Posted on 10 Mar 2025 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.174160442.25339601/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

Genomic evaluation of assisted gene flow options in an endangered rattlesnake

Samarth Mathur[1] and H. Lisle Gibbs[2]

[1]Ohio State Univ
[2]Ohio State University

March 10, 2025

Abstract

Assisted gene flow is used to counteract genetic erosion in small populations of endangered species, yet an evaluation of genetic compatibility of potential donor populations and recipient populations is rare. We developed new metrics for assessing genetic impact of genetic augmentation based on genotype identity of functional variants between donor and recipient genomes and used these metrics to evaluate options for assisted gene flow in Eastern Massasauaga rattlesnake (Sistrutus catenatus) populations in Ohio, USA. We used putatively deleterious variants, and genetic variants likely under positive selection (termed "adaptive" variants) as the two components of functional variation. For potential donor and recipient populations, we estimated three key aspects of genetic compatibility – (a) introduction of novel variants, (b) masking or unmasking of existing deleterious variants, and (c) potential for outbreeding depression through disruption of local adaptation. The main impact of augmentation from diverse donor populations was to introduce novel deleterious variants and to a lesser extent novel adaptive variants into each recipient population. Both donor populations had a similar minor impact in terms of masking existing deleterious variants. Finally, only ˜ 7% of adaptive variants show evidence for local adaptation arguing that the negative effect of outbreeding depression would be small. These results draw attention to the importance of considering the potential impact of both deleterious and adaptive genetic variants in augmentation efforts and suggest that in case of these endangered rattlesnakes, the net effect of proposed assisted gene flow may lead to an increase in absolute levels of mutation load.

Genomic evaluation of assisted gene flow options in an endangered rattlesnake.

Samarth Mathur[1,2*], and H. Lisle Gibbs[1,2*]

[1]*Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, OH, USA*

[2]*Ohio Biodiversity Conservation Partnership, The Ohio State University, Columbus, OH, USA*

* Corresponding author: H. Lisle Gibbs.

Email: gibbs.128@osu.edu

ORCIDs:

Samarth Mathur: 0000-0002-6446-5718

H. Lisle Gibbs: 0000-0001-7461-3393

Keywords: genetic rescue; assisted gene flow; functional genomic variation; conservation genomics; *Sistrurus catenatus*.

Abstract

Assisted gene flow is used to counteract genetic erosion in small populations of endangered species, yet an evaluation of genetic compatibility of potential donor populations and recipient populations is rare. We developed new metrics for assessing genetic impact of genetic augmentation based on genotype identity of functional variants between donor and recipient genomes and used these metrics to evaluate options for assisted gene flow in Eastern Massasauaga rattlesnake (*Sistrutus catenatus*) populations in Ohio, USA. We used putatively deleterious variants, and genetic variants likely under positive selection (termed "adaptive" variants) as the two components of functional variation. For potential donor and recipient populations, we estimated three key aspects of genetic compatibility – (a) introduction of novel variants, (b) masking or unmasking of existing deleterious variants, and (c) potential for outbreeding depression through disruption of local adaptation. The main impact of augmentation from diverse donor populations was to introduce novel deleterious variants and to a lesser extent novel adaptive variants into each recipient population. Both donor populations had a similar minor impact in terms of masking existing deleterious variants. Finally, only ~ 7% of adaptive variants show evidence for local adaptation arguing that the negative effect of outbreeding depression would be small. These results draw attention to the importance of considering the potential impact of both deleterious and adaptive genetic variants in augmentation efforts and suggest that in case of these endangered rattlesnakes, the net effect of proposed assisted gene flow may lead to an increase in absolute levels of mutation load.

Introduction

Threatened and endangered species often exist in small populations that suffer from high levels of genetic erosion which can negatively impact their future viability (Frankham et al., 2017; Leroy et al., 2017). Assisted gene flow represents an *in-situ* management strategy in which the genetic makeup of small recipient populations can be potentially improved by the introduction of individuals from larger and more genetically diverse populations. In theory, these introductions can mitigate the effects of high levels of genetic load and low levels of adaptive variation in recipient populations (Fitzpatrick et al., 2023; Frankham et al., 2011; Weeks et al., 2011) leading to the genetic rescue of recipient populations (Fitzpatrick et al., 2023; Whiteley et al., 2015). However, the positive effect of assisted gene flow is dependent on the genetic complementarity of donor and recipient populations in terms of whether the genetic input from donor individuals will minimize the negative impacts of genetic load while also increasing future viability by replenishing adaptive variation (Fitzpatrick & Funk, 2021). Thus, evaluating the genetic compatibility of potential donor and recipient populations is a key step in identifying potentially successful genetic rescue options for a given species.

When considering appropriate strategies for assisted gene flow, it is important to evaluate the positive or negative genetic impacts that source populations will potentially have through the introduction of novel deleterious mutations and/or adaptive genetic variants (Fitzpatrick & Funk, 2021). Positive impacts include masking of recessive deleterious variants (in heterozygotes) that occur at high frequencies in the small recipient population and the introduction of adaptive variation previously lost due to genetic drift. However, these need to be balanced against potential risks which include the introduction of new deleterious mutations leading to increased inbreeding depression and the introduction of variants that disrupt local adaptation leading to outbreeding depression (Edmands, 2017; Frankham et al., 2011). Until recently, the main options to identify suitable donor populations were to estimate population characteristics such as effective population size, levels of inbreeding, and gene flow or, use levels of neutral variation as a proxy for functional variation (Fitzpatrick & Funk, 2021). However, these approaches lack information on the identity and abundance of the specific genetic variants that determine the levels of genetic erosion in each population. This makes it difficult to determine key aspects of the genetic impact of potential donor individuals such as the number of novel mutations (positive or negative) that will be introduced, and whether the existing deleterious mutations being expressed in the recipient population due to inbreeding will be masked in heterozygotes. The recent availably of genomic information and methods for characterizing positive and negative functional variation at a genome-wide scale now makes it possible to indirectly assess the potential genetic impacts of specific donor populations on the functional genetic profile of recipient populations (e.g. Khan et al. (2021)). These developments are timely as more species stand to benefit from a conservation strategy of assisted migration for genetic rescue than have previously been considered (Fitzpatrick et al., 2023).

Here, we apply a genomic-based approach to evaluate assisted gene flow options for small populations by assessing allele-specific impacts on genetic load and levels of adaptive variation. We identify putatively deleterious and adaptive functional variation in the genomes of multiple populations of the endangered Eastern Massasauga rattlesnake as previously described in Mathur et al. (2023) and use new metrics to assess the genomic compatibility between potential donor and recipient populations. This genomic compatibility is based on comparing the impacts of

translocating individuals from specific donor populations to specific recipient populations in terms of the absolute number of new deleterious and adaptive variants introduced, the masking of existing deleterious mutations, and the levels of local adaptation present in positively selected variation.

The Federally threatened Eastern Massasauga Rattlesnake (*Sistrurus catenatus*) currently exists as a series of small, isolated populations across its range in the US and Canada (Fig. 1A). Recent work has shown that small populations of these snakes show high frequencies of deleterious mutations (Ochoa & Gibbs, 2021) and reduced levels of adaptive variation (Ochoa et al., 2020). In Ohio, there are three populations which have exceptionally small recent effective population sizes (Mathur et al., 2023) and are potentially on the verge of extirpation (CEBO, SPVY, PRDF; Fig. 1B). In contrast, there are several large populations which are sufficiently genetically diverse that they could serve as potential donor populations (KLDR and WLRD). The first steps in conservation plan that includes assisted gene flow have been taken for one of these populations (G. Lipps, pers. comm.) yet no evaluation of the potential genetic consequences this action is available. Our goal is to provide a genetic evaluation of various translocation scenarios that involve these populations (Fig. 1C) which can then be considered with other relevant political, ecological, and logistic considerations for the implementation and interpretation of the assisted gene flow efforts.

