced genetic load in present-day populations. Such a pattern has been observed in populations of endangered species which have experienced long-term bottlenecks (Benazzo et al., 2017) suggesting the importance of incorporating historical patterns of population size in explanations for genetic load.

Whole genome resequencing offers new data and methods of analysis to evaluate inbreeding and genetic load in endangered species (Brüniche-Olsen, Kellner, Anderson, & DeWoody, 2018). For assessing inbreeding, individual genome sequences provide information on runs of homozygosity (ROHs) which are genomic regions that contains significantly less nucleotide variation than expected based on the genome-wide average for an individual (Ceballos, Joshi, Clark, Ramsay, & Wilson, 2018). These ROH regions arise in a genome due to the transmission of genome segments that are identical by descent (IBD) from parents to offspring (Ceballos et al., 2018). ROH segments can be used to quantify inbreeding reflected by genomic autozygosity because ROH distributions of regions that are IBD often match pedigree histories (Kardos, Qvarnström, & Ellegren, 2017). Distributions of different size ROHs also provide insights into recent demography because under random mating the length of ROH regions is expected to decrease with increasing number of generations to the most recent common ancestor due to recombination and de novo mutations (Bosse et al., 2012). Therefore, the number and length of ROH reflect the timing and intensity of a population bottleneck, with longer ROH originating from recent inbreeding, and shorter ROH from ancestral bottlenecks (Ceballos et al., 2018).

For assessing genetic load, new analytical methods (Adzhubei et al., 2010; Choi, Sims, Murphy, Miller, & Chan, 2012) now permit strong inference of the mutational effects of substitutions in coding sequences assumed to be under strong purifying selection. As such data from these approaches can be used to estimate levels of accumulation of putatively deleterious mutations at the population or species level in non-model species and explore the process that drive variation in genetic load in natural populations (Allendorf, 2017; Kardos, Taylor, Ellegren, Luikart, & Allendorf, 2016).

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 United States Endangered Species Act (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 high levels of population genetic structure (Chiocchi & Gibbs, 2010; Sovic, Fries, Martin, & Gibbs, 2019). Thus, the relevant management units within this species are individual populations.

The negative genetic impacts drift on population viability are potentially a significant conservation issue for *S. catenatus*. Sovic et al. (2019) have recently shown that the contemporary effective values of most populations of *S. catenatus* are < 50 , suggesting that genetic drift and the negative genetic impacts of inbreeding could be significant yet evidence in support of this possibility is unclear. Specifically, heterozygosity-fitness correlations based on neutral genetic markers and body condition show few positive relationships consistent with the negative effects of drift and inbreeding but the power of these tests may be limited because of body condition is an indirect measure of individual fitness (Gibbs & Chiocchi, 2012; Sovic et al., 2019). Resolving how genetic factors impact the future viability of *S. catenatus* at the species and individual population level is an important goal of management plans at state and federal levels (Szymanski et al., 2016). Genome-level quantification of levels of inbreeding and genetic load similar to those conducted in other threatened species in similar conservation situations would address this need.

To this end, we use data from resequenced genomes from 90 individual *S. catenatus* to conducted a detailed analyses of inbreeding, population demography and levels of deleterious genetic load at the species and population level. To provide necessary context for our results, we also generated similar data for 10 individuals

from a single population of a closely-related yet non-threatened species, the Western Massasauga rattlesnake (*S. tergeminus*). In contrast, to *S. catenatus*, *S. tergeminus* is relatively common through much of its range and genetic studies show no evidence for high levels of population genetic structure supporting the idea that it represents an outbred species that has experienced far fewer anthropogenic impacts on its genetic makeup (Bylsma, 2020; McCluskey & Bender, 2015; Ryberg, Harvey, Blick, Hibbitts, & Voelker, 2015).

We used these data to address the following questions: What are the genome-wide levels of inbreeding and is inbreeding higher *S. catenatus* as compared to *S. tergeminus*? Based demographic analyses that use ROH size and abundance distributions, what are trends in population size over recent timescales and how do these aid in interpreting observed levels of inbreeding and levels of genetic load in each species? What are levels of different types of deleterious mutations in both species and among *S. catenatus* populations and what evolutionary processes underlie differences in load between species and *S. catenatus* populations? Our results support an emerging idea that the historical demography of a threatened species has a significant impact on the type of genetic load present which impacts conservation actions such as genetic rescue and also identify populations of *S. catenatus* whose viability may be at risk due to genetic factors.

2 | MATERIALS AND METHODS

2.1 | Samples and DNA sequencing

We analyzed resequenced whole genome data from 90 *S. catenatus* individuals from nine of the populations described in Ochoa et al. (2020) (see Table S1 for population names, codes, and numbers of individuals analyzed, and Figure 1 in Ochoa et al. [2020] for map showing population locations). We also analyzed samples from a population of *S. tergeminus* from Cheyenne Bottoms, Kansas (Table S1) as a reference for expected patterns of variation in a closely-related yet outbred species (see above).

We used genomic DNA previously extracted from blood samples (see Sovic et al. [2019]) to prepare paired-end (PE) shotgun libraries with 500-bp inserts and to generate PE 150-bp reads using Illumina sequencing platforms. Most samples were sequenced at medium or low coverage ($< 15\times$), but a small number of samples were sequenced at high coverage ($> 15\times$) (see Table S1 for mean coverage statistics for individual samples).

2.2 | Data processing and mapping

We used Trim Galore v.0.4.5 (<https://github.com/FelixKrueger/TrimGalore>) to remove the adapter sequences from the PE reads, trim low-quality 5' and 3' ends (i.e., ends with missing information 'N', Phred-scaled scores < 20 , or poly-G tails), and discard PE reads if any or both were < 100 bp after trimming. We then used BWA v.0.7.15 (Li & Durbin, 2009) and the 'MEM' algorithm to map the processed PE reads to a *S. catenatus* reference genome (~ 1.6 Gb; 7,469 scaffolds; scaffold N50 = 1,045 Mb; Broe et al., in prep.). Next, we used SAMtools v.1.3.1 (Li et al., 2009) to remove potential PCR duplicates, unmapped reads, and reads with unmapped partners from the resulting BAM files. Finally, we used GATK v.3.8 (McKenna et al., 2010) to realign indels by applying the 'RealignerTargetCreator' and 'IndelRealigner' functions.

2.3 | Genetic diversity and inbreeding

We compared levels of genetic diversity, size and abundance of ROHs, and relatedness between combined samples of *S. catenatus* and *S. tergeminus* and between individual populations of *S. catenatus*. We used ROHan v.1.0 (Renaud, Hanghøj, Korneliussen, Willerslev, & Orlando, 2019) to estimate genome-wide heterozygosity (Θ_W), fraction of the genome in ROH (F_{ROH}), and number of ROHs (N_{ROH}) for each sample. This program combines local Bayesian and hidden Markov models to generate reliable estimates of Θ_W and identify ROHs from low-coverage samples; furthermore, it does not require stringent mapping and base quality filters, since these metrics are informative for the models (Renaud et al., 2019).

Following Benazzo et al. (2017), we defined ROHs as genomic regions > 50 kb with a heterozygosity rate $[?] \ 5 \times 10^{-4}$ (i.e., $[?] \ 25$ heterozygous genotypes in 50-kb sliding windows), thus accounting for potential sequencing errors. For this analysis we used individual BAM files downsampled to 5x coverage to make samples statistically comparable. Also, to maximize our ability to detect long ROHs (see below), we limited

searches to the 135 scaffolds [?] 2 Mb in length, which comprised ~24% of the *S. catenatus* genome assembly (Broe et al., in prep.).

We quantified the impact of inbreeding on genome-wide levels of variation by comparing individual estimates of F_{ROH} with individual estimates of Θ_{W} and N_{ROH} . Inbred individuals should show a high F_{ROH} and low Θ_{W} (Saremi et al., 2019), whereas individuals from populations that have experienced a recent bottleneck should show an excess of N_{ROH} (Ceballos et al., 2018). To make cross-study comparisons with other threatened and endangered species and subspecies (Benazzo et al., 2017; Grossen et al., 2020; Robinson et al., 2019; Saremi et al., 2019; van der Valk et al., 2019), we also recalculated F_{ROH} for ROH sizes of [?] 0.1, 1, 2, and 2.5 Mb.

Finally, estimates of individual relatedness can provide an additional evaluation of the level of inbreeding occurring within populations. We used ANGSD v.0.930 (Korneliussen, Albrechtsen, & Nielsen, 2014) and NgsRelate v.2 (Korneliussen & Moltke, 2015) to calculate relatedness for pairs of individuals from the same population based on genotype likelihood distributions from low-coverage data. We focused on the r_{xy} statistic (Hedrick & Lacy, 2015) because it is designed for estimating individual relatedness in populations where inbreeding occurs. We used the downsampled BAM files (with reads mapped only to the 135 scaffolds [?] 2 Mb) grouped into populations for the analysis, while ANGSD/NgsRelate filtering parameters consisted of setting mapping and base qualities [?] 20, SNP P [?] 1×10^{-6} , and MAF [?] 0.05.

2.4 Recent demographic history

The size distribution of ROH tract lengths can provide insights into the recent demographic history of individuals (Ceballos et al., 2018). Since homozygous tract length declines exponentially as a function of recombination rate and time, ROHs due to background inbreeding during an ancestral bottleneck are expected to be shorter than ROHs caused by recent inbreeding (Pool and Nielsen, 2009). ROH tract length can then be correlated with the expected number of generations since the individual's maternal and paternal lineages shared a common ancestor. Clusters of different lengths in the overall distribution can then be used to identify and date recent populations bottlenecks due to close inbreeding and/or population declines (Pemberton et al., 2012; Saremi et al., 2019).

