Report

Current Biology

Genomic and Fitness Consequences of Genetic Rescue in Wild Populations

Highlights

- New gene flow into small, isolated guppy populations led to increases in abundance
- Mark-recapture and pedigree data show high hybrid survival and reproductive success
- Candidate adaptive alleles resist introgression more than neutral expectations
- Gene flow can rescue small populations without erasing adaptive variation

Authors

Sarah W. Fitzpatrick, Gideon S. Bradburd, Colin T. Kremer, Patricia E. Salerno, Lisa M. Angeloni, W. Chris Funk

Correspondence

sfitz@msu.edu

In Brief

Gene flow can limit adaptation but may also rescue small populations. Fitzpatrick et al. document genetic rescue in small populations of Trinidadian guppies. Wild pedigrees and mark-recapture data reveal high hybrid fitness and maintenance of putative adaptive alleles, suggesting assisted gene flow may be an effective conservation strategy.

Fitzpatrick et al., 2020, Current Biology *30*, 1–6
February 3, 2020
https://doi.org/10.1016/j.cub.2019.11.062

CellPress

Please cite this article in press as: Fitzpatrick et al., Genomic and Fitness Consequences of Genetic Rescue in Wild Populations, Current Biology (2019), https://doi.org/10.1016/j.cub.2019.11.062

Current Biology
Report

Genomic and Fitness Consequences of Genetic Rescue in Wild Populations

Sarah W. Fitzpatrick,[1,2,3,7,*] Gideon S. Bradburd,[2,3] Colin T. Kremer,[1] Patricia E. Salerno,[4,5] Lisa M. Angeloni,[4,6] and W. Chris Funk[4,6]

[1]W.K. Kellogg Biological Station, Michigan State University, 3700 E. Gull Lake Drive, Hickory Corners, MI 49060, USA
[2]Department of Integrative Biology, Michigan State University, 288 Farm Lane, East Lansing, MI 48824, USA
[3]Ecology, Evolutionary Biology, and Behavior Program, Michigan State University, East Lansing, MI 48824, USA
[4]Department of Biology, Colorado State University, 1878 Campus Delivery, Fort Collins, CO 80523, USA
[5]Universidad Regional Amazónica Ikiam, Km 7 Vía Muyuna, Tena, Napo, Ecuador
[6]Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO 80523, USA
[7]Lead Contact
*Correspondence: sfitz@msu.edu
https://doi.org/10.1016/j.cub.2019.11.062

SUMMARY

Gene flow is an enigmatic evolutionary force because it can limit adaptation but may also rescue small populations from inbreeding depression [1–3]. Several iconic examples of genetic rescue—increased population growth caused by gene flow [4, 5]—have reversed population declines [6, 7]. However, concerns about outbreeding depression and maladaptive gene flow limit the use of human-mediated gene flow in conservation [8, 9]. Rescue effects of immigration through demographic and/or genetic mechanisms have received theoretical and empirical support, but studies that monitor initial and long-term effects of gene flow on individuals and populations in the wild are lacking. Here, we used individual-based mark-recapture, multigenerational pedigrees, and genomics to test the demographic and evolutionary consequences of manipulating gene flow in two isolated, wild Trinidadian guppy populations. Recipient and source populations originated from environments with different predation, flow, and resource regimes [10]. We documented 10-fold increases in population size following gene flow and found that, on average, hybrids lived longer and reproduced more than residents and immigrants. Despite overall genomic homogenization, alleles potentially associated with local adaptation were not entirely swamped by gene flow. Our results suggest that genetic rescue was caused not just by increasing individual genetic diversity, rather new genomic variation from immigrants combined with alleles from the recipient population resulted in highly fit hybrids and subsequent increases in population size. Contrary to the classic view of maladaptive gene flow, our study reveals conditions under which immigration can produce long-term fitness benefits in small populations without entirely swamping adaptive variation.

RESULTS

Replicated translocations of guppies originating from a mainstem site moved into headwater sites located upstream of native, recipient guppy populations [11] (Figure 1A) provided the opportunity to study the demographic and evolutionary effects of gene flow between adaptively differentiated populations. We captured, uniquely marked, and monitored 9,590 Trinidadian guppies in focal recipient populations in the Caigual and Taylor Rivers, located on the south slope of the Northern Range Mountains in Trinidad, over the course of 29 consecutive months (~eight to ten guppy generations). Detection probabilities were high in both streams with monthly averages of 0.83 in Taylor and 0.86 in Caigual [12]. The first three capture events occurred prior to upstream translocations. Microsatellite genotypes for every individual captured during the first 17 months of the study were used to reconstruct pedigrees (n = 2,831 fish spanning six guppy generations), assign parentage, and estimate hybrid indices ranging from 0 (pure recipient genotype) to 1 (pure immigrant genotype), where an intermediate hybrid index of 0.5 reflects an individual that is maximally hybrid (i.e., F_1 hybrid). Throughout the course of our study, we captured 63 fish with immigrant genotypes in the Caigual recipient population and 753 immigrant genotypes in the Taylor recipient population, inferred by either elastomer mark (recognizable because different color and position combinations were used in the translocated site) or whether hybrid index equaled one. Observed differences between the two study sites in number of captured immigrant individuals were most likely due to differences in proximity between our focal sites, chosen based on the upstream-most extent of native guppies, and the translocation sites. Our Caigual focal site was located ~700 m downstream of the translocation site and only ~5 m downstream of the translocation site in the Taylor.

Sustained Population Growth

Following the onset of gene flow, population sizes increased nearly 10-fold throughout the 2 years in which Caigual and Taylor recipient populations were censused [12] (Figure 1B), with modest fluctuations driven by typical wet/dry season dynamics [13]. Prior to upstream translocations, Caigual and Taylor

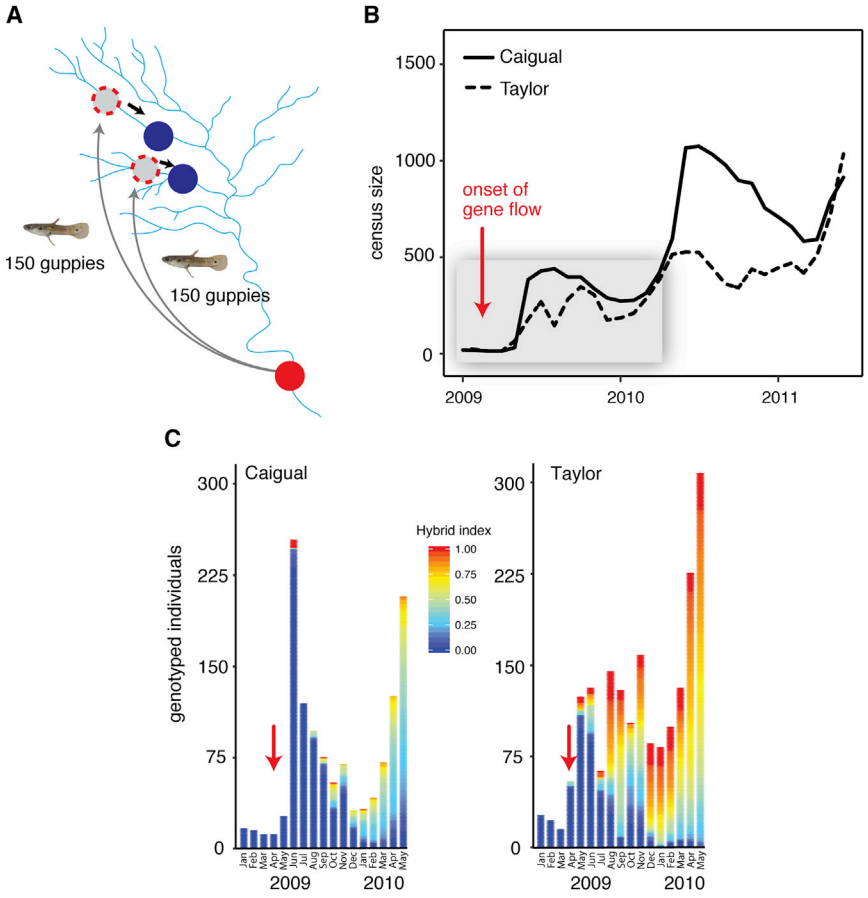

Figure 1. Gene Flow Manipulation Experiments in Trinidad

(A) Map of the Guanapo River drainage. In 2009, guppies were translocated from a downstream high-predation locality (red) into two headwater sites (dashed red) that were upstream of native recipient populations in low-predation environments (dark blue). Unidirectional, downstream gene flow began shortly after the introductions, indicated by black arrows.

(B) Census sizes in Caigual (solid) and Taylor (dashed) following the onset of gene flow from the upstream introduction sites. Gray box indicates the time span in which all captured individuals were genotyped at 12 microsatellite loci.

(C) Temporal patterns of continuous hybrid index assignments throughout the first 17 months of the study (~four to six guppy generations). Individuals from recipient populations prior to gene flow had a hybrid index = 0, and pure immigrant individuals had a hybrid index = 1. Hybrid indices were assigned using data from 12 microsatellite loci. Red arrows indicate the onset of gene flow.

populations numbered less than one hundred guppies each and were composed entirely of pure recipient genotypes (Figure 1C). Immigration led to an increase in the frequency of hybrid genotypes throughout the study's duration. By the end of the 17-month period for which we had individual genotype data, both populations were composed mostly of hybrid and immigrant individuals.