Materials and Methods

Samples, sequencing, and sequence processing.

In this study, we analyzed whole genome sequences from 152 *S. catenatus* individuals from 14 different study sites in the U.S.A and Canada (Fig. 1A; Data S1). Our focus was on the isolated populations in Ohio (Fig. 1B) especially to understand the consequences of assisted gene flow of moving snakes from larger and more genetically diverse ("donor") populations into the smaller and less diverse ("recipient") populations. We compared all possible donor-recipient scenarios (Fig.1C) of the two donor populations (KLDR and WLRD) and the three recipient populations (CEBO, PRDF, and SPVY) to identify which donor population is more suitable and which recipient population would benefit the most from assisted gene flow.

We combined publicly available *S. catenatus* genome sequences (NCBI BioProject Accession Nos. PRJNA750087 and PRJNA975611) and new individuals sequenced for this study. More details about DNA extraction and Illumina® paired-end (PE) 150-bp reads sequencing are provided in Supplementary Methods. All sequences were mapped to the *S. catenatus* Hi-C reference genome (GenBank Accession No. JASPKD000000000.1) using the germline pipeline available in NVIDIA Clara Parabricks v3.5 which is compatible with GATK4 best practices ((Van der Auwera et al., 2018); Supplementary Methods). The depth of coverage per individual was 15.3 ± 3.3X (mean ± SD) with 77.8 ± 9.6 % of the reference genome covered at least 10X depth (Data S1).

Genotype calling, variant filtration, and variant annotation.

We identified SNP variants using NVIDIA Clara Parabricks v3.5 germline pipeline and performed joint genotyping on all 152 samples using GATK v4.1.2.0 (((McKenna et al., 2010); Supplementary Methods). We used "QD < 2.0;

109 MQ < 20.0;|MQRankSum| > 3.0; |ReadPosRankSum| > 3.0; SOR > 5.0" to filter low-quality SNPs (Grossen et al.,
110 2020). Only biallelic SNPs were retained and SNPs with genotypes missing in >20% of samples, singletons and
111 private doubletons were removed using VCFTools v0.1.16 (Danecek et al., 2011) resulting in the final set of
112 18,923,075 SNPs across the genome. We functionally annotated genome-wide SNPs using SnpEff v.4.3 (Cingolani
113 et al., 2014) and classified substitutions based on their regions (exons, introns, intergenic) and functional class
114 (missense, nonsense, and silent). After filtering out low confidence annotations, we identified 68,041 missense,
115 1,221 nonsense (hereafter referred to as "loss of function (LoF)"), and 93,195 silent substitutions in our dataset.
116
117 *Identifying functional variants and estimating levels of functional genomic variation.*
118 To identify missense mutations whose amino acid change can have a negative impact on overall protein function, we
119 used the PROVEAN (Protein Variation Effect Analyzer) framework (de Brevern et al., 2012) as implemented in
120 ePAT (extended PROVEAN annotation tool) software (Ito, Yoshitake, & Iwata, 2021). PROVEAN predicts the
121 impact of an amino acid substitution on the biological function of a protein by first aligning the protein sequences
122 from related species using BLAST and then providing a quantitative measure (i.e. a PROVEAN score) of how
123 phylogenetically conserved a missense allele is among homologous protein sequences. ePAT allows estimating
124 PROVEAN scores for multiple variants simultaneously in a single step using batch processing. If the PROVEAN
125 score is < -2.5, the missense mutation is considered as "damaging" indicating a possibly negative impact of amino
126 acid change whereas, missense mutation with score > -2.5 were considered "neutral" with no effect of animo acid
127 change on the protein function (Ito et al., 2021). Out of 68,041 missense mutations within our dataset, only 6,623
128 mutations (~10%) were classified as damaging. We considered both damaging missense and LoF substitutions as
129 deleterious mutations because both these types of substitutions can negatively impact individual fitness (possibly
130 sublethal and lethal respectively) and thus contribute to the overall mutation load. We estimated mutation load of
131 deleterious mutations at an individual level by partitioning the two components by the zygosity of deleterious
132 alleles: Masked load (L_{mask}) and realized load ($L_{realized}$). For each individual genome, L_{mask} is the total number of
133 deleterious SNPs that are present as heterozygotes whereas $L_{realized}$ is the total number of SNPs that are homozygous
134 for the deleterious allele (Bertorelle et al., 2022; Mathur & DeWoody, 2021). We took the mean individual load to
135 estimate population levels of mutation load.
136 To identify putative adaptive variation, we first designated missense substitutions in protein-coding regions
137 of genes previously identified as under positive selection in *S. catenatus* based on the Direction of Selection ("DoS")
138 values ((Stoletzki & Eyre-Walker, 2010); see Mathur et al. (2023) for details) (Supplementary Methods). Here, we
139 refer to these missense substitutions as positively selected variants to better represent the uncertainty about whether
140 they do in fact represent adaptive variation (see Mathur et al. 2023 for discussion).
141
142 *Assessing the genomic compatibility between donor and recipient populations.*
143 Mean levels of mutational load and positively selected variation within each population fail to account for the
144 genotype identity of these substitutions between a pair of genomes and their overall abundance in donor and
145 recipient populations. This is important because evaluating the impact of a potential donor on a recipient population

includes identifying whether novel mutations (positive or negative) will be introduced and if a deleterious mutation that exists as homozygote (i.e. realized) in a recipient individual will be masked as heterozygote in the next (F1) generation (Khan et al., 2021). Genomic information provides this added nuance by comparing genotypic differences in various donor and recipient genomes at positive and negative functional loci.

For this purpose, we developed two new indices that compares the genomic compatibility between a pair of donor and recipient populations based on the genotypes of different functional mutations as outlined in Figure 2 : (a) mutations added (M_{add}) i.e., the number for novel negative or positive mutations added by a donor to the recipient population, and (b) masking potential (P_{mask}) i.e., the net effect of the number of exposed negative variants in recipient population that can be masked by the donor population and the number of already masked negative variants in recipient population that can be unmasked or exposed by the donor population in the next generation (Fig. 2B). Exposed mutations can be masked at different probabilities (P=1, P=0.5) in the F1 generation depending on donor genotypes and similarly, masked mutations can be exposed at different probabilities as well (P=0.5, P= 0.25; Fig. 2C), assuming Mendelian inheritance. P_{mask} estimates a cumulative score of the probabilities of a donor genome masking homozygous alleles ($P_{mask} > 0$) and unmasking heterozygote alleles ($P_{mask} < 0$) when crossed with a recipient genome, considering the probability weights of masked and unmasked mutations. We then applied these metrics to evaluate the potential consequences of assisted gene flow for the different donor-recipient scenarios as described in Figure 1C.

Evidence for local adaptation driving genetic differentiation between Ohio populations.