To identify recent demographic events in *S. catenatus* and *S. tergeminus* using this approach we used mclust v.3 in R (R Core Team, 2020) to identify distinct ROH size clusters (see Pemberton et al. [2012]) from a pooled distribution of *S. catenatus* samples. Schield et al. (2020) have recently documented significant differences in recombination rates on snake macro- and microchromosomes. As such, we restricted data to scaffolds that aligned to macrochromosomes identified in a prairie rattlesnake (*Crotalus viridis*) genome assembly (Schield et al., 2019). We did this by mapping the 135 scaffolds [?] 2 Mb to the *C. viridis* assembly using [?] 95% sequence identity for 5-kb sliding window tracts in MashMap v.2.0 (Jain, Koren, Dilthey, Phillippy, & Aluru, 2018).

We dated the resulting ROH size clusters using the equation $g = 100/(2rL)$, where r is given in cM Mb⁻¹ and L in Mb (Saremi et al., 2019). Because there are no estimates of r for any reptilian species, we used a recombination rate estimate of 2.8 cM Mb⁻¹ for chicken (*Gallus gallus*) macrochromosomes (International Chicken Genome Sequencing Consortium, 2004). We used an estimated generation time of 3 years for both *Sistrurus* species as done previously for *S. catenatus* by Sovic et al. (2019).

2.5 | Long-term demographic history

We also assessed the long-term demographic history of each *Sistrurus* species using PSMC v.0.6.4 (Li & Durbin, 2011), which infers changes in effective sizes (N_e) over extended periods of time based on the physical distance of heterozygous genotypes across a single diploid genome. We used the highest-coverage sample from each species (sca1038 and ste0121; Table S1) and GATK's (McKenna et al., 2010) best practices pipeline (www.broadinstitute.org) to call genome-wide SNPs after excluding genomic regions containing repetitive elements as defined by RepeatMasker v.4.0.7 (Smit, Hubley, & Green, 2010). As such, in GATK we used DP < 8, DP > twice the individual coverage, FS > 60.0, MQ < 20.0, MQRankSum < -12.5, QD <

2.0, and ReadPosRankSum < -8.0 as filtering parameters. We ran PSMC with these data for 100 bootstrap replicates and assumed a neutral mutation rate of 7.2×10^{-9} per generation (see squamates in Green et al., [2014]) for each species.

2.6 | Genetic load

We evaluated genetic load at the species and population level using variants detected in a set of conserved single copy coding loci (Benchmarking Universal Single-Copy Orthologs or BUSCOs; Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) that are expected to be under purifying selection due to evolutionary constraints across specific lineages. To assess genetic load at the species level we focused on derived nonsynonymous substitutions at sites that were polymorphic in either one or both species (“Species level mutations”). Using these sites will capture differences in genome-wide heterozygosity between species, which can substantially impact estimates of load (van der Valk et al., 2019). To estimate genetic load at the population level we used sites with derived nonsynonymous substitutions that were polymorphic within specific populations (“Population level mutations”). These variants capture a different aspect of genetic load that reflects the evolutionary process shaping mutation variation within populations independent of species level effects.

To assess variation in BUSCO loci, we first performed homology searches for these loci in the *S. catenatus* reference genome. We used BUSCO v.3.1.0 (Simão et al., 2015) and the Tetrapoda v.10 database to identify homologs for 5,310 query BUSCO loci. We then used ANGSD to characterize SNPs and genotypes for these loci in downsampled (5x) sequences for both *S. catenatus* and *S. tergeminus* samples. ANGSD runs used the ‘sites’ parameter to restrict searches to BUSCO annotations, eliminate reads with mapping qualities < 20, remove bases with quality scores < 20, retain SNPs identified with P -values < 1×10^{-6} , and call genotypes with posterior likelihoods of $P > 0.95$. To identify derived nonsynonymous substitutions in BUSCO loci for both *S. catenatus* and *S. tergeminus*, we compared detected variants with those identified in a genome assembly of a closely-related outgroup (*S. miliarius*; Broe et al., in prep.) (for detailed bioinformatic methods see Supplemental Material). This analysis identified both Species and Population level mutations (see above).

Next, we used PROVEAN v.1.1 (Choi & Chan, 2015) to classify the potential deleterious impact of specific derived nonsynonymous alleles in the BUSCO loci (Perrier, Ferchaud, Sirois, Thibault, & Bernatchez, 2017; Renaut & Rieseberg, 2015). PROVEAN uses BLAST searches of protein databases and subsequent alignments of related protein sequences from other related species to generate a measure (“Delta alignment score”) of how phylogenetically conserved a given nonsynonymous variant is among homologous protein sequences. To characterize load, we used Delta scores to pool nonsynonymous mutations into “benign” (PROVEAN scores > -2.5) and “deleterious” (PROVEAN scores < -2.5). We also identified substitutions that resulted in intermediate stop codons or “loss of function” mutations—not accounted for by PROVEAN—as an additional category of potentially highly deleterious mutations.

To estimate overall levels of deleterious mutations present, we calculated the number and mean proportion of derived alleles present in an individual that were deleterious or loss of function with respect to the benign ones. For comparisons using Species level mutations, we compared these values averaged across all *S. catenatus* and *S. tergeminus* individuals. For comparisons that used Population level mutations, we generated mean values for the proportion of all derived deleterious alleles for a specific mutation category across all individuals from single *S. catenatus* populations and the only *S. tergeminus* population analyzed. Finally, for Species level mutations only, we compared the mean PROVEAN scores for all deleterious mutations found in an individual and compared this value separately for heterozygous and homozygous genotypes between individuals from each species. For this analysis we coded loss of function mutations as having a PROVEAN score of -12.

We used the nonparametric Mann-Whitney U -test for statistical comparisons between groups. We further used the false discovery rate method (Benjamini & Hochberg, 1995) for P -value adjustments across multiple tests.

3 | RESULTS

3.1 | Genetic diversity and inbreeding

Fine-scale representations of the distribution of ROHs across the longest scaffold analyzed suggest that such tracts are longer and more abundant in threatened *S. catenatus* individuals relative to non-threatened *S. tergeminus* individuals (Figure 1). Further analyses confirmed that *S. catenatus* shows low genetic diversity and high inbreeding levels with respect to *S. tergeminus* at the genome-wide level (Figure 2). Specifically, when averaged across all individuals, *S. catenatus* had a mean Θ_W (scaled to 10^3 hereafter) of 2.71 (SE = 0.01) and F_{ROH} of 0.557 (SE = 0.013), while *S. tergeminus* had a mean Θ_W of 4.44 (SE = 0.02) and F_{ROH} of 0.067 (SE = 0.001) (Figure 2a); differences between species for both metrics were highly significant ($P < 0.001$). These patterns are consistent with repeated reductions in effective population size and higher levels of inbreeding in *S. catenatus* compared to *S. tergeminus*.

A comparison of adjusted F_{ROH} values across studies suggests that *S. catenatus* is moderately inbred with respect to other threatened and endangered species and subspecies (Table S2 – see also Brüniche-Olsen et al., 2018). For these comparisons we recalculated F_{ROH} using data from our study to be comparable with the different size criteria used in different studies (see Table S2). We also recognize that such comparisons are problematic because of varying criteria used to define ROH tracts (e.g., maximum number of heterozygous genotypes allowed to account for sequencing errors). However, based on these adjusted comparisons, *S. catenatus* is half as inbred as Apennine brown bears (*Ursus arctos marsicanus*) that have existed in isolation for hundreds of years (Bennazo et al., 2017), but shows similar levels of inbreeding to Alpine ibex (*Capra ibex*) that have recently undergone serial population bottlenecks (Grossen et al., 2020) (Table S2). Our interpretation that *S. catenatus* is moderately inbred is also consistent with F_{ROH} values generated from 100 mammals representing species with varying population sizes (van de Valk et al., 2019). Based on a size criterion of ROH > 0.1 Mb, the estimated F_{ROH} value for *S. catenatus* (0.552) is in the upper third of values for species and subspecies analyzed. On the other hand, F_{ROH} values for *S. tergeminus* are always representative of an outbred species that shows little inbreeding (Table S2).

In contrast, to species-level comparisons of Θ_W and F_{ROH} across the nine *S. catenatus* populations examined show few significant differences for both metrics (Figures 2a and S1). Population-level values of Θ_W ranged from 2.58 (SE = 0.05) in ROME to 2.78 (SE = 0.01) in KPWA, with only two pairwise comparisons among populations being significant ($P < 0.05$; Figures 2a and S1). Likewise, values of F_{ROH} ranged from 0.513 (SE = 0.045) in PROF to 0.608 (SE = 0.033) in CEBO, with only three pairwise comparisons among populations being significant ($P < 0.05$; Figures 2a and S1). Overall, these results suggest a remarkable degree of uniformity across *S. catenatus* populations with respect to their long-term demographic trajectories.

S. catenatus genomes were also composed of a substantially greater number of ROHs ($N_{ROH} = 364$, SE = 10) relative to *S. tergeminus* genomes ($N_{ROH} = 150$, SE = 8) ($P < 0.001$; Figure 2b). This two-fold increase in N_{ROH} (combined with the eight-fold increase in F_{ROH} as above) is consistent with the effects of recent bottlenecks leading to inbreeding in *S. catenatus* populations, as it implies that ROH tracts have increased in mean length (Figure 2b; also see Figure 1 in Ceballos et al. [2018]).