High Hybrid Fitness

Hybrid index was a strong predictor of variation in fitness in both streams. Hybrids and/or pure immigrants lived longer (Figures 2A and 2B) and had higher lifetime reproductive success (Figures 2C and 2D) than pure recipient individuals. Quadratic models relating fitness to hybrid index consistently outperformed linear or constant models. In some cases, these quadratic relationships clearly showed that hybrid genotypes lived longer (e.g., males in Caigual and both sexes in Taylor; Figures 2A and 2B) and had higher reproductive success (hybrids in Taylor; Figure 2D) than individuals with pure recipient or immigrant genotypes. In other cases, hybrids and pure immigrant individuals may have had comparable success (Figure 2C; females in Figure 2A). We found some evidence of zero inflation in longevity (i.e., more fish failed to survive beyond their initial capture than expected given the negative binomial distribution) in both Taylor and Caigual. In Caigual, zero inflation did not appear to vary between sexes or with hybrid index (Tables S2 and S5). However, in Taylor, zero inflation peaked at intermediate hybrid indices and was more common among females (Figure 3). We also found significant zero inflation in lifetime reproductive success of Taylor fish (i.e., more fish failed to reproduce than expected given the negative binomial distribution), especially those with recipient genotypes (indicated by significant zero inflation peaking at low hybrid indices; Figure 3).

Increased Diversity and Homogenized Variation at Most Loci

RAD sequencing (RAD-seq) genotyping of 12,407 SNPs (an average of one locus per 58,542 bp) in pre- and post-gene flow Caigual and Taylor populations and the mainstem source population revealed increased genomic variation within recipient populations and substantial genomic homogenization among all populations following the onset of gene flow. Before gene flow, the recipient populations were highly distinct from each other and from the source population (dark blue versus red in Figure 4A) and showed extremely low levels of genomic variation (Figure 4B). Genomic differentiation between recipient and source populations decreased dramatically after the onset of gene flow (light blue versus red in Figure 4A). Genome-wide average F_{st} between the recipient and source populations decreased from 0.29 to 0.01 in Caigual and from 0.31 to 0.02 in Taylor. Recipient Caigual and Taylor populations showed nearly entirely homozygous genomes and extremely low nucleotide diversity, followed by substantial increases in both metrics after the onset of gene flow (Figure 4B). Ninety-five percent of SNPs were monomorphic in Caigual and 96% in Taylor prior to gene flow, compared to 22% and 24% monomorphic loci after gene flow. Average nucleotide diversity (π) increased from 0.01 to 0.22 in Caigual and from 0.01 to 0.21 in Taylor.

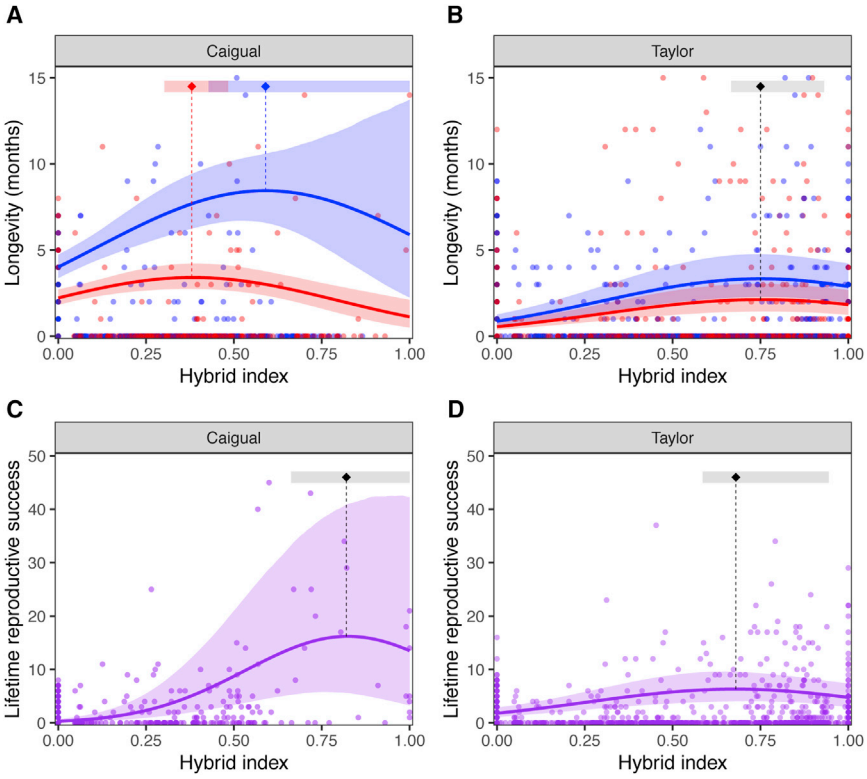

Figure 2. Relationships between Hybrid Index and Fitness

Fitness metrics (longevity and total lifetime reproductive success [LRS]) varied quadratically with hybrid index (0, pure recipient genotype; 1, pure immigrant genotype). Maxima of the quadratic functions are indicated by vertical dashed lines/diamonds; uncertainty in their positions is indicated by (horizontal) 95% confidence bars. Shading around regression lines displays approximate 95% confidence bands obtained through simulation. Model details appear in Tables S1–S5.

(A and B) Longevity differed between males (red) and females (blue). Generally, females lived longer than males, and fish in (A) Caigual lived longer than those in (B) Taylor. In Taylor, male and female longevity had quadratic relationships with hybrid index that differed in magnitude but peaked at similar parameter estimates; this differed by sex in Caigual (A versus B).

(C and D) LRS varied quadratically with hybrid index, and this trend did not differ between males and females. Individuals from Taylor generally had lower LRS than Caigual (C versus D) and were more likely to not reproduce at all (i.e., we detected significant zero inflation), especially those with recipient genotypes (hybrid indices near zero).

Plotted regressions in (A), (B), and (D) display the average predicted success of individuals after controlling for zero inflation. LRS in Caigual individuals (C) showed no strong evidence of zero inflation.

Locally Adaptive Variation Maintained

To study the maintenance of locally adaptive variation, we first needed to identify alleles involved in adaptation to headwater, low-predation (LP) environments. We assumed such alleles were favored in both LP recipient populations and selected against in the downstream, high-predation (HP) population. Using an arbitrary sample frequency difference cutoff of ≥ 0.9 between headwater populations and the mainstem (and ≤ 0.1 sample frequency difference between the pre-gene flow headwater populations), we identified a set of 146 such loci spanning all 23 linkage groups that were strong candidates for alleles involved in local adaptation (or linked to putatively adaptive alleles) to the LP environment. An excess of pre-gene flow ancestry at these loci (versus neutral expectations) in the post-gene flow populations would be evidence of selection for the maintenance of locally adaptive variation in the face of gene flow.

Despite overall genomic homogenization, we found evidence for selective maintenance of alleles at these candidate loci in Caigual and Taylor post-gene flow populations. Using simulations, we inferred that the frequencies of this set of candidate alleles were significantly more similar to the inferred ancestral LP population allele frequencies than expected compared to frequency-matched non-candidate loci (Figure 4C). In other words, an excess of pre-gene flow ancestry at candidate adaptive loci in post-gene flow populations is consistent with selection for the maintenance of locally adaptive variation that was present in pre-gene flow recipient populations. Furthermore, 71 of the 146 candidate alleles had positive ancestry deviations (i.e., resisted introgression more than neutral expectations) in both populations, suggesting some similarity in genomic responses to gene flow. BLAST query results against the Trinidadian guppy genome showed that 129 out of the 146 candidate loci were located in a gene or within 10 kb of a gene. However, there was no intersection between these BLAST hits and a set of 40 genes previously identified [14] as potential contributors to guppy phenotypes known to differ between headwater and mainstem environments, such as growth, vision, and pigment pattern development. We found no significant gene ontology enrichment terms among our candidate adaptive loci.

DISCUSSION

Gene flow between adaptively differentiated populations is typically assumed to swamp local adaptation and reduce fitness, but few studies have mechanistically tested multigenerational fitness effects of gene flow into small and isolated populations in the wild. We documented high hybrid fitness resulting in sustained population growth over multiple generations in replicated populations of Trinidadian guppies. Contrary to the prediction that small populations are especially vulnerable to genomic swamping, we showed that some portions of the recipient genome (associated with the local environment) were maintained, suggesting that genetic load was reduced without compromising potentially important adaptive variation.

Empirical tests of the phenotypic and fitness effects of gene flow in wild populations tend to yield idiosyncratic responses, giving rise to the prevailing wisdom that phenotypic effects of gene flow are trait specific and net fitness effects are difficult

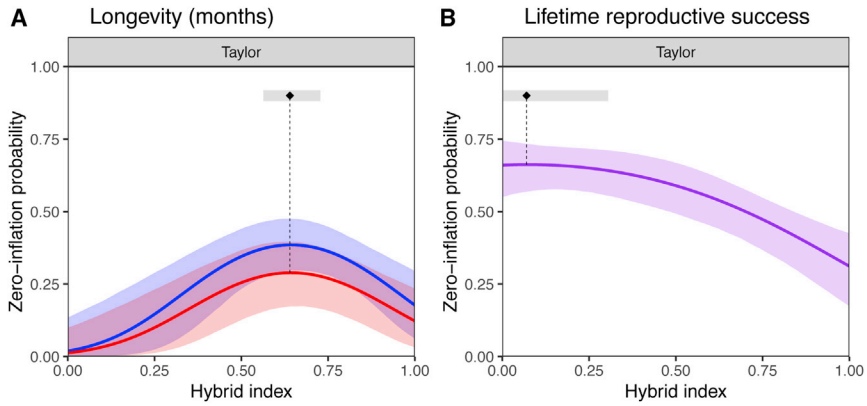

Figure 3. Zero Inflation Regressions

(A) Fish with intermediate hybrid indices, especially females (blue), showed elevated zero inflation probabilities for longevity in Taylor.