The introduction of a novel adaptive allele to a recipient population is generally assumed to have a positive impact on future population viability (Khan et al., 2021). However, if local adaption is high among populations, the introduction of novel adaptive alleles could potentially have a negative impact through the disruption locally adapted gene complexes or the introduction of maladapted alleles leading to outbreeding depression (Brooks et al., 2024; Edmands, 2007). Identifying local adaptation as a selective pressure leading to adaptive differences between isolated populations is challenging (Hoban et al., 2016). However, we can assess the possible role of local adaption as a driver of population differences by comparing the differentiation in positive variants against background differentiation in neutral loci (which reflect nonadaptive processes alone). Functional variants that are positively selected in *S. catenatus* and also show greater levels of divergence than that observed at neutral loci are candidates for functional loci that may also have contributed to local adaptation in different populations. The relative abundance of such variants represents an estimate of the importance of local adaptation as an evolutionary process in these populations (Margres et al., 2017).

We performed an outlier test by estimating the F_{ST} value of each adaptive locus among all Ohio populations and comparing it to the F_{ST} value of loci in the gene desert regions (i.e. regions that lie in intergenic regions outside of any possible physical linkage to any protein coding regions; Supplementary Methods). Loci evolving via local adaptation among these populations were identified as having F_{ST} values that were greater than the 95th percentile of the neutral locus F_{ST} distribution (Margres et al., 2017; Ochoa et al., 2020).

182 **Results**

183 *Levels of functional diversity.*

184 We first compared population levels estimates of negative and positive functional diversity among all *S. catenatus*
185 populations to get a baseline estimate of different components of genomic diversity. We identified that differences in
186 genetic erosion are present among different Ohio populations (Fig. S1; Table S1). Mean populations levels of
187 realized genetic load for both high impact (LoF; N = 1,221) and lower impact (damaging missense based on
188 PROVEAN scores; N = 6,623) deleterious mutations are lower in the two potential donor populations compared to
189 each of the three recipient populations (Fig. S1) and snakes from at least one potential recipient populations (CEBO)
190 had high mean realized load (Table S1). This implies that the three potential recipient populations can benefit from
191 assisted gene flow via reduction in the negative effects of inbreeding depression. Comparisons of mean adaptive
192 variation are similar among the different populations (Table S1). The KLDR donor population shows higher mean
193 levels of diversity in positive variants than mean values for any of the recipient populations. In contrast, the WLRD
194 population has level of diversity in positive variants similar to (CEBO) or less than (PRDF and SPVY) the mean
195 values for the potential recipient populations (Table S1). Based on these results KLDR individuals are predicted to
196 provide an advantage over WLRD individuals in terms of potentially increasing adaptive diversity in recipient
197 populations. However, these comparisons fail to consider the impact of the identity of specific genetic variants in
198 donor and recipient populations.

199

200 *Introduction of novel mutations.*

201 We compared the effect of assisted gene flow of moving snakes from donor (KLDR and WLRD) to recipient
202 populations (CEBO, PRDF, SPVY) using the different genetic compatibility metrics (Fig. 2). First, we estimated the
203 assisted gene flow impact in terms of introducing novel functional variants (M_{add}), both deleterious and positive. Our
204 analysis shows that introducing individuals from either donor population would add novel positive and negative
205 functional variants which can impact the genetic fitness of the recipient populations (Table S2) and that the number
206 of each type of mutation introduced is positively correlated (Fig. 3). A significant positive relationship between the
207 number of novel positive mutations ($M_{add\text{-}Adaptive}$) and deleterious mutations (both LoF $M_{add\text{-}LoF}$ and damaging
208 missense $M_{add\text{-}damaging}$) added by donor genomes means that the donor population that would add more novel
209 adaptive variants into the recipient population would also lead to the introduction of more new deleterious mutations
210 (Fig. 3). However, in terms of absolute numbers most of the novel variants added (~80%) will be deleterious
211 mutations in the form of damaging missense variants (Fig. 3B and D) with a smaller number of LoF mutations also
212 added (Fig. 3A and C; Table S2). Thus, assisted gene flow will have both negative and positive impacts but that in
213 terms of the numbers of negative variants added, the number of low-impact variants introduced would be about 5-
214 fold higher than the number of high-impact deleterious mutations.

215 In terms of relative impacts of the two donors, an average KLDR snake would add more new LoF
216 mutations (100 – 115) into the three recipient populations relative to an average WLDR individual (90 – 104 new
217 mutations) and more lower impact missense mutations (536 – 578) relative to a WLRD individual (470 – 500 new
218 mutations; Table S2). Relative to total functional mutations already present in recipient genomes, using KLDR

219 snakes would result in 28-34% more novel LoF and 34-39% damaging missense mutations being introduced to the
220 recipient populations whereas WLRD would add 26-32% novel LoF and 31-36% damaging missense mutations
221 (Table S3). In contrast, KLDR individuals will add more positive variants (132 – 146; 32-36%) to recipient
222 populations than would an average WLDR individual (115 – 124; 29-33%) although the difference is small. Overall,
223 these results indicate that the impacts on functional genomic variation in donor populations of choosing one donor
224 population over the other are small in terms of adding novel mutations.
225
226 *Masking and unmasking of deleterious mutations, and overall masking potential.*
227 Overall masking potential (P_{mask}) of a donor population is dependent on how many deleterious alleles that are
228 exposed in homozygotes would get masked in heterozygotes and how many deleterious alleles that are already
229 masked would get exposed in the next generation (F1 generation). This metric accounts for the differences in
230 masking/unmasking probabilities between the genotype identity of each locus between the donor and recipient
231 genome (Fig. 2 and Supplementary Methods). Fig. S2 and Table S4 report the number of deleterious mutations that
232 are potentially masked (at P = 1 and 0.5; see Fig. 2C) and unmasked (at P = 0.5 and 0.25; Fig. 2C) for both LoF and
233 missense mutations. Overall, when considering the mutation load (masked and realized) which is already present in
234 each individual recipient genome, 27-34% of the exposed LoF and 34-42% of damaging missense mutations will
235 likely be masked (at P =1) by KLDR genomes (Table S5). Similarly, only 14.1-14.3% of the masked LoF and 7.4-
236 8.1% damaging missense mutations are likely to get exposed (at P = 0.5) by introducing KLDR genomes.
237 Proportions were similar but slightly lower for WLRD donors (Table S5).
238 We used P_{mask} to summarize the net potential of an average donor genome to mask or unmask deleterious
239 mutations across the different probabilities described in Figure 2C (Fig. S3; Table S6). The results show that the
240 degree of masking varies markedly across potential recipient populations but is similar between potential donor
241 populations. For LoF mutations, the degree of masking is highest for the CEBO, intermediate for SPVY, and lowest
242 for the PRDF recipient population. There was a similar pattern for damaging missense mutations.
243
244 *Net impact of assisted gene flow.*
245 Fig. 4 summarizes the net impact of potential assisted gene flow by KLDR and WLDR individuals across recipient
246 populations in terms of changes in numbers of negative and positive genetic variants. When used as a donor, an
247 average individual from KLDR will introduce 139 ± 7 novel positive variants and 553 ± 22 novel damaging
248 missense and 108 ± 8 LoF variants. Additionally, when the degree of masking is considered (P=1), KLDR genomes
249 will mask pre-existing 28 ± 6 LoF and 109 ± 22 damaging missense mutations that are exposed in the recipient
250 populations. Values for donor individuals from WLRD are smaller but similar in relative magnitude: 120 ± 5 new
251 positive variants, 486 ± 15 damaging missense, and 99 ± 8 LoF mutations would be introduced; 31 ± 6 pre-existing
252 LoF and 122 ± 22 damaging missense mutations that are exposed would be masked. Overall, our mutation counting
253 approach indicates that the introduction of donor individuals from both source populations results in a > 4.5-fold
254 increase in the relative number of negative (high and low-impact) to positive genetic mutations in potential donor
255 populations. Most of these variants are low impact missense mutations. When only high impact LoF negative