We also found substantial variation among individuals for both F_{ROH} and N_{ROH} metrics. For example, there is a two-fold difference in F_{ROH} and even greater variation in N_{ROH} among individuals from the SPVA population (Figure S2). Some of this variation could be due to differences in the degree of relatedness among individuals within single populations, since, unlike the only *S. tergeminus* population analyzed, *S. catenatus* populations presented small but perhaps significant (mean = 0.082, SE = 0.012) proportions of close relatives ($r_{xy} > 0.125$; Figure 3). These patterns are also consistent with a recent onset of inbreeding in *S. catenatus* populations.

3.2 | Recent demographic history

The presence of distinct ROH size clusters represent time-stamps for when population bottlenecks have occurred in the recent past (Bosse et al., 2012; Pemberton et al., 2012). Analysis of the overall distribution of ROH lengths in *S. catenatus* shows evidence for three size classes: short (0.05-0.2 Mb), medium (0.2-0.7 Mb), and long (0.7-7.9 Mb) (Figures S3 and S4). Assuming a recombination rate of 2.8 cM Mb⁻¹ and

a generation time of 3 years (see Materials and Methods), we estimated the following time-frames for the demographic events responsible for each ROH size class: short (268-1071 years before present [ybp]), medium (77-268 ybp), and long (7-77 ybp). The latter two time-frames are consistent with possible anthropogenic impacts on this species (Sovic et al., 2019).

The effects of these events can be assessed by comparing differences in class N_{ROH} between species. As such, N_{ROH} for the three ROH size classes (i.e., short, medium, long) was always greater in *S. catenatus* relative to *S. tergestinus*, which is consistent with a greater impact of bottlenecks in *S. catenatus* during the recent past (Figure 4a). Specifically, *S. catenatus* genomes had a greater number of short ($N_{\text{ROH}} = 177$, $\text{SE} = 7$), medium ($N_{\text{ROH}} = 107$, $\text{SE} = 4$), and long ($N_{\text{ROH}} = 65$, $\text{SE} = 2$) ROH tracts than *S. tergestinus* genomes ($N_{\text{ROH}} [\text{short}] = 120$, $\text{SE} = 7$; $N_{\text{ROH}} [\text{medium}] = 18$, $\text{SE} = 1$; $N_{\text{ROH}} [\text{long}] = 6$, $\text{SE} < 1$) ($P < 0.05$ in all cases; Figure 4a). Fold increases in each class suggest that, in *S. catenatus*, ROH tract gains have been most substantial lately (i.e., 7-77 ybp), followed by ROH tract gains 77-268 and 268-1071 ybp.

We can also use the relative proportion of each size class to measure the relative impact of each event in a way that accounts for overall differences in inbreeding between the species (Figure 4c). Using this approach, small ROHs associated with the oldest bottleneck made up 93% of all ROH segments in *S. tergestinus* as compared to 77% in *S. catenatus* ($P < 0.05$), whereas the relative proportion of medium (*S. catenatus* : 17%; *S. tergestinus* : 5%; $P < 0.001$) and large (*S. catenatus* : 5%; *S. tergestinus* : 1%; $P < 0.001$) ROHs was significantly greater in *S. catenatus* (Figure 4c). These proportional increases indicate that more recent demographic events have had a substantially greater impact on *S. catenatus*.

Finally, *S. catenatus* populations showed few differences, if any, in class N_{ROH} profiles (Figures 4b and S5a-c). Overall, 14, 33, and 0% of N_{ROH} pairwise comparisons among populations were significant ($P < 0.05$) for short, medium, and long ROH tracts, respectively (Figures 4b and S5a-c). We obtained similar species- and population-level results using relative proportion metrics for each ROH size class (Figures 4d and S5d-f). Broadly speaking, this suggests similar magnitudes of population declines across different populations spanning the range of the species.

3.3 | Long-term demographic history

Simulations with PSMC indicate that, since the Middle Pleistocene (ca. 500,000 ybp), *S. catenatus* has had a consistently smaller N_e than *S. tergestinus* (Figure 5). These results show that *S. tergestinus* reached a maximum N_e of 180,000 individuals ca. 400,000 ybp, and had a stable N_e of ~60,000 individuals ca. 40,000-110,000 ybp, during the Upper Pleistocene (Figure 5). After this time, *S. tergestinus* experienced declines in N_e , reaching a size of ~10,000 individuals ca. 10,000 ybp, in the Holocene (Figure 5).

In contrast, *S. catenatus* reached a larger maximum N_e of 230,000 individuals ca. 1,000,000 ybp, during the Lower Pleistocene (Figure 5). This species then experienced consistent declines in N_e until reaching a size of ~20,000 individuals ca. 25,000-110,000 ybp, during the Upper Pleistocene (Figure 5). Similar to *S. tergestinus*, our demographic simulations suggest that *S. catenatus* had an N_e of < 10,000 individuals ca. 10,000 ybp (Figure 5).

3.4 | Genetic load

At the interspecific level (*S. catenatus* and *S. tergestinus* samples combined), we identified 28,493 sites containing derived nonsynonymous substitutions from 3,800 BUSCOs. When we analyzed the degree of evolutionary conservation of these sites using PROVEAN, we characterized 91% (26,035) of these sites as benign mutations, 8% (2,189) as deleterious mutations, and only 1% (269) as loss of function mutations. Collectively these represent Species level mutations in our analyses.

Based on these variants, on average, the genomes of inbred *S. catenatus* individuals had lower absolute numbers and lower relative proportion of all classes of mutations compared to the genomes of outbred *S. tergestinus* individuals ($P < 0.001$ in all cases; Figures 6a-b). There were ~2-fold differences in absolute numbers of benign (*S. tergestinus* : mean = 7054, $\text{SE} = 94$; *S. catenatus* : mean = 4999, $\text{SE} = 47$) and loss of function (*S. tergestinus* : mean = 65, $\text{SE} = 2$; *S. catenatus* : mean = 24, $\text{SE} < 1$) mutations compared to

a smaller, but still significant, difference in deleterious mutations (*S. tergeminus* : mean = 290, SE = 6; *S. catenatus* : mean = 253, SE = 3). Differences in the relative proportion of deleterious variants are highest for loss of function alleles (*S. tergeminus* : mean = 0.147, SE = 0.003; *S. catenatus* : mean = 0.047, SE = 0.001), intermediate for benign mutations (*S. tergeminus* : mean = 0.154, SE = 0.001; *S. catenatus* : mean = 0.102, SE = 0.001), and lowest for deleterious mutations (*S. tergeminus* : mean = 0.071, SE = 0.001; *S. catenatus* : mean = 0.061, SE = 0.001) (Figure 6b).

Under the assumption that most deleterious mutations that arise in a population are recessive (Agrawal & Whitlock, 2011) the genotypic configuration of mutations can have a substantial impact on their realized negative impacts on individual fitness (Mathur & DeWoody, 2021). In this respect, *S. catenatus* has a significantly greater proportion of homozygotes relative to *S. tergeminus*, demonstrating greater levels of realized genetic load on a per-individual basis for two of three classes of mutations (Figure 6c). The proportion of deleterious homozygous genotypes was > 2-fold higher in *S. catenatus* compared to *S. tergeminus* (*S. tergeminus* : mean = 0.017, SE = 0.001; *S. catenatus* : mean = 0.037, SE = 0.001), while benign homozygous genotypes were also found at a significantly greater frequency in *S. catenatus* (*S. tergeminus* : mean = 0.060, SE = 0.001; *S. catenatus* : mean = 0.074, SE = 0.001) (Figure 6c). The exception is loss of function homozygous genotypes, which are significantly less often found in *S. catenatus* (*S. tergeminus* : mean = 0.043, SE = 0.003; *S. catenatus* : mean = 0.025, SE = 0.001; Figure 6c). Our interpretation is that this may represent extreme purging of highly deleterious mutations in small *S. catenatus* populations, as has been found in extremely small and repeatedly bottlenecked populations of Alpine ibex (Grossen et al., 2020).

Finally, although *S. tergeminus* had overall more deleterious mutations than *S. catenatus*, the fitness related impacts of mutations may be reduced because a greater proportion of all classes of deleterious mutations are found in heterozygous genotypes in *S. tergeminus* individuals (Figure 6d). This difference is highly significant for all three mutation classes ($P < 0.001$). This result supports the interpretation that the greater number of derived deleterious mutations in *S. tergeminus* may have reduced phenotypic expression because they are more often found in heterozygotes and less often found in homozygotes.

Mean PROVEAN scores of alleles found in both homozygote and heterozygote genotypes are also significantly lower (more deleterious) in *S. catenatus* compared to *S. tergeminus* (Figure 6e). These results support the interpretation that high levels of drift present in small *S. catenatus* populations have allowed the persistence of mutations with greater potential negative impacts regardless of genotype class.

In contrast, analysis that use Population level mutations that capture population level processes show the opposite pattern in the form of higher levels of deleterious mutations with greater impact in *S. catenatus* (Fig. 7). For example, when data is pooled across all individuals within each species, the proportion of all variants made up by all classes of Population specific deleterious mutations is significantly higher in *S. catenatus* compared to *S. tergeminus* (Fig. 7a) as is the frequency of homozygous genotypes (Fig. 7b). Patterns of variation in heterozygotes are less clear: there is the expected increases in heterozygote frequency of beneficial mutations but no significant difference in heterozygote frequency between species for other two more negative classes of mutations possibly due to the high variation present among individuals (Fig. 7c). Nonetheless, the overall pattern is that, as a class, individual deleterious mutations that escape the “drift sieve” at the species level are more abundant within individual populations and potentially have more impact on individual fitness.