(B) Zero inflation also occurred in the lifetime reproductive success of Taylor fish, independent of their sex. In both panels, bands around regression lines display approximate 95% confidence bands around regressions, obtained through simulation.

Model details appear in Tables S1–S5.

to predict [1]. Here, we observed similar fitness and genomic responses to gene flow in two neighboring headwater populations, suggesting that the demographic and evolutionary responses to gene flow under similar conditions are not inherently idiosyncratic. This has important conservation implications: once the relevant factors (i.e., within and among population genetic and phenotypic characteristics) are understood and assessed, the design of successful assisted gene flow programs for threatened populations may be more feasible than expected.

Our pedigree analysis included fish up to six generations following gene flow, and we found that fish with intermediate hybrid indices (i.e., maximally hybrid) had the highest longevity and lifetime reproductive success (LRS). Given that guppies breed year round and have overlapping generations, it was not possible to directly compare fitness of different hybrid classes (i.e., F_1 to F_2). We also cannot untangle the specific genetic mechanisms underlying high hybrid fitness, which could be caused by dominance or overdominance of alleles in the hybrids or a combination of the two across different loci. Epistatic interactions among loci could also play a role as new interactions are presumably established in hybrid individuals. We do know that immigrant and hybrid groups showed similar levels of heterozygosity, but hybrids had higher fitness on average, suggesting that increased fitness was not solely due to increased genetic variation. Individuals with a hybrid index between 0.6 and 0.8 had the highest fitness in all cases except for male longevity in Caigual (fitness peak at 0.35), suggesting that high hybrid fitness extended beyond heterosis in the F_1 generation (hybrid index = 0.5). Interestingly, the average hybrid index of individuals with maximal fitness (0.8 in Caigual; 0.7 in Taylor) was similar to the average genome-wide hybrid index sampled at the end of our study, 8−10 generations after gene flow, suggesting individuals with hybrid genomes continued to contribute disproportionately to the observed increases in population size. Elevated probability of zero inflation at intermediate hybrid indices in the Taylor longevity model potentially represents a subset of hybrids that were less fit due to genomic incompatibilities or another uniting characteristic. However, the hybrids in this stream that did survive more than one capture ultimately lived longer than pure recipient or immigrant genotypes.

The extent to which adaptation in small populations is limited by genetic drift and facilitated by occasional gene flow is a question in need of more attention. In our study, recipient populations underwent substantial genomic changes following gene flow, generally conforming to expectations that gene flow increases genetic variation within populations and homogenizes differentiation among populations. However, whereas microsatellite analysis revealed genetic swamping by the immigrant genotype [12], genomic analysis revealed maintenance of candidate adaptive alleles at higher than expected frequencies. This result suggests that gene flow into small populations does not inevitably swamp locally important variation and highlights advantages of using genomic data to untangle the complexities of hybridized wild populations [15]. We note that, although candidate adaptive allele frequencies resisted introgression more than our neutral expectations, they did undergo substantial shifts toward the mainstem source population. Unlike the single pulse of immigration typically implemented when assisted gene flow is used for augmenting small populations, rates of gene flow in this study were high and continuous. Under the lower rates of migration recommended for assisted gene flow in management, strong selection might maintain adaptive alleles at higher frequency than we observed [2, 8].

We also note that the candidate alleles we identified likely do not represent the full extent of adaptive variation in our focal populations given the sparseness of our genotyping across the genome [16] and our stringent selection criteria that required similar allele frequencies in both headwater populations prior to gene flow. This approach excludes variants that might underlie adaptation in one site, but not the other. We opted for this conservative approach so as to restrict our analysis to the loci in which we had highest confidence about their environmental association. We also do not yet know the functional significance of alleles that resisted introgression, which is unsurprising given limited understanding of the genomic architecture of local adaptation in guppies. Our candidate loci might be located in (or linked to) relevant genes whose functions are unknown or they could affect uncharacterized traits involved in local adaption to the headwater environment (e.g., physiological and metabolic traits) that have not been mapped. Given these limitations, we emphasize that it is the signature of selection in the face of such high migration that is itself interesting. Further investigation of differential rates of introgression throughout the genome with higher resolution genomic data will help identify the genomic architecture of local adaptation to headwater environments in guppies. This task will be additionally strengthened by directly linking variable patterns of introgression to changes in traits and individual fitness.

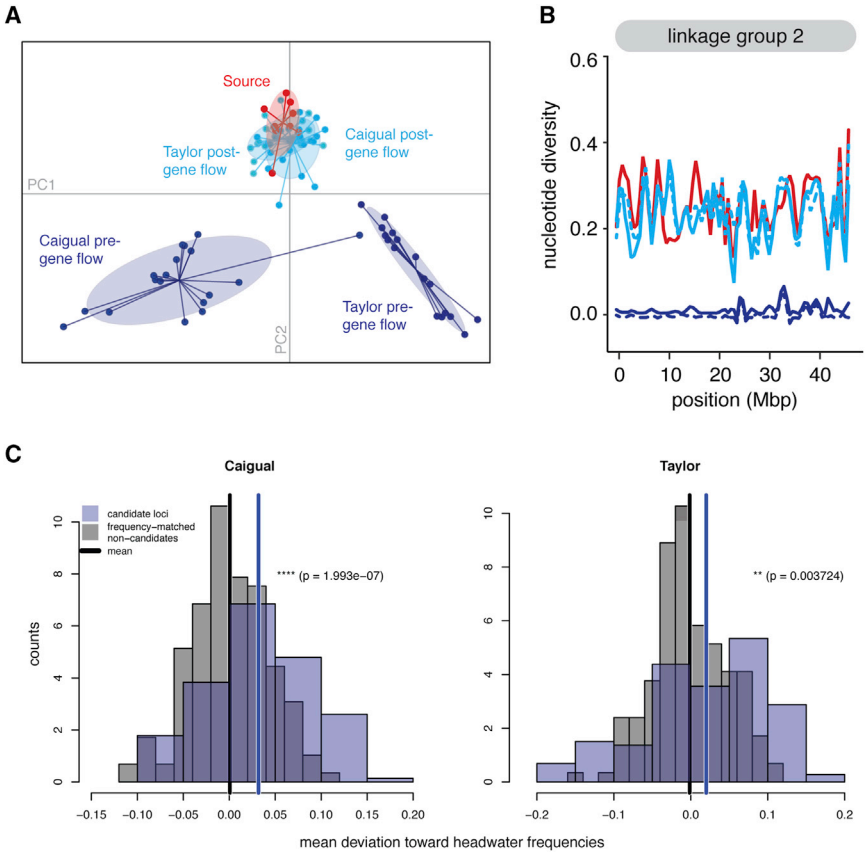

Figure 4. Genomic Consequences of Gene Flow

New gene flow caused overall genomic homogenization, but candidate adaptive alleles were maintained at higher than expected frequencies.

(A) PCA plot showing overall population differentiation based on polymorphic SNP loci from the RAD-seq data.

(B) Comparison of nucleotide diversity patterns along linkage group two among pre-gene flow (dark blue) and post-gene flow (light blue) Caigual (solid) and Taylor (dashed) populations and the introduction source (red). Similar patterns were found across all 23 linkage groups.

(C) Distributions of ancestry-polarized deviations in candidate loci versus frequency-matched non-candidates for both populations. In each stream, the allele frequencies of the candidate loci were significantly closer to the headwater ancestral frequency compared to a set of frequency-matched non-candidates.

In our view, the scenario studied here represents an ideal management outcome in which gene flow into small, inbred populations caused substantial increases in genomic variation, individual fitness, and population size but did not wipe out variation presumed to be locally adaptive. To what extent this scenario translates to other organisms, including species of conservation concern, is unknown. However, our results agree with a growing body of literature supporting the idea that gene flow from a closely related source into small, genetically depauperate populations can produce substantial demographic benefits [17–20]. These studies support a proposed paradigm shift in the genetic management of small populations from the current default of inaction to a new policy that considers restoring gene flow to recently fragmented populations [21]. Selection alone may be unlikely to counteract maladaptation caused by rapid global change, especially in small populations with low genetic variation [22]. In these cases, gene flow may be essential for providing the necessary variation for populations to persist and adapt to fast-paced environmental change.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Capture-mark-recapture
  - Microsatellite genotyping and pedigree reconstruction
  - RADseq data collection
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Fitness estimates and GLMMs
  - Detecting selection on locally adaptive variation
  - Workflow (1): Clustering analyses
  - Workflow (2): Simulating admixed populations
  - Workflow (3): Calculating ancestry deviation
  - Workflow (4): Identifying excess pre-gene flow ancestry at putatively locally adapted loci
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2019.11.062.