256 mutations are considered, the numbers of negative versus positive mutations are similar with slightly more positive
257 variants being added.
258
259 *Impact of local adaptation*
260 Our outlier analysis suggests that relatively few variants putatively under positive selection in *S. catenatus* show
261 evidence for local adaptation among Ohio populations and so possible negative effects due to outbreeding
262 depression are likely small. Figure 5 shows distributions of F_{ST} values for neutral variants from "gene desert"
263 noncoding regions and for missense variants in coding regions of genes under strong positive selection (Mathur et
264 al., 2023) for Ohio populations. Only a small number of positive variants show elevated F_{ST} values consistent with
265 local adaptation - specifically, only 78 of 1053 total positive variants (7.4%) showed F_{ST} values greater than the 95%
266 cutoff based on the null distribution for neutral variants in these populations. These results suggest that local
267 adaption may only be responsible for differentiation in ~ 7% of the putative positive variants in our study and so the
268 potential for outbreeding depression because of assisted gene flow is small.
269

270 **Discussion**
271 *Potential genetic impacts of assisted gene flow*
272 Traditionally, the potential genetic impacts of population translocations into small, inbred populations of endangered
273 species have relied on indirect measures of key population genetic parameters (Fitzpatrick & Funk, 2021).
274 Evaluation of specific assisted gene flow options using genomic information enables more direct analyses of genetic
275 consequences of these options in terms of specific 'good' and 'bad' functional variants (Fitzpatrick & Funk, 2021;
276 Hedrick, Allendorf, & Waples, 2013; Khan et al., 2021). This leads to quantitative estimates of the relative
277 magnitude of specific genomic contributions that potentially affect short- and long-term population survival.
278 Our newly developed genomic compatibility approach based on the comparison of functional genotypes
279 between donor and recipient genomes (Fig. 2) indicates that the net effect of proposed assisted gene flow options for
280 Eastern Massasauga rattlesnake populations in Ohio is: (a) a large increase in the absolute number of novel
281 deleterious mutations introduced into recipient populations, (b) a relatively smaller increase in the number of novel
282 putatively adaptive variants introduced, and (c) minor impacts of the masking or unmasking of deleterious alleles in
283 the first generation post augmentation.
284 The impacts on the levels of functional genetic variation in the recipient genomes would be substantial.
285 Based on the average number of variants present genome-wide in an recipient individual (positive = 274 ± 38;
286 damaging missense = 978 ± 133, LoF = 234 ± 28), KLDR donors will lead to a 33.7 +/- 2.3% increase in the
287 numbers of positive variants, a 36.2 +/- 2.5% increase in damaging missense mutations, and a 31.6 +/- 2.9% increase
288 in LoF mutations in recipient individuals (Fig. 4). However, the absolute number of mutations present is more
289 relevant to genome-wide fitness effects and, a > 4.5-fold increase in the absolute number of negative versus positive
290 mutations in recipient populations suggests that, based on absolute numbers alone the net impact of augmentation

291 would be to increase the mutation load in recipient populations. These insights would not be possible without a
292 locus-by-locus assessment of the genotypic identity of functional variants and their possible fate upon augmentation.
293 Currently, we have no way to evaluate the negative and positive impacts of specific mutations on survival
294 and reproduction in these snakes and so our inferences are based on numbers of mutations alone. Yet these fitness
295 differences clearly exist as incorporated in empirical distributions of fitness effects for model organisms (Eyre-
296 Walker & Keightley, 2007). As such, we cannot exclude the possibility that the average fitness effect of an adaptive
297 variant is higher than the negative fitness effect of a deleterious mutations. This could lead to the situation whereby
298 assisted gene flow may lead to a net fitness benefit to recipient populations despite the lower number of adaptive
299 variants introduced relative to the number of deleterious variants. Many empirical studies of assisted gene flow to
300 date show an increase in the viability of recipient populations (Clarke, Smith, & Cullingham, 2024; Ralls et al.,
301 2020) which is consistent with a broad positive impact of this management action in small populations through
302 either the addition of new adaptive variants or the masking of existing deleterious mutations in heterozygotes. It is
303 also consistent with an overwhelming benefit of demographic effects alone through the increased buffering of small
304 recipient populations against stochastic events and Allele effects due to increase in population sizes (Hufbauer et al.,
305 2015).
306 For negative mutations we attempted to add context by classifying variants into high impact (LoF) and
307 lower impact (damaging missense) mutations, but these classifications are not linked to demographic data which is
308 difficult to obtain in these cryptic species. An alternative for assessing the fitness effect of deleterious alleles would
309 be to use an approach such as the Combined Annotation-Dependent Depletion (CADD) framework ((Rentzsch et al.,
310 2019) – see for example (Speak et al., 2024)) but this requires information on the phenotypic impact of specific
311 mutations for a related model organism which doesn't currently exit for a reptile species.
312 Estimating plausible positive or negative fitness effects for substitutions that are computationally inferred is
313 a general problem in using genomic-based approaches for assessing functional variation (Galtier & Duret, 2007;
314 Kardos, Keller, & Funk, 2024; Orr, 2010; Robinson et al., 2023). As discussed in Mathur et al. (2023), the impact of
315 a given missense mutation considered in this study could be influenced by a number of factors. These include
316 compensatory mutations which could modify the effect of substitutions our scheme classifies and factors such as
317 dominance, epistasis, pleiotropy, and purging may also complicate the relationship between mutation load, adaptive
318 diversity and fitness (Grossen et al., 2020). Geographic variation in habitat can also alter the impact of deleterious or
319 adaptive mutations among populations (Mee & Yeaman, 2019) although the impacts of local adaption on our
320 positive variants seems limited (Fig. 5). Finally, it also seems unlikely that our approach detects all adaptive variants
321 that could arise under different types of selection such as frequency dependent selection because we have focused on
322 measuring variants that increase levels of functional polymorphism in populations (Mathur et al., 2023).
323 Our quantitative approach addresses a long-standing question about the relative importance of the
324 introduction of novel deleterious alleles compared to masking of pre-existing deleterious variants (Fitzpatrick &
325 Funk, 2021; Khan et al., 2021). Our results demonstrate that, for this species, masking also represents a moderate
326 advantage of assisted gene flow resulting in the neutralizing of 26-36% of LoF mutations and 34-46% of damaging
327 missense variants (Table S5) present in small populations of these snakes. Based on numbers alone, introduction of