Comparisons based on Population level mutations across individual *S. catenatus* populations also show substantial and often significant differences in genome-wide proportion of deleterious mutations, and homozygote and heterozygote frequency across all mutation classes (Figure 8). If we use the proportion of each class of deleterious mutations as a measure of load (Figure 8a-c) then there is > 2-fold variation across populations for each class of deleterious mutation (benign: 0.384 [KPWA] – 0.585 [MOSQ]; deleterious: 0.238 [SSSP] – 0.493 [MOSQ]; loss of function: (0.228 [KPWA] – 0.410 [MOSQ])). These analyses identify specific populations that have high levels of load and are potentially subject to significant negative genetic impacts (MOSQ, CEBO, and SPVA) but also populations with the lowest levels of load (KPWA).

Finally, to assess possible drivers of the high level of variation in load across populations, we estimated correlations between estimates of contemporary effective size (Sovic et al., 2019) and mean proportion of deleterious mutations and levels of inbreeding (estimated as mean FROH) and mean frequency of homozygote deleterious genotypes. There were negative and marginally significant correlations between effective size and each class of mutations, supporting the predicted impact of drift acting through differences in population size on overall levels of genetic load (benign: Spearman rank correlation (R) = -0.714; P = 0.06; deleterious: R = -0.667; P = 0.08; loss of function: R = -0.381; P = 0.3). There were also the expected positive correlations between F_{ROH} and frequency of homozygotes, but none approached significance (results not shown).

4 | Discussion

4.1 | Inbreeding and demography

Demographic analyses indicate that increased inbreeding in *S. catenatus* is likely due to decreased population size relative to *S. tergeminus* over both long- and short term timescales. Over evolutionary timescales (1000's of years), a coalescent-based PSMC analysis shows that for much of the last 100,000 years *S. catenatus* has consistently had an effective size that is $> 2 - 3X$ smaller than *S. tergeminus* even in the face of shifting distributions due to impacts of ice sheet expansion and contraction during the late Pleistocene (Pielou, 1992). Fundamental differences in the ecology of these species may account for the long-term differences in effective size. *S. catenatus* is a habitat specialist which occupies wetlands or mesic prairie habitats (Szymanski et al., 2016) which are patchily distributed leading to greater population structure and smaller effective sizes (Sovic et al., 2019). In contrast, *S. tergeminus* is found in more wide-spread continuous xeric grasslands (Stebbins, 1966) and this is reflected in the lack of genetic structure and large effective sizes observed in nature (McClusky & Bender, 2015).

Bottlenecks over recent timescales have also increased inbreeding in *S. catenatus* relative to *S. tergeminus*. The oldest event is most consistent with the timescale of bottlenecks detected by Sovic et al. (2019) and had a proportionally greater impact on *S. tergeminus* based on the relative abundance of small ROHs in each species. The causes of this decline are unknown, but most likely to be related to large-scale environmental changes related to climate that have previously been hypothesized to impact the distribution of this (Cook, 1992) and other species in this region of North America (Soltis, Morris, McLachlan, Manos, & Soltis, 2006).

In contrast, we estimate that the two most recent bottleneck events occurred within the last 200 years and the greater relative abundance of both medium and long ROHs in *S. catenatus* suggest that relative declines in population size have been greater in this species than in *S. tergeminus*. This timeframe coincides with the colonization and subsequent landscape modification by European settlers in North America (Pielou, 1992; Schmidt, 1938) which has resulted in the habitat destruction and fragmentation that has played a key role in recent range-wide declines of *S. catenatus* (Szymanski et al., 2016). Overall our analyses provide new historical evidence in support of the idea that anthropogenic impacts have had a consistently greater impact on *S. catenatus*.

Compared to the species level results, population level comparisons show few significant differences in mean individual FROH and theta between populations. We suggest that this is because all populations have small contemporary effective sizes (Sovic et al., 2019) implying strong inbreeding effects in all populations and high variance in these parameters between individuals within populations (Figure 2a) that reduces statistical power to detect differences. Surprisingly, despite this lack of detectable differences in genome level indications of inbreeding we nonetheless still detect population differences in population specific genetic load (see below). Regardless, our results argue that inbreeding effects in *S. catenatus* are most strongly manifested at the species rather than the population level.

4.2 | Genetic Load

Our ability to estimate genetic load in endangered species has been transformed by applying novel analytical methods to infer the negative impacts of variants detected in genomic datasets (van Oosterhout, 2020). Our results show that the perspectives provided by mutations with different evolutionary histories is significant

for understanding variation in genetic load at the species and populations levels.

Specifically, we estimated species level genetic load using derived nonsynonymous variants at sites that were polymorphic in either one or both species (Species level mutations). Using criterion to define sites used for estimating species level load is important because incorporates sites that reflect the almost two-fold difference in genome-wide heterozygosity present between species which can substantially impact estimates of load (Figure 6). Based on these sites we found that, on average, the genomes of outbred *S. tergeminus* individuals had significantly more deleterious mutations compared to more inbred *S. catenatus* individuals when genetic load was estimated in terms of mean numbers of mutation or proportion of total mutations (Figure 6). These results argue that based on absolute numbers of deleterious mutations the inbred threatened species (*S. catenatus*) has lower levels of genetic load than the more outbred common species (*S. tergeminus*). This is contrary to the long-standing idea that small inbred populations of threatened species will incur genetic erosion in the form of increased genetic load (Frankham et al., 2017).

However, the impact of deleterious mutations is highly dependent the genotypic segregation patterns of deleterious alleles and the assumed dominance coefficients of mutations (Charlesworth, 2009). In particular, empirical evidence suggests that many deleterious mutations are recessive and are only expressed if they segregate in homozygous genotypes (Agrawal & Whitlock, 2011). If this is the case then the impact of mutation is lessened if they mainly found in heterozygote genotypes. This is the case in outbred *S. tergeminus*: heterozygotes containing deleterious mutations are from ~ 3 – 5-fold more common in *S. tergeminus* than *S. catenatus* genomes (Figure 6). In addition, on a per site basis, the frequency of deleterious mutations is between 40 – 68% less in *S. tergeminus* (data not shown). Together these patterns argue that although *S. tergeminus* has higher “Potential” genetic load in the form of absolute numbers of deleterious mutations, the extent to which the negative effects of these mutations (“Realized” load) is blunted by the fact that proportionately more occur in presumably unexpressed heterozygous genotypes and are also found at lower frequencies on a per site basis (Mathur & DeWoody, 2021).

Recently, van der Valk et al. (2019) have argued that species level differences in genetic load need to take into account the impact of long term demographic patterns on levels of deleterious mutations. Specifically, if threatened species have always persisted at low effective sizes over evolutionary time-scales then selective purging of deleterious mutations along with other types of variants due to sustained genetic drift effects can explain reduced genetic load in threatened species that currently exist in small populations. The long and short term demographic differences documented here support this explanation for the decreased genetic load shown in *S. catenatus* compared to *S. tergeminus*. The ROH analyses shows that recent population bottlenecks have occurred in both species with more recent events having a greater impact on *S. catenatus*. In addition, coalescent-based analyses demonstrate that *S. catenatus* has consistently had a smaller effective size over the past 100,000 ybp. These results argue that drift effects have been consistently stronger in the species with lower absolute levels of genetic load which is consistent with the explanation that lower load is due to more sustained purging of mutations over short and long time-scales.

We also examined genetic load at the population level using a different set of sites containing derived mutations were polymorphic found within populations (Population level mutations). These variants capture a different aspect of genetic load other than that due to differences in long term genetic purging between species. In contrast to species level patterns *S. catenatus* populations have, on average, higher proportions of deleterious mutations on a per genome basis across all impact classes and all classes of mutations are expressed at higher levels in homozygote genotypes compared to *S. tergeminus*. Finally, genetic load measured as proportion of deleterious mutations is inversely related to short term effective population size. These patterns mirror those found in other threatened species with small populations (Grossen et al., 2020; Robinson et al., 2016; Robinson et al., 2019) and reflect the negative impact of genetic drift on the efficiency with which natural selection removes deleterious mutations from small populations (Lohmueller, 2014).

These mechanisms extend to explaining the significant differences in genetic load levels between *S. catenatus* populations that are potentially responsible for population differences in negative genetic effects such as inbreeding depression (Kardos et al., 2016). The inverse correlation between contemporary effective size

and genetic load signals a strong impact of recent levels of genetic drift on observed levels of load in *S. catenatus* populations. These effects are likely to also have a historical component because contemporary and historical effective sizes are positively correlated in this species (Ochoa et al., 2020). This is relevant because allele frequency changes likely are the result of processes that take place over longer timescale than the few generations reflected in contemporary estimates of effective size.

4.3 Conservation Implications

We see implications from our results for managing genetic erosion in *S. catenatus* due to range-wide declines in population numbers. First, at the species level, reduced genome-wide genetic load due to long term purging suggests that future population declines in less impacted regions such as the northern parts of the species range will have fewer negative genetic impacts than comparable declines in threatened snakes with historically larger effective sizes. *S. catenatus* may be more resistant to the process of mutational meltdown that is proposed to occur in small populations of threatened species (Frankham et al., 2017) leading to a focus on the greater importance of ecological factors to long term population viability (Lande, 1988).