ACKNOWLEDGMENTS

We thank P. Bois and the many field assistants who contributed to our guppy capture-mark-recapture study in Trinidad. We thank Cameron Ghalambor and David Reznick for their intellectual contribution to this work. Experimental methods were approved by the Colorado State University Institutional Animal Care and Use Committee (protocol no. 12-3818A). Collection permits were graciously provided by the Fisheries division of Trinidad's Ministry of Food Production, Land and Marine Affairs. This work was supported by Michigan State University, The American Society of Naturalist's Student Research Award to S.W.F., and National Science Foundation grants DEB-0846175 to W.C.F.

and L.M.A. and DEB-1722621 to S.W.F. This is W.K. Kellogg Biological Station contribution no. 2144.

AUTHOR CONTRIBUTIONS

Conceptualization, S.W.F., L.M.A., and W.C.F.; Methodology, S.W.F., G.S.B., C.T.K., P.E.S., L.M.A., and W.C.F.; Formal Analysis, S.W.F., G.S.B., C.T.K., and P.E.S.; Investigation, S.W.F., G.S.B., C.T.K., and P.E.S.; Writing – Original Draft, S.W.F., G.S.B., and C.T.K.; Writing – Review & Editing, S.W.F., G.S.B., C.T.K., P.E.S., L.M.A., and W.C.F.; Funding Acquisition, S.W.F., L.M.A., and W.C.F.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 2, 2019
Revised: October 11, 2019
Accepted: November 20, 2019
Published: January 2, 2020

REFERENCES

1. Garant, D., Forde, S.E., and Hendry, A.P. (2007). The multifarious effects of dispersal and gene flow on contemporary adaptation. Funct. Ecol. *21*, 434–443.
2. Wright, S. (1931). Evolution in Mendelian populations. Genetics *16*, 97–159.
3. Frankham, R. (2015). Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. Mol. Ecol. *24*, 2610–2618.
4. Tallmon, D.A., Luikart, G., and Waples, R.S. (2004). The alluring simplicity and complex reality of genetic rescue. Trends Ecol. Evol. *19*, 489–496.
5. Whiteley, A.R., Fitzpatrick, S.W., Funk, W.C., and Tallmon, D.A. (2015). Genetic rescue to the rescue. Trends Ecol. Evol. *30*, 42–49.
6. Westemeier, R.L., Brawn, J.D., Simpson, S.A., Esker, T.L., Jansen, R.W., Walk, J.W., Kershner, E.L., Bouzat, J.L., and Paige, K.N. (1998). Tracking the long-term decline and recovery of an isolated population. Science *282*, 1695–1698.
7. Johnson, W.E., Onorato, D.P., Roelke, M.E., Land, E.D., Cunningham, M., Belden, R.C., McBride, R., Jansen, D., Lotz, M., Shindle, D., et al. (2010). Genetic restoration of the Florida panther. Science *329*, 1641–1645.
8. Mills, L.S., and Allendorf, F.W. (1996). The one-migrant-per-generation rule in conservation and management. Conserv. Biol. *10*, 1509–1518.
9. Rhymer, J.M., and Simberloff, D. (1996). Extinction by hybridization and introgression. Annu. Rev. Ecol. Syst. *27*, 83–109.
10. Gilliam, J.F., Fraser, D.F., and Alkins-Koo, M. (1993). Structure of a tropical stream fish community: a role for biotic interactions. Ecology *74*, 1856–1870.
11. Travis, J., Reznick, D., Bassar, R.D., López-Sepulcre, A., Ferriere, R., and Coulson, T. (2014). Do eco-evo feedbacks help us understand nature? Answers from studies of the Trinidadian guppy. Adv. Ecol. Res *50*, 1–40.
12. Fitzpatrick, S.W., Gerberich, J.C., Angeloni, L.M., Bailey, L.L., Broder, E.D., Torres-Dowdall, J., Handelsman, C.A., López-Sepulcre, A., Reznick, D.N., Ghalambor, C.K., and Chris Funk, W. (2016). Gene flow from an adaptively divergent source causes rescue through genetic and demographic factors in two wild populations of Trinidadian guppies. Evol. Appl. *9*, 879–891.
13. Grether, G.F., Millie, D.F., Bryant, M.J., Reznick, D.N., and Mayea, W. (2001). Rain forest canopy cover, resource availability, and life history evolution in guppies. Ecology *82*, 1546–1559.
14. Künstner, A., Hoffmann, M., Fraser, B.A., Kottler, V.A., Sharma, E., Weigel, D., and Dreyer, C. (2016). The genome of the Trinidadian guppy, *Poecilia reticulata*, and variation in the Guanapo population. PLoS ONE *11*, e0169087.
15. Fitzpatrick, S.W., and Funk, W.C. (2019). Genomics for genetic rescue. In Population Genomics (Springer).
16. Hoban, S., Kelley, J.L., Lotterhos, K.E., Antolin, M.F., Bradburd, G., Lowry, D.B., Poss, M.L., Reed, L.K., Storfer, A., and Whitlock, M.C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. Am. Nat. *188*, 379–397.
17. Hufbauer, R.A., Szűcs, M., Kasyon, E., Youngberg, C., Koontz, M.J., Richards, C., Tuff, T., and Melbourne, B.A. (2015). Three types of rescue can avert extinction in a changing environment. Proc. Natl. Acad. Sci. USA *112*, 10557–10562.
18. Robinson, Z.L., Coombs, J.A., Hudy, M., Nislow, K.H., Letcher, B.H., and Whiteley, A.R. (2017). Experimental test of genetic rescue in isolated populations of brook trout. Mol. Ecol. *26*, 4418–4433.
19. Hasselgren, M., Angerbjörn, A., Eide, N.E., Erlandsson, R., Flagstad, Ø., Landa, A., Wallén, J., and Norén, K. (2018). Genetic rescue in an inbred arctic fox (*Vulpes lagopus*) population. Proc. Biol. Sci. *285*, 20172814.
20. Kronenberger, J.A., Gerberich, J.C., Fitzpatrick, S.W., Broder, E.D., Angeloni, L.M., and Funk, W.C. (2018). An experimental test of alternative population augmentation scenarios. Conserv. Biol. *32*, 838–848.
21. Ralls, K., Ballou, J.D., Dudash, M.R., Eldridge, M.D.B., Fenster, C.B., Lacy, R.C., Sunnucks, P., and Frankham, R. (2018). Call for a paradigm shift in the genetic management of fragmented populations. Conserv. Lett. *11*, e12412.
22. Ceballos, G., Ehrlich, P.R., and Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proc. Natl. Acad. Sci. USA *114*, E6089–E6096.
23. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.
24. Meirmans, P.G., and Van Tienderen, P.H. (2004). GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol. Ecol. Notes *4*, 792–794.
25. Jones, O.R., and Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. Mol. Ecol. Resour. *10*, 551–555.
26. Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., and Cresko, W.A. (2013). Stacks: an analysis tool set for population genomics. Mol. Ecol. *22*, 3124–3140.
27. Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics *26*, 873–881.
28. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.
29. Buerkle, C.A. (2005). Maximum-likelihood estimation of a hybrid index based on molecular markers. Mol. Ecol. Notes *5*, 684–687.
30. Etter, P.D., and Johnson, E. (2012). RAD paired-end sequencing for local de novo assembly and SNP discovery in non-model organisms. Methods Mol. Biol. *888*, 135–151.
31. Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Mächler, M., and Bolker, B.M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J. *9*, 378–400.
32. Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. *45* (D1), D183–D189.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
9,590 Trinidadian guppy (*Poecilia reticulata*) scale samples and photographs	Authors	N/A
Deposited Data		
Demographic and fitness data reported in this paper	Authors	https://github.com/ctkremer/guppy
VCF file of SNPs used in genomic analyses	Authors	https://github.com/gbradburd/guppy_seln
Software and Algorithms		
R: A language and environment of statistical computing	R	https://www.r-project.org
GENEIOUS 7.1.7	[23]	https://www.ncbi.nlm.nih.gov/pubmed/22543367
GenoDive v.2	[24]	http://www.bentleydrummer.nl/software/software/GenoDive.html
Colony2	[25]	https://www.zsl.org/science/software/colony
Stacks v.1.09	[26]	http://catchenlab.life.illinois.edu/stacks/
GSnap	[27]	http://research-pub.gene.com/gmap/src/README
ADMIXTURE	[28]	N/A
Hybrid fitness analysis code	this paper	https://github.com/ctkremer/guppy
Selection analysis code	this paper	https://github.com/gbradburd/guppy_seln

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Sarah Fitzpatrick (sfitz@msu.edu). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We studied two wild populations of Trinidadian guppies (*Poecilia reticulata*) in headwater environments in the Caigual and Taylor Rivers located in the Guanapo drainage on the south slope of the Northern Range Mountains in Trinidad. Focal populations were located downstream of translocation sites where guppies originating from a mainstem site on the Guanapo River were introduced ([11]; Figure 1A). In total, we handled 9,590 individual guppies throughout 29 months (4,880 in Caigual; 4,710 in Taylor). Animal care and experimental procedures were approved by Colorado State University's Institutional Animal Care and Use Committee (protocol no. 12-3818A) and by the Fisheries division of Trinidad's Ministry of Food Production, Land and Marine Affairs.