328 novel deleterious alleles has a greater impact on the genetic makeup of small populations and so should be of greater
329 focus when designing and assessing assisted gene flow plans for this species.
330 Finally, our analyses suggests that the potential for outbreeding depression is low in this set of populations
331 because relatively few of the positively selected variants in *S. catenatus* that we have analyzed show evidence for
332 local adaptation in Ohio populations (Fig. 5) at least over the limited spatial scale that we conducted these analyses.
333 This is significant because reptiles in general show limited dispersal which can lead to fine-scale local adaptation
334 (Brooks et al., 2024). More generally, a commonly cited reason for not pursing assisted gene flow in endangered
335 species is the fear of disrupting local adaption in recipient populations (Bell et al., 2019; Ralls et al., 2020). Our
336 analyses used methods for assessing local adaptation over evolutionary timescales and our results suggested that
337 concern about outbreeding depression in these snakes is likely minimal. We recognize that detecting location
338 adaption using F_{ST} outlier tests has limitations, and the results presented here need to be confirmed with addition
339 analyses (Hoban et al., 2016). This inference would need to be reevaluated if translocations between most distant
340 populations were proposed since at a larger geographic scale, local adaptation may play a larger role as a process
341 impacting adaptive variation.
342 Our results add to the increasing recognition among conservation biologists that using donors from large
343 genetic diverse populations can have both positive and negative genetic impacts on recipient populations (Hedrick et
344 al., 2013; Khan et al., 2021). A long-held view is that large genetically diverse populations are good source
345 populations for assisted gene flow because of the increase in diversity that will occur in small, depleted populations
346 (e.g. (Fitzpatrick & Funk, 2021)). This reasoning makes the implicit assumption that adaptive variation will be the
347 primary form of functional variation that will be transferred. However, large genetically diverse populations also
348 contain high levels of potential inbreeding load in the form of deleterious variants hidden in heterozygous
349 individuals (Mathur & DeWoody, 2021) that are then exposed in small populations through interbreeding with
350 inbred recipient individuals (Kardos et al., 2021; Mathur et al., 2023). We suspect that these hidden variants are the
351 primary source of the new deleterious variants that will be transferred from donor to recipient populations in these
352 snakes.
353 Recognizing that large genetically diverse populations contain higher levels of both good and bad variation
354 is important when assessing assisted gene flow options in other endangered species. Our genomic compatibility
355 approach allows a quantitative assessment of the relative impacts of both components of functional variation
356 between donor and recipient populations and indicates that, at least in this species with this genetic architecture,
357 substantially more deleterious mutations would be introduced than would adaptive variants. Future extensions of this
358 work could include using individual-based simulations with estimates of Distribution of Fitness Effects (DFE) to
359 explore how the net impact of positive and negative mutations in donor individuals vary under different
360 demographic scenarios (e.g. Kyriazis et al. (2024)) and genetic backgrounds in recipient populations.
361
362 *Implications for assisted gene flow in snake populations*
363 Our results suggest that the net effect of proposed assisted gene flow for the three small Ohio populations of *S.*
364 *catenatus* will be an increase in the total number of mostly small effect deleterious mutations and a smaller increase

365 in the number of presumed adaptive variants. Based on variant counts alone, this suggests that in this specific
366 situation augmentation will have a negative impact on the genetic diversity present in these small populations.
367 However, this conclusion is qualified by the fact that there is a lack of difficult-to-obtain phenotypic data which is
368 necessary to evaluate the fitness impacts of specific functional mutations. These impacts are similar between the two
369 potential donor populations analyzed indicating that from a purely genetic perspective there is little basis for
370 choosing individuals from one population over the other for use as donors. We also emphasize that genetic
371 considerations are only one aspect in the choice of whether assisted gene flow should be pursued in a given situation
372 and ecological, political and logistic considerations are also important components of when deciding the most
373 suitable donor population(s) (Seddon et al., 2014). Finally, the complete isolation between many existing *S.*
374 *catenatus* populations throughout much of their range means that translocations may be the only way in which
375 specific populations may survive through the positive impacts of demographic rescue even in the face of possible
376 negative genetic impacts. This emphasizes the need to evaluate all potential risks and benefits – genetic and
377 demographic - using an appropriate risk-management framework (Liddell, Sunnucks, & Cook, 2021) when making
378 decisions about whether population augmentation should be pursued for a given species including Eastern
379 Masassauga rattlesnakes.
380

Acknowledgments

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, and we are deeply appreciative. Many individuals assisted, but we especially thank Jeff Davis, Michael Dreslik, Brian Fedorko, Tony Frazier, Kim Frolich, Dan Harvey, Matt Kowalski, Greg Lipps, Scott Martin, 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 thank Peri Bolton, Emily Bode, Greg Lipps, Andrew Mason, Marissa Roseman, for discussion and Marty Kardos, Andrew DeWoody and Greg Lipps for comments on the manuscript. We also thank and Kendra Wecker, Kate Parsons, Erin Hazlett, Carolyn Caldwell from the Ohio Division of Wildlife for their long-standing support of conservation genetics work by HLG on endangered snakes. 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. Computational analyses were performed using resources provided by the Ohio Supercomputer Center.

Data Availability Statement

Raw sequence data used in this study are available in NCBI's Short Read Archive BioProject Accession No PRJNA1220313. The scripts developed for data analysis and visualization and additional data can be publicly accessed on GitHub at https://github.com/samarth8392/genomic_compatibility.

Author Contributions

SM and HLG designed research; SM performed research and analyzed the data; SM and HLG wrote the paper.

Competing Interest Statement

The authors declare no competing financial interests.

References

Bell, D. A., Robinson, Z. L., Funk, W. C., Fitzpatrick, S. W., Allendorf, F. W., Tallmon, D. A., & Whiteley, A. R. (2019). The Exciting Potential and Remaining Uncertainties of Genetic Rescue. *Trends in Ecology & Evolution, 34*(12), 1070-1079. doi:10.1016/j.tree.2019.06.006

Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., . . . van Oosterhout, C. (2022). Genetic load: genomic estimates and applications in non-model animals. *Nature Reviews Genetics, 23*(8), 492-503. doi:10.1038/s41576-022-00448-x

Brooks, G. C., Chandler, H. C., Haas, C. A., & Kindsvater, H. K. (2024). Ten Principles from Evolutionary Ecology for the Effective Conservation of Reptiles and Amphibians. *Journal of Herpetology, 58*(3). doi:10.1670/2330062

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2014). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly, 6*(2), 80-92. doi:10.4161/fly.19695

Clarke, J. G., Smith, A. C., & Cullingham, C. I. (2024). Genetic rescue often leads to higher fitness as a result of increased heterozygosity across animal taxa. *Molecular Ecology, 33*(19). doi:10.1111/mec.17532

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158. doi:10.1093/bioinformatics/btr330

de Brevern, A. G., 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). doi:10.1371/journal.pone.0046688

Edmands, S. (2007). Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology, 16*(3), 463-475. doi:10.1111/j.1365-294X.2006.03148.x

Edmands, S. (2017). Heterosis and Outbreeding Depression in Interpopulation Crosses Spanning a Wide Range of Divergence. *Evolution, 53*(6), 1757-1768. doi:10.1111/j.1558-5646.1999.tb04560.x

Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics, 8*(8), 610-618. doi:10.1038/nrg2146

Fitzpatrick, S. W., & Funk, W. C. (2021). Genomics for Genetic Rescue. In P. A. Hohenlohe & O. P. Rajora (Eds.), *Population Genomics: Wildlife* (pp. 437-471). Cham: Springer International Publishing.

Fitzpatrick, S. W., Mittan-Moreau, C., Miller, M., Judson, J. M., & Zamudio, K. (2023). Genetic rescue remains underused for aiding recovery of federally listed vertebrates in the United States. *Journal of Heredity, 114*(4), 354-366. doi:10.1093/jhered/esad002

Frankham, R., Ballou, J. D., Eldridge, M. D. B., Lacy, R. C., Ralls, K., Dudash, M. R., & Fenster, C. B. (2011). Predicting the Probability of Outbreeding Depression. *Conservation Biology, 25*(3), 465-475. doi:10.1111/j.1523-1739.2011.01662.x

Frankham, R., Ballou, J. D., Ralls, K., Eldridge, M. D. B., Dudash, M. R., Fenster, C. B., . . . Sunnucks, P. (2017). *Genetic management of fragmented animal and plant populations* (First edition. ed.). Oxford: Oxford University Press.