Second, consideration of the potential effects of long-term purging on mutational load have led to suggestions that conventional guidelines for conducting genetic rescue of small inbred populations of threatened species need to be reconsidered (Kyriazis et al., 2020; van der Valk et al., 2019; but see Ralls et al., 2020). Specifically, Kyriazis et al. (2020) argue that the typical approach for managing these populations is to maintain high genetic diversity through the transfer of individuals from large genetically diverse populations but that this carries a risk of introducing large numbers of deleterious mutations that can be exposed by inbreeding. Our empirical results show that, in fact, there is an inverse relationship between levels of mutational load and effective size among *S. catenatus* populations or that large potential donor populations have fewer not more potential deleterious mutations present. This observation, combined with the fact that donor individuals from larger populations will have greater numbers of potentially adaptive variants (Ochoa et al., 2020) broadly supports the idea that choosing donor individuals from populations with large effective sizes is a sound strategy for genetic rescue in this species. This claim comes with two important caveats. It assumes a high degree of sharing of deleterious mutations between populations such that the inter-population transfer of individuals will not introduce large numbers of new distinct negative mutations. Second, it also assumes that adaptive variation that evolves through local adaptation will have similar positive effects on individual fitness in populations receiving translocated individuals.

More broadly, our study provides an example of how whole genome sequence from threatened and endangered species can provide new approaches to assess patterns of functional variation that impact genetic erosion in these species (Leroy et al., 2018). When combined with the results of Ochoa et al. (2020) it provides a rare comprehensive assessment of two of the key aspects of genetic erosion in a single endangered species namely the magnitude and evolutionary mechanisms shaping negative genetic load and adaptive genetic variation. As such it represents an important realization of the application of genomic tools and analyses for addressing genetic issues related to conservation of biodiversity at the species and population levels (Funk et al., 2019; Hohenlohe, Funk, & Rajora, 2021).

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 Villeneuve, and Doug Wynn for their enthusiastic and generous assistance with finding snakes and/or providing samples. We thank Anna Brüniche-Olsen for introducing us to the idea of ROHs and encouraging us to think about analyzing them in our snakes, Scott Martin, Drew Schield and Todd Castoe for help with data and analyses and Anna Brüniche-Olsen, J. Andrew DeWoody, Gideon Bradburd, Robert Fitak, Katerina Guschanski, and Christine Grossen for discussions and comments on the manuscript. We also thank Kate Parsons and Carolyn Caldwell for their long-time support of our conservation genetics

work on endangered snakes. Computational analyses were performed on the Ohio Biodiversity Conservation Partnership Computing Cluster and using resources provided by the Ohio Supercomputer Center. This work was 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 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 1638872.

Data Availability Statement :

The data that support the findings of this study are openly available in [to be determined] at [http://doi.org/\[doi\]](http://doi.org/[doi]), reference number [to be determined].

Author Contributions:

Designed research: AO and HLG

Performed research: AO and HLG

Analyzed data: AO

Wrote the paper: AO and HLG

REFERENCES

- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., ... Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods* , 7 (4), 248-249. doi: 10.1038/nmeth0410-248
- Agrawal, A. F., & Whitlock, M. C. (2011). Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* , 187 (2), 553-566. doi: 10.1534/genetics.110.124560
- Allendorf, F. W. (2017). Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology* , 26 (2), 420-430. doi: 10.1111/mec.13948
- Benazzo, A., Trucchi, E., Cahill, J. A., Maisano Delser, P., Mona, S., Fumagalli, M., ... Bertorelle, G. (2017). Survival and divergence in a small group: the extraordinary genomic history of the endangered Apennine brown bear stragglers. *Proceedings of the National Academy of Sciences* , 114 (45), E9589-E9597. doi: 10.1073/pnas.1707279114
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B* , 57 (1), 289-300. doi:10.2307/2346101
- Bosse, M., Megens, H. J., Madsen, O., Paudel, Y., Frantz, L. A., Schook, L. B., ... Groenen, M. A. (2012). Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genetics* , 8 (11), e1003100. doi: 10.1371/journal.pgen.1003100
- Bruniche-Olsen, A., Kellner, K.F., Anderson, C.J., & DeWoody, J. A. (2018). Runs of homozygosity have utility in mammalian conservation and evolutionary studies. *Conservation Genetics* , 19 (6), 1295-1307. doi: 10.1007/s10592-018-1099-y
- Bylsma, R. R. (2020). *Population genetic and genomic analyses of western massasauga (Sistrurus tergeminus ssp.): subspecies delimitation and conservation status* (Unpublished MS thesis). Purdue University.
- Ceballos, F. C., Joshi, PK, Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics* , 19 (4), 220-234. doi:10.1038/nrg.2017.109
- Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* , 10 , 195-205. doi: 10.1038/nrg2526

- Chiucchi, J. E., & Gibbs, H. L. (2010). Similarity of contemporary and historical gene flow among highly fragmented populations of an endangered rattlesnake. *Molecular Ecology* , 19 (24), 5345-5358. doi: 10.1111/j.1365-294X.2010.04860.x
- Choi, Y., Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* , 31 (16), 2745-2747. doi: 10.1093/bioinformatics/btv195
- Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One* , 7 (10), e46688. doi: 10.1371/journal.pone.0046688
- Cook, F. R. (1992). After an ice age: zoogeography of the massasauga within a Canadian herpetological perspective. In B. Johnson & V. Menzies (Eds.), *International Symposium and Workshop on the Conservation of the Eastern Massasauga Rattlesnake, Sistrurus catenatus catenatus* (pp. 19-25). Toronto, Canada: Toronto Zoo.
- Frankham, R., Ballou, J. D., Ralls, K., Eldridge, M., Dudash, M. R., Fenster, C. B., ... Sunnucks, P. (2017). *Genetic management of fragmented animal and plant populations* . Oxford, UK: Oxford University Press.
- Funk, W. C., Forester, B. R., Converse, S. J., Darst, C., & Morey, S. (2019). Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists. *Conservation Genetics* , 20 (1), 115–134. doi: 10.1007/s10592-018-1096-1
- Garcia-Dorado, A. (2012). Understanding and predicting the fitness decline of shrunk populations: inbreeding, purging, mutation, and standard selection. *Genetics* , 190 (4), 1461-1476. doi: 10.1534/genetics.111.135541.
- Gibbs, H. L., & Chiucchi, J. E. (2012). Inbreeding, body condition, and heterozygosity-fitness correlations in isolated populations of the endangered eastern massasauga rattlesnake (*Sistrurus c. catenatus*). *Conservation Genetics* , 13 (4), 1133–1143. doi: 10.1007/s10592-012-0360-z
- Glemin, S. (2003). How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution* , 57 (12), 2678–2687. doi: 10.1111/j.0014-3820.2003.tb01512.x
- Government of Canada (2009). *Species at risk public registry* . Retrieved from <https://www.sararegistry.gc.ca>
- Green, R. E., Braun, E. L., Armstrong, J., Earl, D., Nguyen, N., Hickey, G., ... Ray, D. A. (2014). Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science* , 346 (6215), 12544-12549. doi:10.1126/science.1254449
- Grossen, C., Guillaume, F., Keller, L. K., & Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nature Communications* , 11 (1), 1001. doi: 10.1038/s41467-020-14803-1
- Hedrick, P. W., & Garcia-Dorado, A. (2016). Understanding inbreeding depression, purging, and genetic rescue. *Trends in Ecology & Evolution* , 31 (12), 940–952. doi: 10.1016/j.tree.2016.09.005
- Hedrick, P. W., & Lacy, R. C. (2015). Measuring relatedness between inbred individuals. *Journal of Heredity* , 106 (1), 20-25. doi:10.1093/jhered/esu072
- Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology* , 30 (1), 62–82. doi: 10.1111/mec.15720
- International Chicken Genome Sequencing Consortium (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* , 432 (7018), 695-716. doi: 10.1038/nature03154
- Jain, C., Koren, S., Dilthey, A., Phillippy, A. M., & Aluru, S. (2018). A fast adaptive algorithm for computing whole-genome homology maps. *Bioinformatics* , 34 (17), i748-i756. doi: 10.1093/bioinformatics/bty597