METHOD DETAILS

Capture-mark-recapture

Focal stream reaches were sampled every month from January 2009 to June 2011. Three sample occasions occurred prior to upstream introductions (see [11] for detailed description of the upstream guppy translocation experiment), followed by 26 additional sample occasions. All guppies greater than 14 mm were caught using nets or minnow traps and transported to the lab in Nalgene (Rochester, NY, USA) bottles filled with stream water. In the lab, fish were housed in aerated tanks separated by location and sex. Fish were anesthetized with dilute MS-222 and processed under a dissecting microscope. New recruits had three scales removed and dried for DNA extraction and were given a unique set of visible implant elastomer marks (Northwest Marine Technologies, Shaw Island, WA, USA) using eight marking sites and 12 possible colors. The set of marks used in our study was distinct from the set used in the upstream translocations. All fish each month were weighed, photographed, and returned to their exact location within the focal stream reach one to two days after initial capture. Lab mortality was less than 0.5%.

Microsatellite genotyping and pedigree reconstruction

We extracted genomic DNA from scale samples from all individuals caught in the first 17 (of 29) capture occasions using Gentra Puregene Tissue Kits (QIAGEN). Guppies were genotyped at 12 microsatellite markers as described in [12]. Briefly, we amplified loci using QIAGEN Type-It Microsatellite Multiplex PCR kits. PCR products combined with HiDi formamide and LIZ size standard were sent to

the Life Sciences Core Laboratory at Cornell University to be read on an ABI 3730xl automated sequencer. Genotypes were visualized and scored using the microsatellite plug-in with GENEIOUS 7.1.7 [23]. We scored two positive controls and one negative control on each plate and found low genotyping error rates (< 0.5%). In total, we genotyped 3,298 guppies (1,491 from Caigual; 1,807 from Taylor) at 12 microsatellite loci. We calculated a continuous hybrid index between 0 (''pure'' recipient genotype) and 1 (''pure'' immigrant genotype) for all individuals using the maximum likelihood method of [29] in the GenoDive software [24]. A set of 20 pre-gene flow fish from Caigual and Taylor and 20 fish from the mainstem source population were used as reference populations to estimate hybrid indices.

We reconstructed wild pedigrees of the Caigual and Taylor focal populations using Colony2 [25], specifying a polygamous mating system without inbreeding, a genotyping error rate of 0.005, and the full-likelihood analysis method with ''high'' likelihood precision and ''medium'' runs that were repeated five times with different random seeds to maximize correct parentage assignment. Parent-offspring relationships were only assigned if the same assignment was made in at least four out of the five runs. Separate analyses were carried out for each stream and for each successive cohort of new recruits. The set of possible parents for a given sample occasion were all guppies caught during previous occasions and the set of possible offspring were all new recruits to the population. The final pedigree consisted of 1,106 individuals in Caigual (458 maternal links, 655 paternal links) and 1,725 individuals in Taylor (975 maternal links, 994 paternal links) spanning 4-6 overlapping generations.

RADseq data collection

We collected genomic data using RADseq for a subset of 96 individuals collected before gene flow (20 from Caigual and 20 from Taylor collected in 2009) and again approximately 10 generations following the onset of gene flow (23 from Caigual and 23 from Taylor collected in 2012), as well as individuals from the downstream source population (10 from Guanapo collected in 2009). Genomic DNA for RADseq library preparation was purified from guppy scales using the DNeasy Blood & Tissue kit with an additional RNase A treatment following the manufacturer's recommended protocols (QIAGEN, CA, USA). Purified DNA was quantified using a Qubit 2.0 Fluorometer (Invitrogen by Life Technologies) and brought to equal concentration. We prepared RAD sequencing libraries for 96 individuals following the protocol of [30]. RAD libraries were sequenced on an Illumina HiSeq 2500 sequencer at the University of Oregon with single-end 100 bp reads.

Raw sequence reads were demultiplexed using the process_radtags program in Stacks v.1.09 [26]. Reads with raw Phred Quality scores greater than 20, the correct barcode, and an unambiguous RAD site were retained. We used the guppy genome produced from a female guppy from the Guanapo source population as a reference sequence (version GCF_000633615.1_Guppy_female_1.0 [14],). Demultiplexed reads were aligned to the reference genome using GSnap [27]. We required unique alignments, allowing for a maximum of five mismatches, the presence of up to two indels, and no terminal alignments. Aligned reads were analyzed using the ref_map.pl program in Stacks, which derived each locus from overlapping GSnap alignments to produce a consensus sequence. SNPs were determined and genotypes called using a maximum-likelihood statistical model implemented in Stacks. We removed four individuals that had greater than 50% missing loci. To include a locus in further analyses, we required it to be genotyped in at least 60% of individuals of each population. Only a single randomly chosen SNP per RAD locus was included. We removed loci with a minor allele frequency cutoff of less than 0.02 and loci below a log likelihood threshold of −30. At this point, we removed an additional six individuals that had an average read depth of less than ten. We did not test for linkage disequilibrium or conformation to Hardy-Weinberg proportions because we expected extensive physical linkage and non-random mating due to the extensive admixture between recipient and source populations.

QUANTIFICATION AND STATISTICAL ANALYSIS

Fitness estimates and GLMMs

We estimated two components of total fitness: longevity and total lifetime reproductive success. Longevity was defined as the maximum number of months an individual was confirmed as present in our study (after the first occurrence) using the full mark-recapture dataset. This value is likely an underestimate of true survival due to imperfect detection probability and because new recruits had to be greater than 14 mm (approximately two months old) to be detected. Lifetime reproductive success was defined as the total number of offspring assigned to an individual using the reconstructed pedigree from individuals caught and genotyped in the first 17 months of data collection. We restricted our estimates to individuals captured in the first 13 months because individuals sampled after this cohort lacked enough time to reproduce and have their offspring identified.

We used generalized linear mixed models (GLMMs) to examine the relationship between an individual's fitness components and their hybrid index, accounting for the month the individual was first captured (i.e., its cohort) as a random effect, and checking for potential effects of sex and zero-inflation. Variation in both fitness components was modeled using a negative binomial distribution; this discrete distribution is appropriate for potentially over-dispersed count data (such as the number of offspring produced) as well as our discretized estimate of lifespan (here, number of months). We considered a suite of competing models describing changes in the negative binomial distribution's mean as a function of hybrid index and/or sex, using a log link function (Tables S1 and S4). A subset of these models also tested for zero-inflation (i.e., zeros in excess of those inherently predicted by the negative binomial distribution) and examined whether the extent of zero-inflation varied with hybrid index and/or sex, using a logit link function. All models were fit using the glmmTMB package [31] in R (version 3.3.3).

From the set of candidate models fit to each separate fitness component and stream, we used AICc comparison to select the best model (Tables S1 and S4). For each of the four resulting models, we also: (i) obtained approximate 95% confidence bands, and (ii) estimated 95% confidence intervals for the value of the hybrid index where longevity or reproductive success was highest (i.e., the maxima of the quadratic function). Both of these analyses relied on simulating 10,000 new datasets based on the original model fits (using glmmTMB's simulate function) and re-fitting the model to each dataset. For each new fit, we calculated the position of the quadratic maxima; all values above 1 were rounded down (as hybrid index ranges from 0 to 1). Approximate 95% confidence intervals were then estimated as falling between the 2.5^{th} and 97.5^{th} quantiles of the resulting distribution. Confidence bands were obtained similarly, by constructing distributions of predicted regression lines at a range of hybrid indices.

Detecting selection on locally adaptive variation

In this section, we describe in detail the procedure used to identify the signature of selection for the maintenance of locally adaptive variation in the face of gene flow. Our logic is as follows: If pre-gene flow headwater populations were locally adapted to their headwater, low predation habitat, and there has been selection for the maintenance of locally adapted variation in the face of gene flow, the signature of that selection would be greater-than-expected amounts of pre-gene flow headwater ancestry in the post-gene flow headwater populations at and around the alleles involved in local adaptation. Identifying this signature can be difficult, as gene flow may homogenize even large allele frequency differences that have arisen between populations due to different directional selection. However, the sampling of individuals from the pre-gene flow headwater and mainstem populations, as well as replication across two independent headwater populations, gives us power to determine whether there has been selection on locally adaptive variation. Here, we briefly lay out the steps of our analysis, and subsequently go into greater depth on each step:

1) We ran the model-based clustering method ADMIXTURE [28] to estimate admixture proportions for the post-gene flow headwater populations as well as allele frequencies in "pure" pre-gene flow headwater and mainstem populations.
2) Using those outputs, we simulated many post-gene flow headwater stream samples with the same admixture make-up and sampling noise.
3) We then calculated an ancestry deviation for each locus in the dataset. This ancestry deviation measures the amount by which an allele's frequency differs from the simulation-based expectation in the direction of the inferred ancestral headwater frequency, suggesting directional selection for the locally adapted allele in the face of gene flow from the mainstem.
4) Finally, we identified a set of loci that were strong candidates for alleles involved in local adaptation to the headwater environment prior to gene flow. To determine whether the observed deviations at our candidate SNP-set differed from the null, we matched each of our candidate alleles by frequency with a non-candidate locus, and compared the distribution of ancestry deviations in our candidate set to that of our frequency-matched non-candidate set.