Galtier, N., & Duret, L. (2007). Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends in Genetics, 23*(6), 273-277. doi:10.1016/j.tig.2007.03.011

Grossen, C., Guillaume, F., Keller, L. F., & Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nature Communications, 11*(1). doi:10.1038/s41467-020-14803-1

Hedrick, P. W., Allendorf, F. W., & Waples, R. S. (2013). Genetic engineering in conservation. *Nature, 502*(7471), 303-303. doi:10.1038/502303b

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., . . . Whitlock, M. C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist, 188*(4), 379-397. doi:10.1086/688018

Hufbauer, R. A., Szűcs, M., Kasyon, E., Youngberg, C., Koontz, M. J., Richards, C., . . . Melbourne, B. A. (2015). Three types of rescue can avert extinction in a changing environment. *Proceedings of the National Academy of Sciences, 112*(33), 10557-10562. doi:10.1073/pnas.1504732112

Ito, T., Yoshitake, K., & Iwata, T. (2021). ePat: extended PROVEAN annotation tool. *bioRxiv*, 2021.2012.2021.468911. doi:10.1101/2021.12.21.468911

Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M., . . . Funk, W. C. (2021). The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences, 118*(48). doi:10.1073/pnas.2104642118

Kardos, M., Keller, L. F., & Funk, W. C. (2024). What Can Genome Sequence Data Reveal About Population Viability? *Molecular Ecology*. doi:10.1111/mec.17608

Khan, A., Patel, K., Shukla, H., Viswanathan, A., van der Valk, T., Borthakur, U., . . . Ramakrishnan, U. (2021). Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. *Proceedings of the National Academy of Sciences, 118*(49). doi:10.1073/pnas.2023018118

Kyriazis, C. C., Serieys, L. E. K., Bishop, J. M., Drouilly, M., Viljoen, S., Wayne, R. K., & Lohmueller, K. E. (2024). The influence of gene flow on population viability in an isolated urban caracal population. *Molecular Ecology, 33*(9). doi:10.1111/mec.17346

Leroy, G., Carroll, E. L., Bruford, M. W., DeWoody, J. A., Strand, A., Waits, L., & Wang, J. (2017). Next-generation metrics for monitoring genetic erosion within populations of conservation concern. *Evolutionary Applications, 11*(7), 1066-1083. doi:10.1111/eva.12564

Liddell, E., Sunnucks, P., & Cook, C. N. (2021). To mix or not to mix gene pools for threatened species management? Few studies use genetic data to examine the risks of both actions, but failing to do so leads disproportionately to recommendations for separate management. *Biological Conservation, 256*. doi:10.1016/j.biocon.2021.109072

Margres, M. J., Wray, K. P., Hassinger, A. T. B., Ward, M. J., McGivern, J. J., Moriarty Lemmon, E., . . . Rokyta, D. R. (2017). 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

Mathur, S., & DeWoody, J. A. (2021). Genetic load has potential in large populations but is realized in small inbred populations. *Evolutionary Applications, 14*(6), 1540-1557. doi:10.1111/eva.13216

Mathur, S., Mason, A. J., Bradburd, G. S., & Gibbs, H. L. (2023). Functional genomic diversity is correlated with neutral genomic diversity in populations of an endangered rattlesnake. *Proceedings of the National Academy of Sciences, 120*(43). doi:10.1073/pnas.2303043120

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), 1297-1303. doi:10.1101/gr.107524.110

Mee, J. A., & Yeaman, S. (2019). Unpacking Conditional Neutrality: Genomic Signatures of Selection on Conditionally Beneficial and Conditionally Deleterious Mutations. *The American Naturalist, 194*(4), 529-540. doi:10.1086/702314

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

Ochoa, A., & Gibbs, H. L. (2021). Genomic signatures of inbreeding and mutation load in a threatened rattlesnake. *Molecular Ecology, 30*(21), 5454-5469. doi:10.1111/mec.16147

Orr, H. A. (2010). The population genetics of beneficial mutations. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1544), 1195-1201. doi:10.1098/rstb.2009.0282

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*. doi:10.1016/j.biocon.2020.108784

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research, 47*(D1), D886-D894. doi:10.1093/nar/gky1016

Robinson, J., Kyriazis, C. C., Yuan, S. C., & Lohmueller, K. E. (2023). Deleterious Variation in Natural Populations and Implications for Conservation Genetics. *Annual Review of Animal Biosciences, 11*(1), 93-114. doi:10.1146/annurev-animal-080522-093311

Seddon, P. J., Griffiths, C. J., Soorae, P. S., & Armstrong, D. P. (2014). Reversing defaunation: Restoring species in a changing world. *Science, 345*(6195), 406-412. doi:10.1126/science.1251818

Speak, S. A., Birley, T., Bortoluzzi, C., Clark, M. D., Percival-Alwyn, L., Morales, H. E., & van Oosterhout, C. (2024). Genomics-informed captive breeding can reduce inbreeding depression and the genetic load in zoo populations. *Molecular Ecology Resources, 24*(7). doi:10.1111/1755-0998.13967

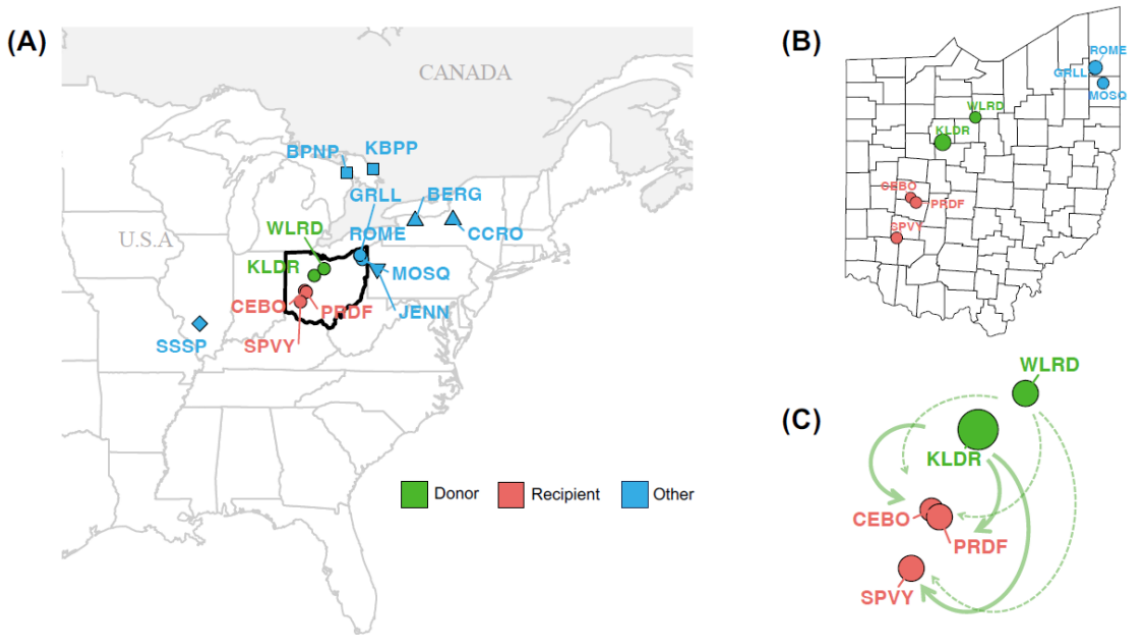
Stoletzki, N., & Eyre-Walker, A. (2010). Estimation of the Neutrality Index. *Molecular Biology and Evolution, 28*(1), 63-70. doi:10.1093/molbev/msq249