- Kardos, M., Qvarnstrom, A., & Ellegren, H. (2017). Inferring individual inbreeding and demographic history from segments of identity by descent in *Ficedula* flycatcher genome sequences. *Genetics* , 205 (3), 1319-1334. doi: 10.1534/genetics.116.198861
- Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., & Allendorf, F. W. (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary Applications* , 9 (10), 1205–1218. doi: 10.1111/eva.12414
- Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution* , 17 (5), 230–241. doi: 10.1016/S0169-5347(02)02489-8
- Kirkpatrick, M., & Jarne, P. (2000). The effects of a bottleneck on inbreeding depression and the genetic load. *American Naturalist* , 155 (2), 154-167. doi: 10.1086/303312
- Korneliussen, T. S., & Moltke, I. (2015). NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics* , 31 (24), 4009-4011. doi:10.1093/bioinformatics/btv509
- Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* , 15 (1), 356. doi:10.1186/s12859-014-0356-4
- Kyriazis, C. C., Wayne, R. K., & Lohmueller, K. E. (2020). Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evolution Letters* , 5 (1), 33–47. <https://doi.org/10.1002/evl3.209>
- Lande, R. (1988). Genetics and demography in biological conservation. *Science* , 241 (4872), 1455-1460. doi:10.1126/science.3420403
- Leroy, G., Carroll, E. L., Bruford, M. W., DeWoody, J. A., Strand, A., Waits, L., & Wang, J. (2018). Next-generation metrics for monitoring genetic erosion within populations of conservation concern. *Evolutionary Applications* , 11 (7), 1066–1083. doi: 10.1111/eva.12564
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , 25 (14), 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature* , 475 (7357), 493–496. doi: 10.1038/nature10231
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics* , 25 (16), 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lohmueller, K. E. (2014). The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genetics* , e1004379. doi: 10.1371/journal.pgen.1004379
- Mathur, S., & DeWoody, J. A. (2021). Genetic load has potential in large populations but is realized in small inbred populations. *Evolutionary Applications* . doi: 10.1111/eva.13216
- McCluskey, E. M., & Bender, D. (2015). Genetic structure of western massasauga rattlesnakes (*Sistrurus catenatus tergeminus*). *Journal of Herpetology* , 49 (3), 343-348. doi: 10.1670/14-016
- 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
- Ochoa, A., Broe, M., Moriarty Lemmon, E., Lemmon, A. R., Rokyta, D. R., & Gibbs, H. L. (2020). Drift, selection and adaptive variation in small populations of a threatened rattlesnake. *Molecular Ecology* , 29 (14), 2612-2625. doi: 10.1111/mec.15517
- Pemberton, T. J., Absher, D., Feldman, M. W., Myers, R. M., Rosenberg, N. A., & Li, J. Z. (2012). Genomic patterns of homozygosity in worldwide human populations. *The American Journal of Human Genetics* , 91 (2), 275-292. doi: 10.1016/j.ajhg.2012.06.014

- Perrier, C., Ferchaud, A. L., Sirois, P., Thibault, I., & Bernatchez, L. (2017). Do genetic drift and accumulation of deleterious mutations preclude adaptation? Empirical investigation using RADseq in a northern lacustrine fish. *Molecular Ecology* , 26 (22), 6317-6335. doi: 10.1111/mec.14361
- Pielou, E. C. (1992). *After the ice age: the return of life to glaciated North America* . Chicago, IL: University of Chicago Press.
- Pool, J. E., & Nielsen, R. (2009). Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics* ,181 (2), 711-719. doi:10.1534/genetics.108.098095
- R Core Team (2020). *R: A language and environment for statistical computing* . Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>
- Ralls, K., Sunnucks, P., Lacy, R. C., & Frankham, R. (2020). Genetic rescue: a critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful genetic variation. *Biological Conservation* , 251 , 108784. doi: 10.1016/j.biocon.2020.108784
- Renaud, G., Hanghoj, K., Korneliussen, T. S., Willerslev, E., & Orlando, L. (2019). Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples. *Genetics* ,212 (3), 587-614. doi: 10.1534/genetics.119.302057
- Renaut, S., & Rieseberg, L. H. (2015). The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other Compositae crops. *Molecular Biology and Evolution* , 32 (9), 2273-2283. doi: 10.1093/molbev/msv106
- Robinson, J. A., Ortega-Del Vecchyo, D., Fan, Z., Kim, B. Y., vonHoldt, B. M., Marsden, C. D., ... Wayne, R. K. (2016). Genomic flatlining in the endangered island fox. *Current Biology* , 26 (9), 1183-1189. doi: 10.1016/j.cub.2016.02.062
- Robinson, J. A., Raikonen, J., Vucetich, L. M., Vucetich, J. A., Peterson, R. O., Lohmueller, K. E., & Wayne, R. K. (2019). Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Science Advances* , 5 (5), eaau0757. doi: 10.1126/sciadv.aau
- Ryberg, W. A., Harvey, J. A., Blick, A., Hibbitts, T. J., & Voelker, G. (2015). Genetic structure is inconsistent with subspecies designations in the western massasauga *Sistrurus tergeminus* . *Journal of Fish and Wildlife Management* , 6 (2), 350-359. doi: 10.3996/122014-JFWM-093
- Saremi, N. F., Supple, M. A., Byrne, A., Cahill, J. A., Coutinho, L. L., Dalen, L., ... Shapiro, B. (2019). Puma genomes from North and South America provide insights into the genomic consequences of inbreeding. *Nature Communications* , 10 (1), 4769. doi: 10.1038/s41467-019-12741-1
- Schild, D. R., Card, D. C., Hales, N. R., Perry, B. W., Pasquesi, G. M., Blackmon, ... Castoe, T. A. (2019). The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome Research* , 29 (4), 590-601. doi:10.1101/gr.240952.118
- Schild, D. R., Pasquesi, G. I. M., Perry, B. W., Adams, R. H., Nikolakis, Z. L., Westfall, A. K., ... Castoe, T. A. (2020). Snake recombination landscapes are concentrated in functional regions despite PRDM9. *Molecular Biology and Evolution* , 37 (5), 1272-1294. doi: 10.1093/molbev/msaa003
- Schmidt, K. P. (1938). Herpetological evidence for the post-glacial extension of the steppe in North America. *Ecology* , 19, 396-407.
- Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* ,31 (19), 3210-3212. doi: 10.1093/bioinformatics/btv351
- Smit, A. F., Hubley, R., & Green, P. (1996-2010). *RepeatMasker Open-3.0* . <http://www.repeatmasker.org>

- Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., Soltis, P. S. (2006). Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* , 15 (14), 4261-4293. doi: 10.1111/j.1365-294X.2006.03061.x
- Sovic, M. G., Fries, A. C., & Gibbs, H. L. (2016). Origin of a cryptic lineage in a threatened reptile through isolation and historical hybridization. *Heredity* , 117 (5), 358–366. doi: 10.1038/hdy.2016.56
- Sovic, M., Fries, A., Martin, S. A., & Gibbs, H. L. (2019). Genetic signatures of small effective population sizes and demographic declines in an endangered rattlesnake, *Sistrurus catenatus* .*Evolutionary Applications* , 12 (4), 664–678. doi: 10.1111/eva.12731
- Stebbins, R.C. (1966). A Field Guide to Western Reptiles and Amphibians. Boston, MA: Houghton Mifflin Co.
- Szymanski, J., Pollack, C., Ragan, L., Redmer, M., Clemency, L., Voorhies, K., & JaKa, J. (2016). *Species status assessment for the eastern massasauga rattlesnake (Sistrurus catenatus)* . Fort Snelling, MN: U.S. Fish and Wildlife Service, Region 3.
- U.S. Fish and Wildlife Service (2016). Endangered and threatened wildlife and plants; threatened species status for the eastern massasauga rattlesnake. *Federal Register* , 81 (190), 67193–67214.
- van der Valk, T., de Manuel, M., Marques-Bonet, T., & Guschanski, K. (2019). Estimates of genetic load in small populations suggest extensive purging of deleterious alleles. *BioRxiv* , 696831. doi: 10.1101/696831
- van Oosterhout, C. (2020). Mutation load is the spectre of species conservation. *Nature Ecology & Evolution* , 4(8), 1004-1006. doi:10.1038/s41559-020-1204-8

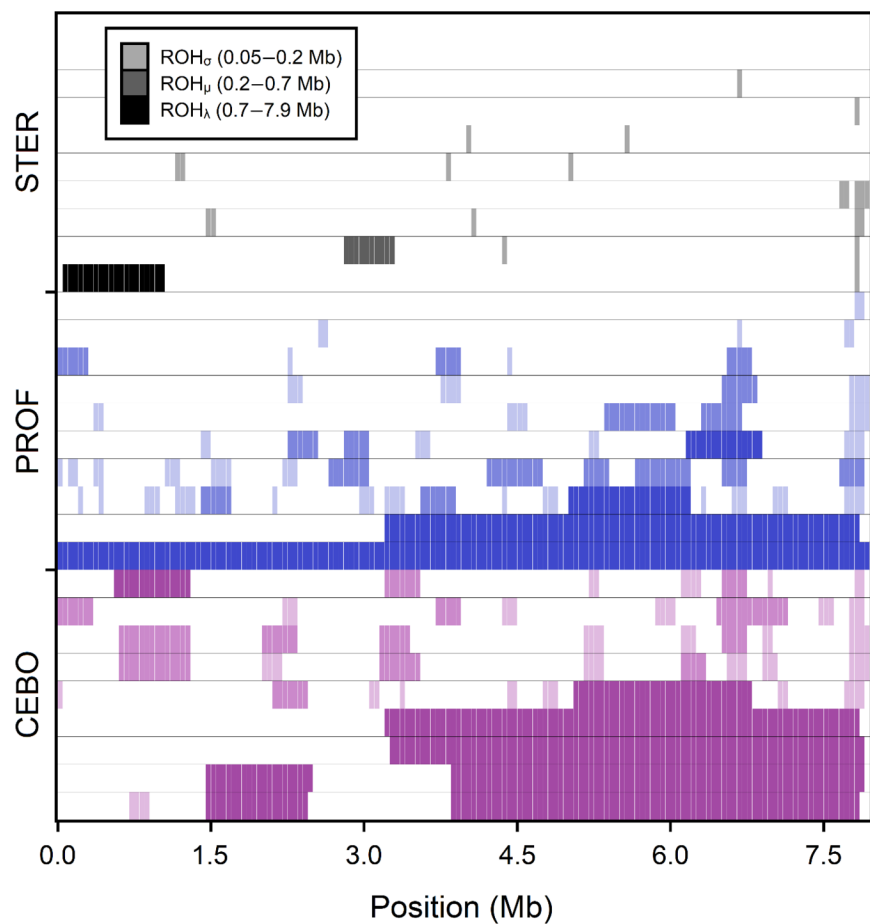


FIGURE 1 Distribution of short ($ROH_{\sigma} = 0.05\text{--}0.2\text{ Mb}$), medium ($ROH_{\mu} = 0.2\text{--}0.7\text{ Mb}$), and long ($ROH_{\lambda} = 0.7\text{--}7.9\text{ Mb}$) runs of homozygosity along the longest scaffold analyzed ($\sim 7.9\text{ Mb}$) for individuals (horizontal bars) from the reference *S. tergestinus* population (STER—black) and two *S. catenatus* populations (PROF—blue ; CEBO—purple) with low and high levels of inbreeding. ROH size classes are indicated by gradual color shadings (see box legend) for each population