We found that our candidate loci were significantly more "headwater" in their frequencies than their frequency-matched null set. This result supports the inference that headwater populations were locally adapted, and that there was selection for the maintenance of adaptive variation in the headwater habitats in the face of gene flow from the mainstem.

Workflow (1): Clustering analyses

For each post-gene flow headwater population, we generated a parametric null hypothesis for the frequency of each SNP, against which we could compare the observed frequencies to determine whether there was greater maintenance of local variation than we would expect. To generate this null for each headwater stream, we first ran ADMIXTURE on a dataset consisting of the two pre-gene flow populations (headwater and mainstem) and the post-gene flow headwater population, modeling individuals as draws from two discrete population clusters ($K = 2$). We ran 10 replicate ADMIXTURE runs on each of the two drainage datasets (Caigual and Taylor drainages), specifying a different random seed for each run to ensure that results were consistent across runs (they were).

We were interested in two quantities from each of these analyses: 1) the matrix of admixture proportions inferred for our post-gene flow populations (one admixture proportion per individual per cluster; Figure S1); and 2) the vector of inferred allele frequencies in each canonical cluster (Figure S2). As expected, individuals from post-gene flow headwater sites were of majority "mainstem" ancestry in both streams (although to slightly differing degrees). The sampled mainstem individuals were inferred to have some admixture with pre-gene flow headwater populations in analyses in both streams, which is biologically plausible given the downstream direction of stream-flow.

In both streams, there were strong correlations between the estimated allele frequencies in the two inferred clusters and the sample allele frequencies observed in the pre-gene flow headwater and mainstem populations. The allele frequencies estimated in "Cluster 2" in both runs, which corresponded most closely to the mainstem population, were quite similar to each other. The allele frequencies estimated for "Cluster 1" in both runs were very highly correlated with the sample frequencies observed in each respective pre-gene flow headwater population. The sample allele frequencies observed in the pre-gene flow headwater populations in each stream (pre_CA and pre_TY) were, for the most part, not strongly correlated with each other, presumably the result of the isolation they have experienced and the independent drift and/or selection they have undergone while isolated.

Workflow (2): Simulating admixed populations

We then used the admixture proportions and cluster allele frequencies estimated in ADMIXTURE to simulate admixed populations to match the observed post-gene flow headwater populations. We had to use the ADMIXTURE-estimated frequencies, rather than observed pre-gene flow population frequencies, because some of our pre-gene flow samples were inferred to be admixed, and so did not offer a clear glimpse of the "pure" parental population frequencies. We describe the simulation procedure for a single stream below, using w_{ik} to denote the estimated admixture proportion of the ith individual in the kth cluster, and f_{lk} to denote the estimated allele frequency at the lth locus in the kth cluster.

To simulate a single haplotype in individual i within a stream, we randomly chose a fraction w_{i1} of all genotyped loci to be of Cluster *1* ancestry, and assigned the remaining 1-w_{i1} fraction of loci to be of Cluster 2 ancestry. At a given locus, we then simulated a haploid genotype as a Bernoulli draw with probability of success f_{lk}; this step approximates the randomness of the sampling procedure used in the original genotyped dataset. We simulated two haplotypes across all loci for each individual, and repeated this procedure for each genotyped individual in the stream. We simulated 1000 replicate datasets for each stream. Note that we were only simulating data for the post-gene flow headwater populations in our sample (Post-gene flow Caigual – post_CA; Post-gene flow Taylor - post_TY). Overall, we saw tight correlations between the simulated and observed data, with the sample frequencies at only 1.01% of loci falling outside the 95% quantile of simulations in Caigual, and 1.53% in Taylor.

Workflow (3): Calculating ancestry deviation

Using these simulations, we calculate an estimated mean allele frequency at each locus in each simulated post-gene flow headwater population, as well as the deviation observed from that expectation at each locus in each headwater stream. The estimated mean allele frequency at a locus was defined as the mean of the simulated frequencies across all simulated replicates, and the deviation was the difference between the observed sample allele frequency in a post-gene flow headwater population (either post_CA or post_TY) and the simulation mean. If our data were well described by the ADMIXTURE model, the distribution of deviations from the model-based expectation should have had mean zero and small variance. In practice, we saw that the distribution of deviations from expectation has mean 2.9 × 10^-3 and standard deviation 4.3 × 10^-2 in Caigual, and mean 2.6 × 10^-3 and standard deviation 4.2 × 10^-2 in Taylor.

We could further characterize the deviation at each locus by its direction: either toward the ancestral headwater frequency or the ancestral mainstem frequency. To do this, we used the allele frequencies estimated in Clusters 1 and 2 from the ADMIXTURE analyses as the ancestral headwater and mainstem population frequencies within each stream, respectively. We defined the ancestry-polarized deviation at a locus as positive when the difference between the observed sample frequency and the simulation-based expectation was in the direction of the ancestral pre-gene flow headwater allele frequency, and negative when it was not (Figure S3). If there was no difference between the observed and expected allele frequencies, the deviation was zero. We could then calculate the ancestry-polarized deviations across all loci in both streams (Figure S3).

Workflow (4): Identifying excess pre-gene flow ancestry at putatively locally adapted loci

To detect a signal of excess pre-gene flow headwater ancestry, we started by identifying alleles that matched our expectations for locally adapted loci. Alleles were included in our candidate list if the difference in sample frequencies between both pre-gene flow headwater populations was less than or equal to 0.1, and the frequency difference between each of the pre-gene flow headwater populations and the mainstem population was greater than or equal to 0.9. These are stringent criteria, so we undoubtedly have a high false negative detection rate, but as we are not interested in these loci individually, and rather in the signal of ancestry deviation aggregated across them, we feel the sacrifice in Type II error is worth the gains in Type I error. Note that other methods for identifying alleles involved in selection between headwater and mainstem habitat are hampered by: a) the apparent extent of genome-wide divergence between headwater and mainstem populations; b) our RADseq dataset, which offers a necessarily limited view of the full genomic dynamics; and c) a lack of hypotheses about loci involved in traits that might be under divergent selection between headwater and mainstem habitats.

In all, 146 loci met our criteria to be considered candidates, and we calculated the ancestry-polarized deviation for each of these loci. However, because the frequencies (both observed and expected) of these alleles affect the distribution of their deviations, it may not be appropriate to simply compare the distribution of ancestry-polarized deviations for these candidate loci to that of all loci, or all other loci.

Instead, we took the approach of comparing the distribution of ancestry-polarized deviations from the 146 candidate loci to that of a set of frequency-matched loci. To match by frequency at a locus, we chose another locus (that was not part of the candidate set) for which the mean simulated frequency fell within 0.05 of the observed post-gene flow allele frequency. We did this sampling without replacement, so that no two candidate loci were frequency-matched to the same locus.

We then calculated the distribution of ancestry-polarized deviation for this null set and compared it to that of our candidate loci to determine whether ancestry at the candidate loci was biased toward ancestral headwater frequencies. To assess significance, we used a one-tailed t test, paired by locus, within each stream (Figure 3C). We found that, in each stream, the frequencies of the candidate loci were significantly more "headwater" (pre-gene flow ancestry) than their frequency-matched null set. In Caigual, the mean deviation from prediction toward pre-gene flow headwater frequencies was 3.182e-2 for candidate loci and 6.566e-4 for non-candidate loci (one tailed paired t test, p = 1.993e-7). In Taylor, the mean deviation from prediction toward pre-gene flow headwater frequencies was 2.002e-2 for candidate loci and 5.089e-3 for non-candidate loci (one tailed paired t test, p = 3.724e-3). In total,

105 of the 146 loci in Caigual, and 88 of 146 in Taylor, had positive deviations toward headwater ancestry, compared to 71/146 of the frequency-matched Caigual null set, and 63/146 the null set from Taylor. All candidate loci were used in a BLAST query against the Trinidadian guppy genome ([14]; NCBI Poecilia reticulata Annotation Release 101; GCF_000633615.1). Using the Zebrafish reference genome we carried out a gene ontology enrichment analysis of the candidate adaptive loci implemented in PANTHER 14.1 [32].

DATA AND CODE AVAILABILITY

Original data and scripts for running the hybrid fitness models are available at (https://github.com/ctkremer/guppy). Scripts for running the tests for selection are available at (https://github.com/gbradburd/guppy_seln).