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., . . . DePristo, M. A. (2018). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics, 43*(1). doi:10.1002/0471250953.bi1110s43

Weeks, A. R., Sgro, C. M., Young, A. G., Frankham, R., Mitchell, N. J., Miller, K. A., . . . Hoffmann, A. A. (2011). Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications, 4*(6), 709-725. doi:10.1111/j.1752-4571.2011.00192.x

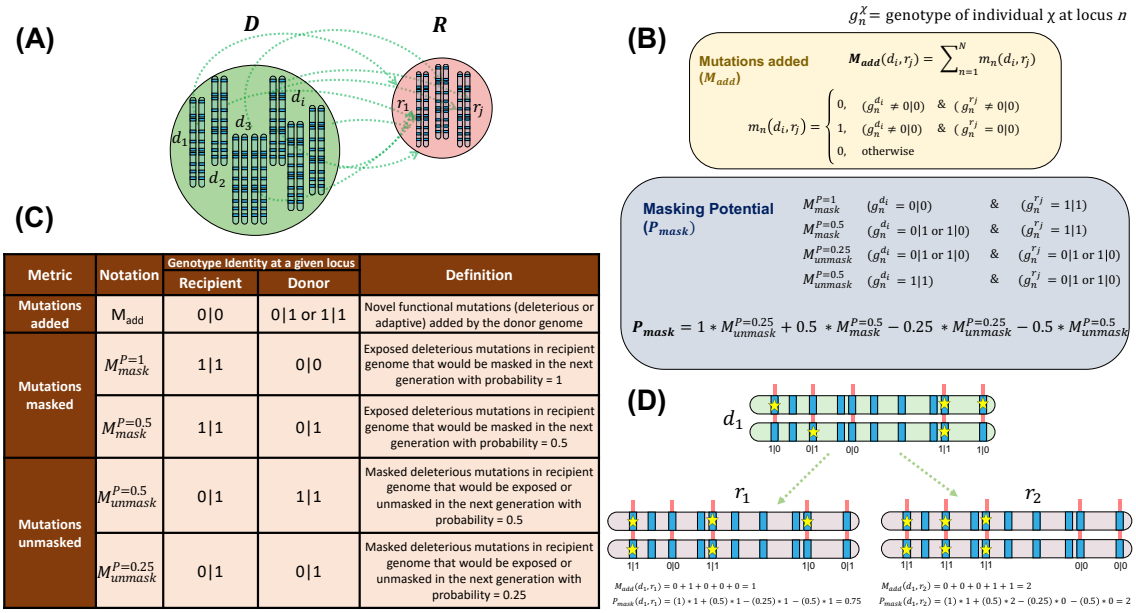
Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., & Tallmon, D. A. (2015). Genetic rescue to the rescue. *Trends in Ecology & Evolution, 30*(1), 42-49. doi:10.1016/j.tree.2014.10.009

524 **Figures**



525 **Figure 1: Sampling sites and potential donor and recipient sites for assisted gene flow of Eastern massasauga**
526 **rattlesnakes (*Sistrurus catenatus*) in the state of Ohio, United States.** (A) 152 whole genomes were analyzed to
527 identify genome-wide SNPs in different *S. catenatus* populations across the species range in the US and Canada
528 with (B) especially focusing on potential donor (green) and recipient (red) populations in the state of Ohio. (C) A
529 schematic showing different assisted gene flow scenarios that are possible (see text). Population labels in red
530 indicate the potential recipient populations, green are potential donor sites, and all other populations are labelled in
531 blue.

532



Metric	Notation	Genotype Identity at a given locus		Definition
		Recipient	Donor	
Mutations added	M_{add}	0|0	0|1 or 1|1	Novel functional mutations (deleterious or adaptive) added by the donor genome
Mutations masked	$M_{mask}^{P=1}$	1|1	0|0	Exposed deleterious mutations in recipient genome that would be masked in the next generation with probability = 1
	$M_{mask}^{P=0.5}$	1|1	0|1	Exposed deleterious mutations in recipient genome that would be masked in the next generation with probability = 0.5
Mutations unmasked	$M_{unmask}^{P=0.5}$	0|1	1|1	Masked deleterious mutations in recipient genome that would be exposed or unmasked in the next generation with probability = 0.5
	$M_{unmask}^{P=0.25}$	0|1	0|1	Masked deleterious mutations in recipient genome that would be exposed or unmasked in the next generation with probability = 0.25

Figure 2: Estimating genomic compatibility for assisted gene flow between a donor and a recipient population. (A) Genomic compatibility is defined between a Donor (D) and Recipient (R) population by comparing the functional genotypes between each pair of donor (d1, d2,…,di) and recipient (r1,r2,…,rj) genomes (Table S1) (B) To estimate genomic compatibility, we defined two metrics: mutations added (Madd) that measures the number of unique deleterious or adaptive alleles present in a donor genome that can be added to the recipient genome, and masking potential (Pmask) that estimates a cumulative score of the probabilities of masking a homozygous allele (Pmask > 0) and unmasking a heterozygote allele (Pmask < 0). (C) Table describing the metrics to count different category of mutations between a donor and recipient genome based on the genotype identity of the functional mutation. See Supplementary Methods for more details. (D) An example illustrating how the two genetic compatibility metrics can be estimated between a donor (d1) and two different recipient (r1, r2) genomes by comparing the genotype combinations of derived allele (star) at different loci of interest.