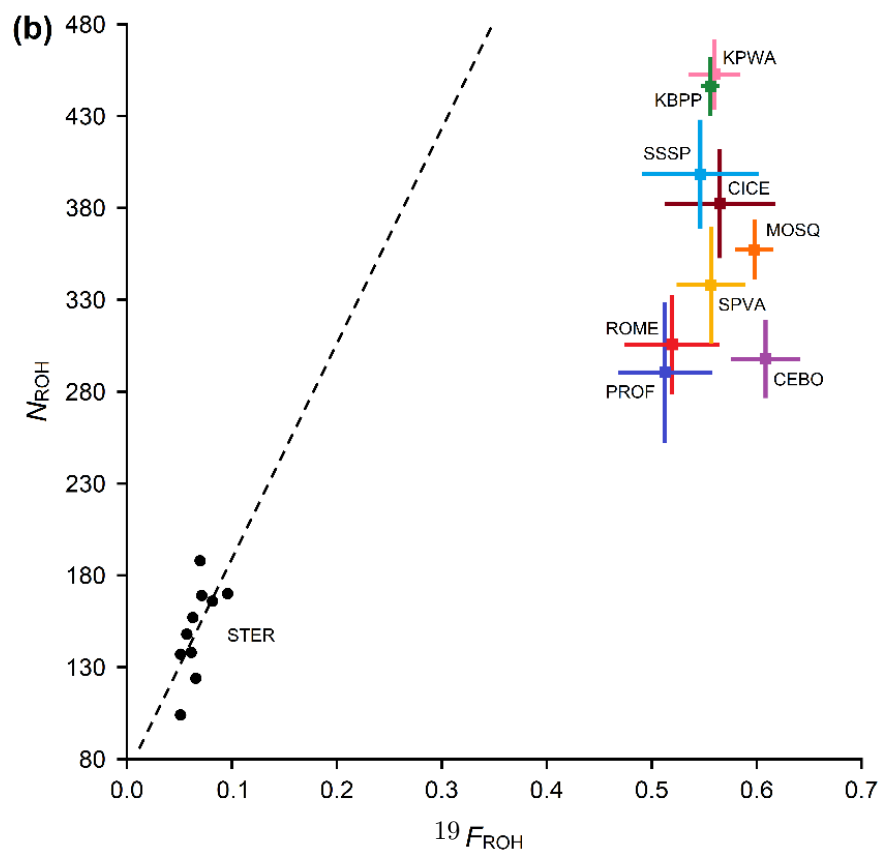
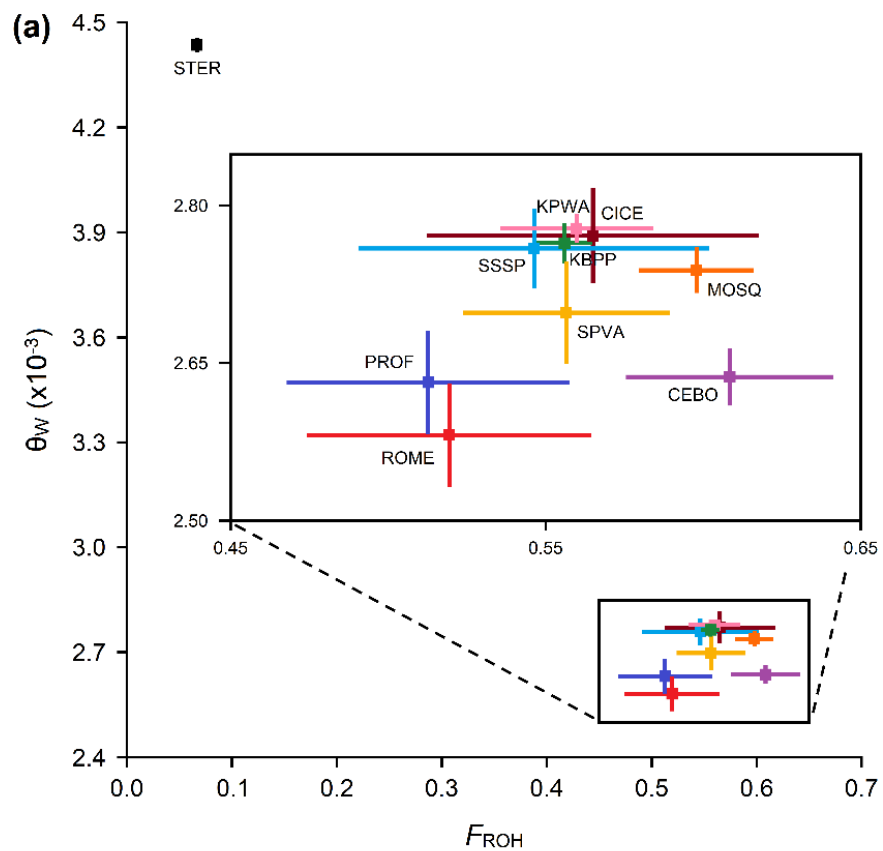


FIGURE 2 Scatter plots of (a) Θ_W and (b) N_{ROH} with respect to F_{ROH} for *S. tergeminus* (STER) and *S. catenatus* (SSSP, SPVA, CEBO, PROF, KPWA, ROME, MOSQ, CICE, KBPP) populations. Vertical and horizontal bars reflect standard errors from the estimated mean (squares) across individuals from the same population. For the (b) plot, each STER individual is represented by a black circle; the dotted line represents the association between N_{ROH} and F_{ROH} for these individuals after performing a linear regression analysis

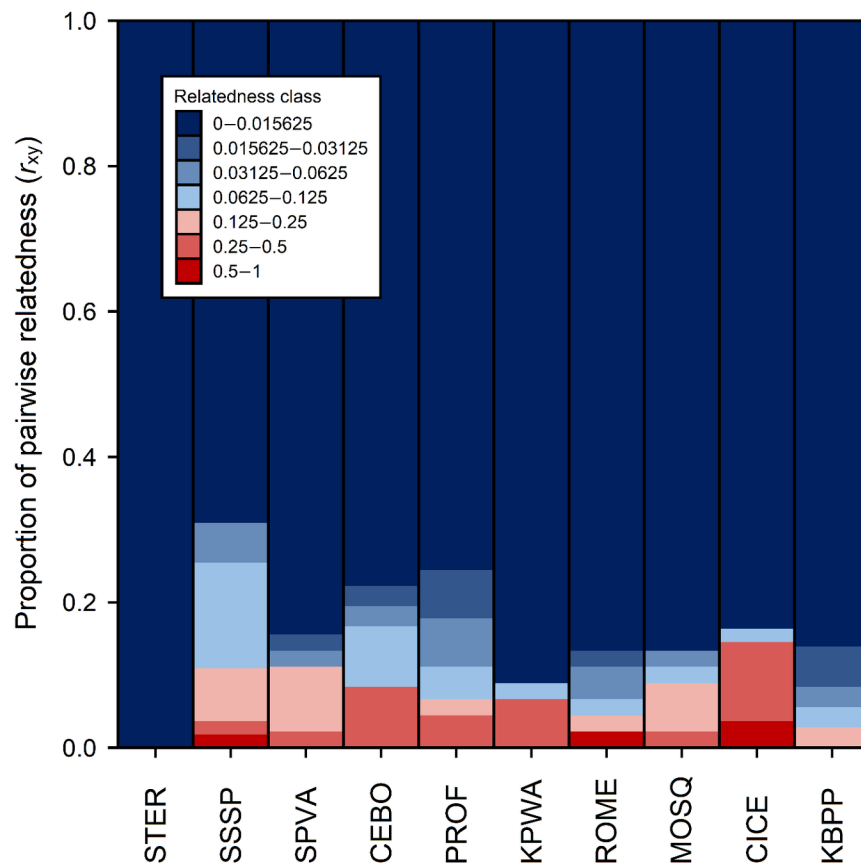


FIGURE 3 Relatedness values (r_{xy}) for pairs of individuals within populations of *S. tergeminus* (STER) and *S. catenatus* (SSSP, SPVA, CEBO, PROF, KPWA, ROME, MOSQ, CICE, KBPP). Colors inside each vertical bar represent the proportion of all pairwise comparisons within a single population that correspond to a particular relatedness class (see box legend)