Current Biology, Volume *30*

Supplemental Information

Genomic and Fitness Consequences of Genetic Rescue in Wild Populations

Sarah W. Fitzpatrick, Gideon S. Bradburd, Colin T. Kremer, Patricia E. Salerno, Lisa M. Angeloni, and W. Chris Funk

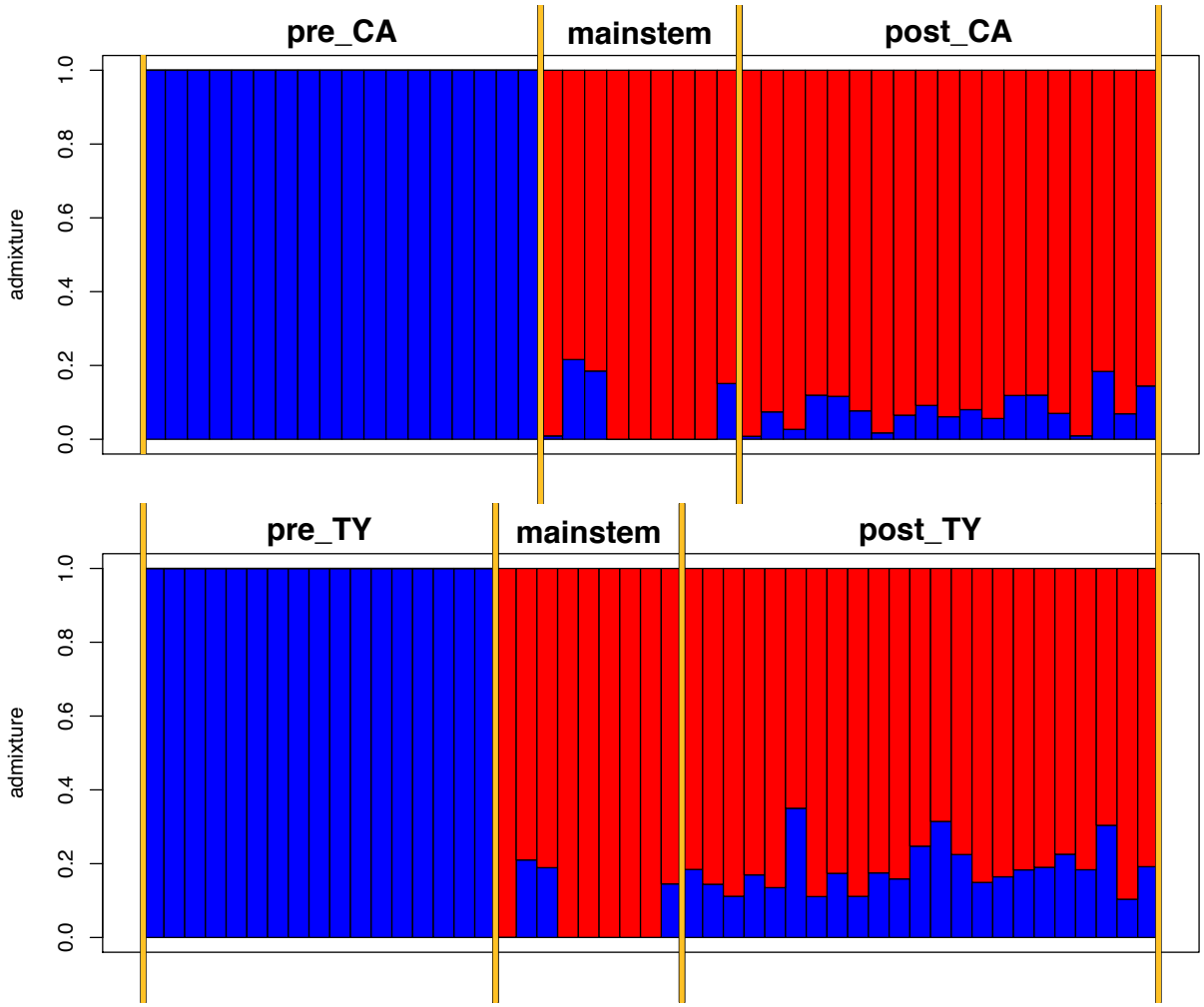

Figure S1. Structure plot of admixture proportions inferred by ADMIXTURE in each stream. Related to STAR Methods. Pre-gene flow headwater populations (pre-gene flow Caigual- pre_CA; pre-gene flow Taylor- pre_TY) were inferred to be of pure "headwater" ancestry, while the downstream mainstem population (mainstem) was inferred to be slightly admixed in analyses of both streams. Both post-gene flow headwater populations were inferred to have majority mainstem ancestry, although to slightly differing degrees.

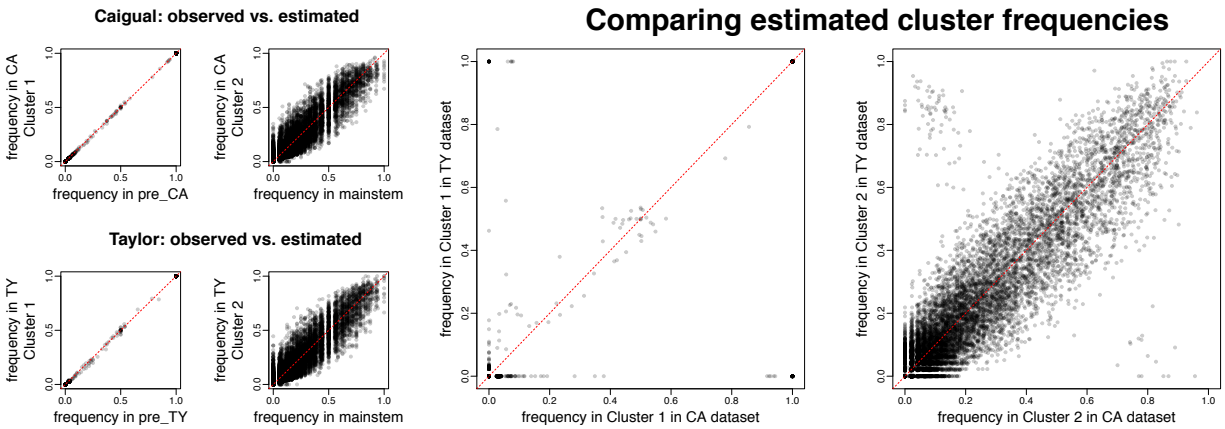

Figure S2. Observed and estimated allele frequencies for all SNPs. Related to STAR Methods. On the left, four small plots show the observed allele frequencies in the mainstem population (mainstem) and both pre-gene flow headwater populations (pre-gene flow Caigual – pre_CA; pre-gene flow Taylor - pre_TY) compared to the allele frequencies estimated by ADMIXTURE in each of the two clusters in each of the two analyses (one for each stream). On the right, two large plots compare the estimated allele frequencies in matched inferred clusters in the two runs (Caigual - CA; Taylor - TY).

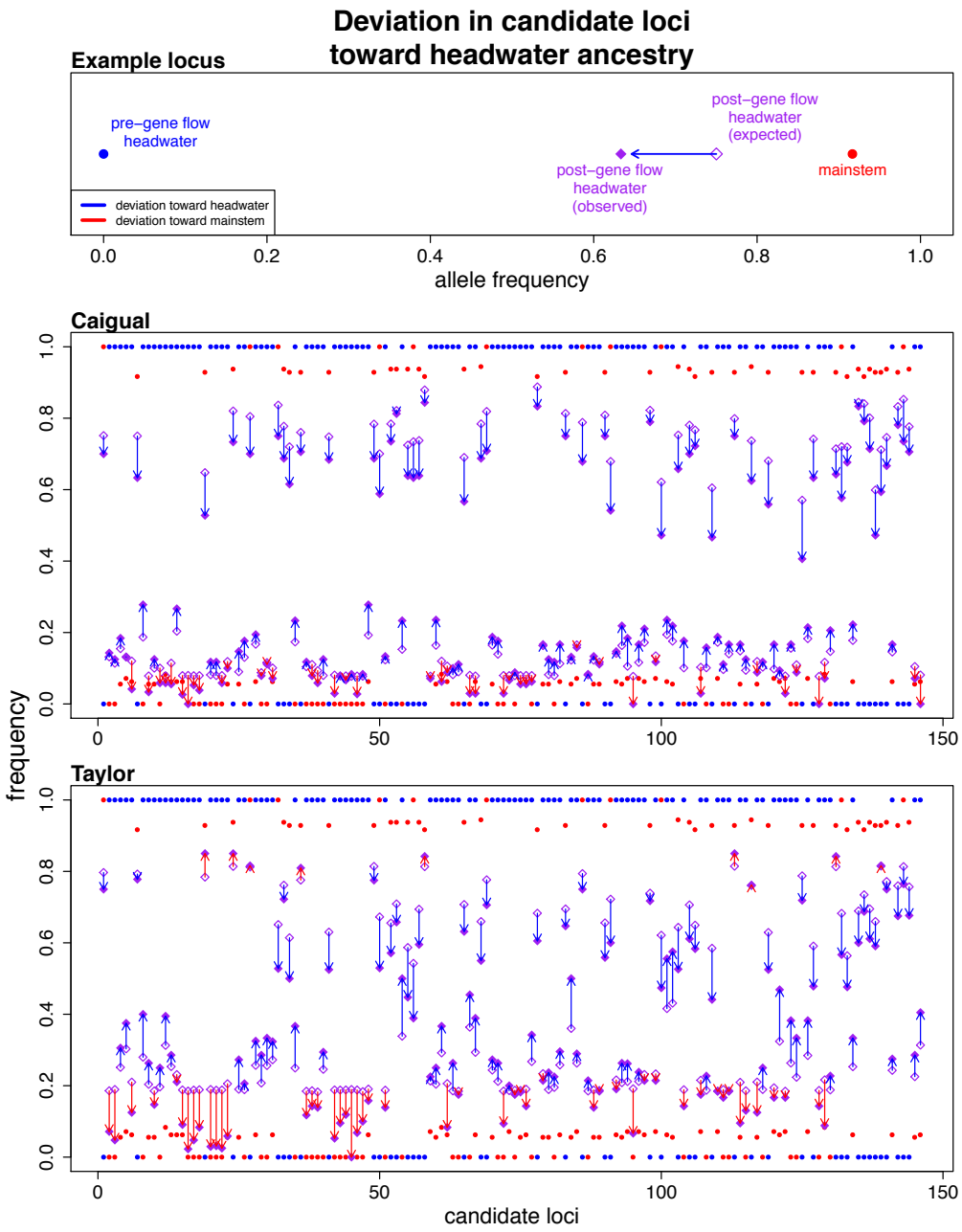

Figure S3. Deviation in candidate loci toward headwater ancestry. Related to STAR Methods. The top panel provides an example of ancestry-polarized deviation. In this example, the deviation between the observed and expected allele frequencies is in the direction of the ancestral pre-gene flow headwater frequency, so the ancestry-polarized deviation would take a positive value. The middle panel is a plot of ancestry deviations across all candidate loci in the Caigual population. The bottom panel is a plot of ancestry deviations across all candidate loci in the Taylor population.