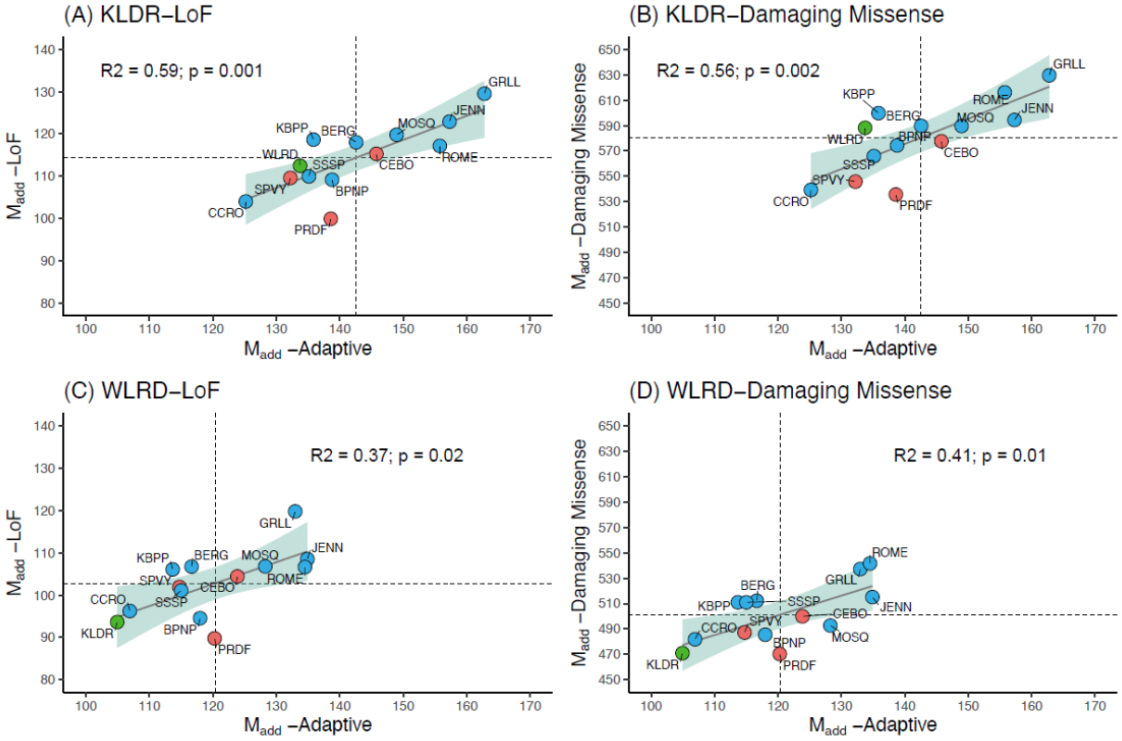


Figure 3: Genetic compatibility assessment based on novel adaptive variants and deleterious mutations added by donor to target recipient (and other) populations. Mean number of novel adaptive variants (x-axis) vs (A) LoF and (B) damaging missense mutations (y-axis) introduced by KLDR genomes that are absent in *S. catenatus* genomes from other populations (M_{add}). (C-D) Similar analysis was performed using WLRD genomes as donors.

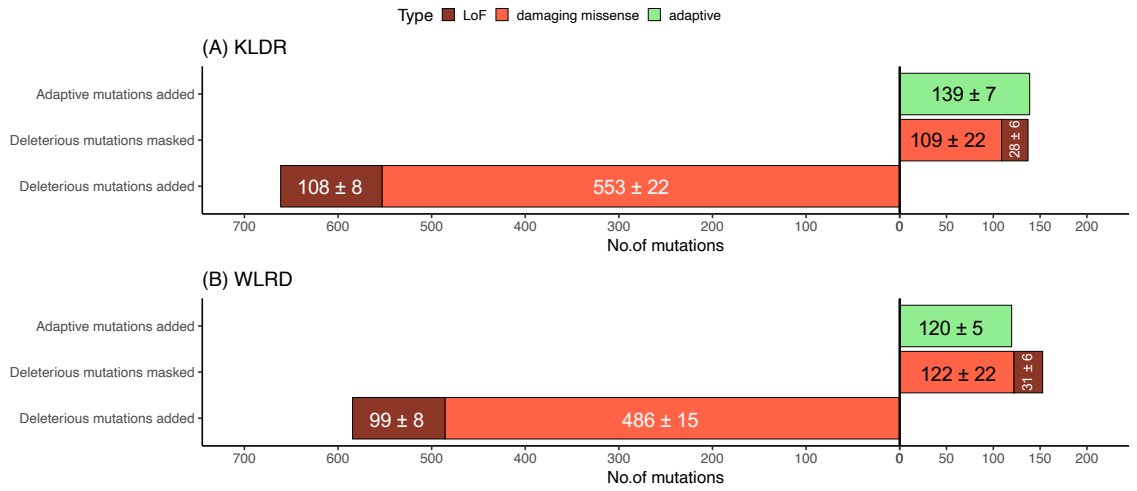


Figure 4: Summary of impact of assisted gene flow on recipient populations in terms of numbers of variants with negative and positive fitness-related effects. The bar plots represent the mean number of variants added or masked (at P = 1) averaged over the three recipient populations when augmented with (A) KLDR or (B) WLRD donor population. Note that the estimate of numbers of deleterious mutations masked is a minimum estimate because only mutations with masking probability of P=1. Numbers indicate mean +/- 1 S.D.

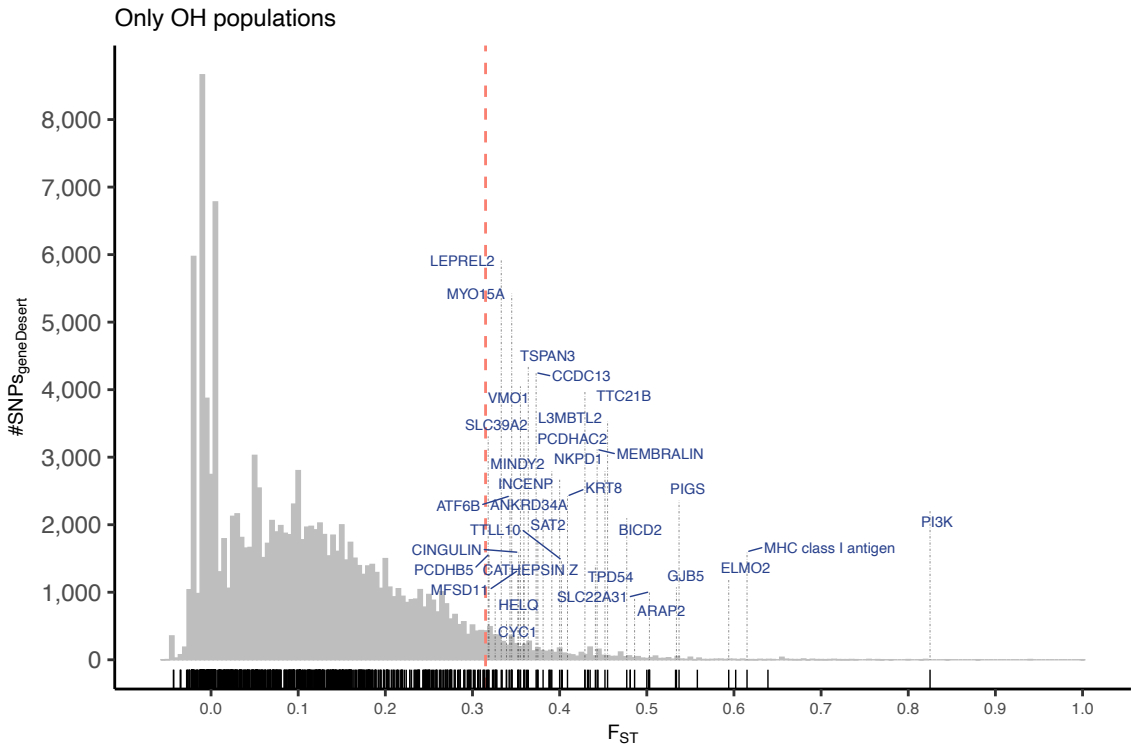


Figure 5: Population outlier analysis based on F_{ST} frequency histograms of functional and neutral sites in *S. catenatus* populations. Light grey bars represent F_{ST} distribution of SNPs in gene-desert regions of the genome; red dashed lines indicate the 95th percentile F_{ST} value in gene-desert regions. Black vertical lines at the bottom of each plot denote F_{ST} estimates at each missense SNP in the *S. catenatus* adaptive genes (Mathur et al. 2023). Adaptive genes that contain SNPs with F_{ST} > 95th %ile of gene-desert SNPs are labelled in blue. The per-SNP F_{ST} for each SNP was calculated only using OH samples ($SNP_{geneDesert}$ = 123,754; $SNP_{adaptive}$ = 1,053; $SNP_{adaptive}$ in top 5% of the gene desert F_{ST} distribution = 78).