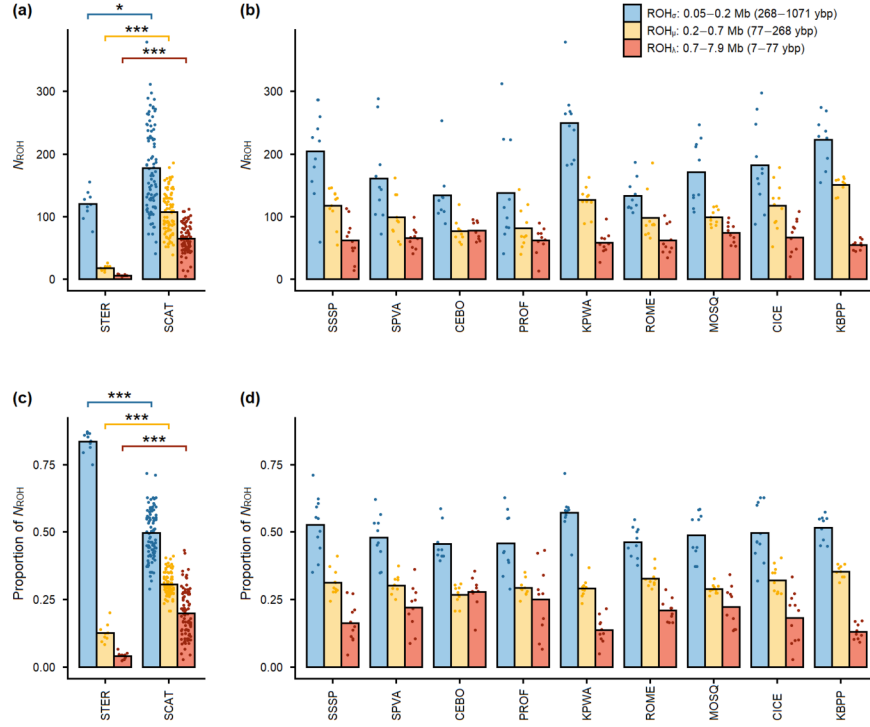


FIGURE 4 Mean N_{ROH} for short ($ROH_{\sigma} = 0.05-0.2$ Mb; blue bars), medium ($ROH_{\mu} = 0.2-0.7$ Mb; yellow bars), and long ($ROH_{\lambda} = 0.7-7.9$ Mb; red bars) ROHs across individuals (circles) from the same (a) species (STER = *S. tergestinus* ; SCAT = *S. catenatus*) and (b) *S. catenatus* population (SSSP, SPVA, CEBO, PROF, KPWA, ROME, MOSQ, CICE, KBPP). Mean relative N_{ROH} metrics for each (c) species and (d) *S. catenatus* population are also shown. Each ROH size class has an associated time of coalescence, which is included in the main legend as years before present (ybp). ROH size class statistical tests are shown at the species level, where *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, and ns = nonsignificant. Statistical tests between *S. catenatus* populations are shown in Figure S5

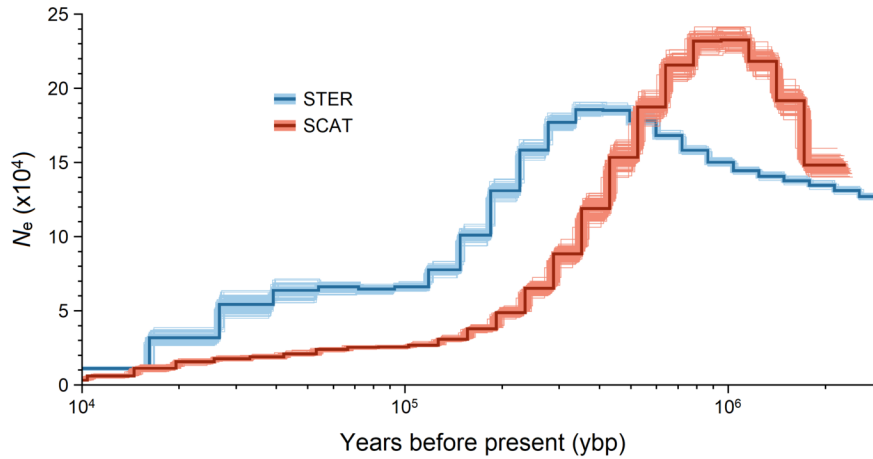


FIGURE 5 Historical changes in long-term effective population sizes (N_e) of *S. tergestinus* (dark blue lines) and *S. catenatus* (dark red lines) based on coalescent analyses. Light blue and light red lines reflect

estimates across 100 bootstrap replicates for each species, respectively

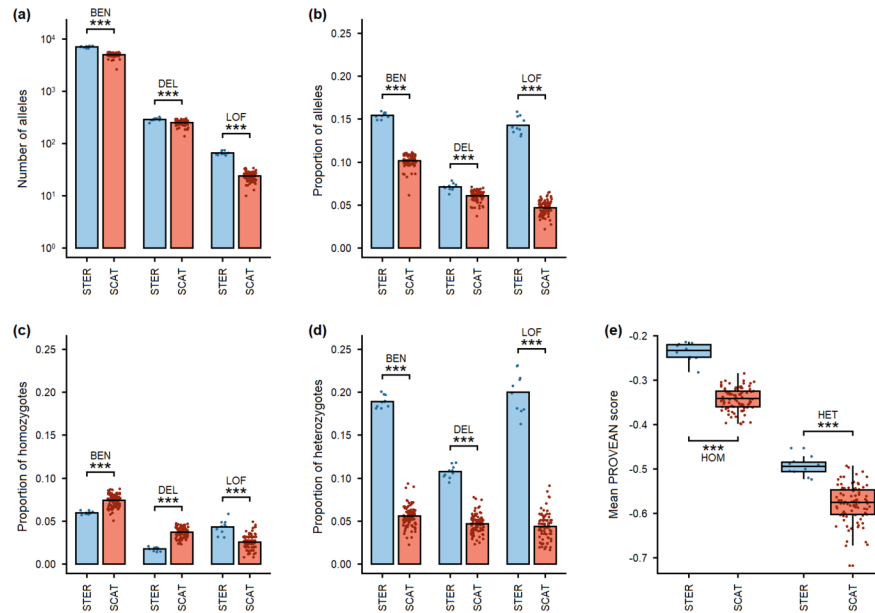


FIGURE 6 Interspecific distribution of benign (BEN), deleterious (DEL), and loss of function (LOF) mutation alleles and genotypes for *S. tergestinus* (STER; blue circles) and *S. catenatus* (SCAT; red circles) individuals across 3,800 conserved genes. Analyses based on Species level mutations (see Methods). (a) Mean number of counts for each allele class. Mean relative proportions of (b) alleles, (c) homozygotes, and (d) heterozygotes for each class site are also presented. (e) Boxplots represent the distribution of mean PROVEAN scores across alleles present in homozygous (HOM) and heterozygous (HET) state (LOF alleles were given a default value of -12). *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; ns = nonsignificant

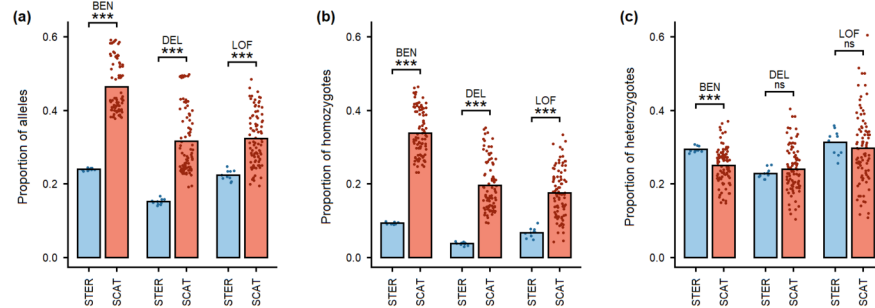


FIGURE 7 Intrapopulation distribution of benign (BEN), deleterious (DEL), and loss of function (LOF) alleles and genotypes for *S. tergestinus* (STER; blue circles) and *S. catenatus* (SCAT; red circles) individuals across 3,800 conserved genes. Analyses based on Population level mutations (see Methods). Mean relative proportions of (a) alleles, (b) homozygotes, and (c) heterozygotes across sites for each class site are represented as bars in each plot. *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; ns = nonsignificant

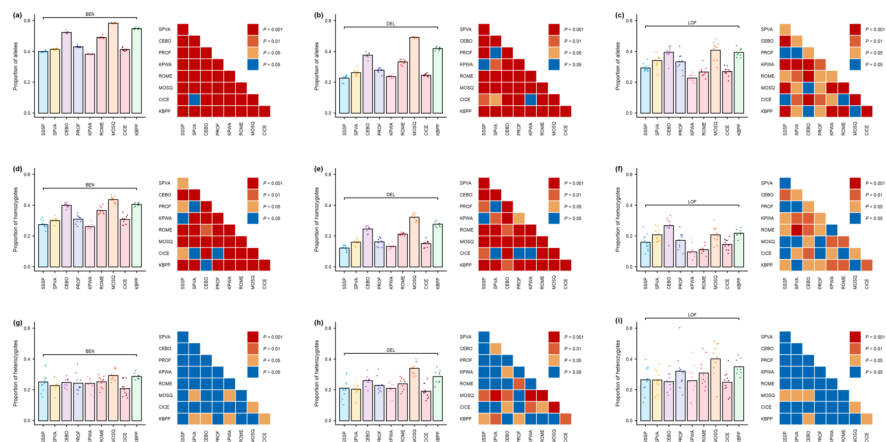


FIGURE 8 Intrapopulation distribution of benign (BEN), deleterious (DEL), and loss of function (LOF) (a-c) alleles and (d-f) homozygous and (g-i) heterozygous genotypes across *S. catenatus* populations. Analyses based on Population level mutations. Bars reflect means and circles individuals. Pairwise comparisons and statistical differences between populations are indicated as colored squares next to each plot