Stream	**Model**		**Comparison**		
Caigual	***Name***	***Fixed***	***df***	***AICc***	***dAICc***
	mzinbC.0	Longevity~hindex	4	7001.1	131.2
	mzinbC.1	Longevity~hindex+hindex^2	5	6993.1	123.2
	mzinbC.2	Longevity~hindex+hindex^2+Sex	6	6891.4	21.5
	mzinbC.3	Longevity~(hindex+hindex^2)*Sex	8	6887.5	17.6
***	mzinbC.4	Longevity~(hindex+hindex^2)*Sex z~1	9	6869.9	0.0
	mzinbC.5	Longevity~(hindex+hindex^2)*Sex z~Sex	10	6870.8	0.9
	mzinbC.6	Longevity~(hindex+hindex^2)*Sex z~hindex+hindex^2	11	6873.4	3.5
	mzinbC.7	Longevity~hindex+hindex^2+Sex z~1	7	6873.5	3.6
	mzinbC.8	Longevity~hindex+hindex^2+Sex z~hindex+hindex^2	9	6875.1	5.2
	mzinbC.9	Longevity~hindex+hindex^2+Sex z~hindex+hindex^2+Sex	10	6876.8	6.9
Taylor	***Name***	***Fixed***	***df***	***AICc***	***dAICc***
	mzinbT.0	Longevity~hindex	4	5290.0	44.8
	mzinbT.1	Longevity~hindex+hindex^2	5	5284.1	38.9
	mzinbT.2	Longevity~hindex+hindex^2+Sex	6	5265.5	20.3
	mzinbT.3	Longevity~(hindex+hindex^2)*Sex	8	5269.0	23.8
	mzinbT.4	Longevity~(hindex+hindex^2)*Sex z~1	9	5271.0	25.8
	mzinbT.5	Longevity~(hindex+hindex^2)*Sex z~Sex	10	5271.4	26.2
	mzinbT.6	Longevity~(hindex+hindex^2)*Sex z~hindex+hindex^2	11	5249.1	3.9
	mzinbT.7	Longevity~hindex+hindex^2+Sex z~1	7	5267.5	22.3
	mzinbT.8	Longevity~hindex+hindex^2+Sex z~hindex+hindex^2	9	5245.5	0.3
***	mzinbT.9	Longevity~hindex+hindex^2+Sex z~hindex+hindex^2+Sex	10	5245.2	0.0

Table S1. AICc comparison for Longevity models. Related to Figure 2. All models included a random effect = 1|cohort. Other variables: z = zero-inflation sub-model; hindex = hybrid index; Sex = male/female. *** indicates the model selected as the best model for subsequent analyses.

Stream	Random effect	Variance	Std. Dev	Nobs	Ngroups
Taylor (mzinbT.9 in Table S1)	Cohort	0.3768	0.6139	1804	14
	Negative binomial over-dispersion parameter	0.889			
	Negative binomial mean model	**Estimate**	**Std. Error**	**z-value**	**Pr(>|z|)**
	Intercept	-0.142	0.195	-0.726	0.468
	hindex	3.583	0.562	6.371	<0.001
	hindex^2	-2.382	0.537	-4.435	<0.001
	Sex(male)	-0.451	0.089	-5.093	<0.001
	Zero-inflation model	**Estimate**	**Std. Error**	**z-value**	**Pr(>|z|)**
	Intercept	-3.940	1.062	-3.708	<0.001
	hindex	10.795	3.101	3.481	<0.001
	hindex^2	-8.392	2.423	-3.463	<0.001
	Sex(male)	-0.432	0.303	-1.425	0.154
Caigual (mzinbC.4 in Table S1)	**Random effect**	**Variance**	**Std. Dev**	**Nobs**	**Ngroups**
	Cohort	0.0450	0.212	1480	13
	Negative binomial over-dispersion parameter	1.34			
	Negative binomial mean model	**Estimate**	**Std. Error**	**z-value**	**Pr(>|z|)**
	Intercept	1.391	0.089	15.633	<0.001
	hindex	2.518	0.674	3.737	<0.001
	hindex^2	-2.134	1.018	-2.096	0.036
	Sex(male)	-0.592	0.081	-7.304	<0.001
	hindex:Sex	-0.283	0.788	-0.359	0.720
	hindex^2:Sex	-0.781	1.281	-0.610	0.542
	Zero-inflation model	**Estimate**	**Std. Error**	**z-value**	**Pr(>|z|)**
	Intercept	-1.907	0.203	-9.383	<0.001

Table S2. Detailed summaries of the best models of longevity by stream. Related to Figure 2.

Fitness component	Stream	Term	Location	95% CI
Longevity	Caigual	Mean, Females	0.590	(0.428, 1.000)
		Mean, Males	0.383	(0.302, 0.483)
	Taylor	Mean, both	0.752	(0.667, 0.931)
		Zero-inflation, both	0.643	(0.563, 0.728)
Lifetime reproductive success	Caigual	Mean	0.824	(0.663, 1.00)
	Taylor	Mean	0.678	(0.585, 0.946)
		Zero-inflation	0.073	(0.000, 0.309)

Table S3. **Position of quadratic maxima in longevity and lifetime reproductive success models (hybrid index values). Related to Figure 2.**

Stream	**Model**		**Comparison**		
Caigual	***Name***	***Fixed***	***df***	***AICc***	***dAICc***
	mzinbC.0	LRS~hindex	4	1697.6	12.9
***	mzinbC.1	LRS~hindex+hindex^2	5	1684.7	0.0
	mzinbC.2	LRS~(hindex+hindex^2)+Sex	6	1685.7	0.9
	mzinbC.3	LRS~(hindex+hindex^2)*Sex	8	1688.1	3.3
	mzinbC.4	LRS~hindex+hindex^2 z~1	6	1686.8	2.0
	mzinbC.5	LRS~hindex+hindex^2 z~hindex	7	1688.8	4.1
Taylor	***Name***	***Fixed***	***df***	***AICc***	***dAICc***
	mzinbT.0	LRS~hindex	4	3095.3	29.2
	mzinbT.1	LRS~hindex+hindex^2	5	3087.5	21.4
	mzinbT.2	LRS~(hindex+hindex^2)+Sex	6	3089.5	23.4
	mzinbT.3	LRS~(hindex+hindex^2)*Sex	8	3092.1	26.0
	mzinbT.4	LRS~hindex+hindex^2 z~1	6	3087.8	21.7
	mzinbT.5	LRS~hindex+hindex^2 z~hindex	7	3066.9	0.8
***	mzinbT.6	LRS~hindex+hindex^2 z~ hindex+hindex^2	8	3066.1	0.0
	mzinbT.7	LRS~hindex+hindex^2 z~hindex+hindex^2+Sex	9	3068.1	2.0

Table S4. **AICc comparison for lifetime reproductive success (LRS) models. Related to Figure 2.** All models included a random effect = 1|cohort. Other variables: z = zero-inflation sub-model; hindex = hybrid index; Sex = male/female. *** indicates the model selected as the best model for subsequent analyses.

Stream	Random effect	Variance	Std. Dev	Nobs	Ngroups
	Negative binomial over-dispersion parameter	0.65			
	Negative binomial conditional mean model	**Estimate**	**Std. Error**	**z-value**	**Pr(>|z|)**
	Intercept	0.586	0.257	2.281	0.023
	hindex	3.706	0.889	4.169	<0.001
	hindex^2	-2.732	0.828	3.300	<0.001
	Zero-inflation model	**Estimate**	**Std. Error**	**z-value**	**Pr(>|z|)**
	Intercept	0.664	0.243	2.728	0.006
	hindex	0.250	0.983	0.255	0.799
	hindex^2	-1.707	1.037	-1.646	0.100
Taylor (mzinbT.6 in Table S5)	Cohort	0.262	0.512	1110	11
Caigual (mzinbC.1 in Table S5)	**Random effect**	**Variance**	**Std. Dev**	**Nobs**	**Ngroups**
	Cohort	1.742	1.32	702	10
	Negative binomial over-dispersion parameter	0.453			
	Negative binomial mean model	**Estimate**	**Std. Error**	**z-value**	**Pr(>|z|)**
	Intercept	-1.209	0.438	-2.757	0.006
	hindex	9.691	1.100	8.807	<0.001
	hindex^2	-5.878	1.411	-4.166	<0.001

Table S5. **Detailed summaries of the best models of lifetime reproductive success by stream. Related to Figure 2.**