Supplementary Information

Specimen collection and processing

JT Garner collected 20 female *Megalonaias nervosa* from Pickwick Reservoir near Muscle Shoals, AL. Specimens were shipped in coolers to UNCC. Four specimens died in shipping and two others had limited response to perturbation on arrival. The other 14 were dissected and tissues were flash frozen in liquid nitrogen. Length (anterior to posterior), depth (left to right shell, measured near the umbo), and height (dorsal to ventral) were measured with digital calipers at the maximum span of each shell. Ring counts were used to estimate age (Table S1). DNA was extracted from tissues using a Qiagen DNEasy column based purification kit. Sequencing libraries were prepared in house using the TruSeq PCR-free library preparation kit and sequenced on 3 lanes of a HiSeq3000/4000 at Duke University Sequencing Core. All specimens later thawed in a -80 meltdown. Population specimens were intractable for RNA extraction and had very low yield for long molecule DNA extraction after thawing.

Genome Scaffolding and Re-annotation

A previously generated a reference genome is available with an N50 of 50 kb , sufficient to capture the majority of gene content in *M. nervosa* (Rogers et al., 2021). Linkage across the genome offers important information for population genetic models. To improve this assembly, we performed 21 phenol-chloroform extractions on adductor tissue for the reference specimen that had thawed in a -80 meltdown. We used bead-based cleanup to enrich for larger DNA fragments. We performed long read sequencing with 1 full run and 3 incomplete runs of Nanopore on a MinIon. Basecalling was performed using the Guppy 3.6.1-cuda basecaller on GPU nodes with 4 cores. Read sizes ranged from 200 bp to 329,099 kb.

We scaffolded the prior assembly from Rogers et al. (2021) in LINKS/1.8.7, which identifies uniquely mapping k-mers to anchor scaffolds using long reads or long inserts. Scaffolding was performed iteratively with subsets of reads for each run to manage memory usage. LINKS was run with distance thresholds of 2,000, 5,000, 10,000, and 20,000 bp using options -k 21 -t 5. This 2.6Gb improved reference assembly has an N50 of 125,303bp and maximum contig length of 992,038bp, far better performance for population genomics. 2.0 Gb out of 2.6 Gb of sequence on 15,643 scaffolds 50 kb or larger and 2.5Gb is on 33,101 scaffolds 10 kb or larger. Hence, the majority of the genome is in contigs large enough for population genetic analysis. BUSCO analysis v. 3.0.2 using the metazoan ortholog data set (Simão et al., 2015) shows 86.1% complete single copy orthologs, 2.0% duplicated orthologs, and 5.6% fragmented buscos. This assembly has 50% fewer fragmented orthologs than the previous assembly attempt (Rogers et al., 2021), but the number of orthologs found in whole or in part remains unchanged (only 6.3% missing). The number of duplicated orthologs remains unchanged. Some sequences removed as putative contamination in the initial assembly now scaffold with mollusc sequences, suggesting prior filtering may have aggressively filtered out short repetitive elements that had incidental matches with contaminant

genomes. Liftover from the initial *M. nervosa* genome assembly, which contains identical contig sequences, was performed using a BLASTn (Altschul et al., 1990). The mtDNA scaffold scaffold21472.30567.f21430Z30567 was identified using a BLASTn (Altschul et al., 1990) against *Quadrula quadrula* mitochondria sequence obtained from NCBI (accession NC 013658.1, downloaded Aug 15 2018).

We re-annotated sequences according to previous work (Rogers et al., 2021). We performed *de novo* transcriptome assembly in Oyster River Protocol (MacManes, 2018). We mapped transcripts to the reference using BLAT (Kent, 2002) and used these as hints for Augustus allowing for isoform detection. Augustus identified 263,584 putative gene sequences. We used Interproscan to identify conserved functional domains (Quevillon et al., 2005). Transposable elements were identified using Interproscan (Quevillon et al., 2005), tBLASTx (Altschul et al., 1990) against RepeatScout TEs, RepBase (Bao et al., 2015), and with RepDeNovo (Chu et al., 2016). These methods identified 22,269 TE-related transcripts, with multiple transcripts identified for some TEs. A total of 824 Mb of sequence matches with RepeatScout TEs in the reference genome, and 562 Mb matches with RepDeNovo repeats. These were removed from putative gene annotations. Orthologs and paralogs were identified in a BLASTp comparison at $E < 10^{-20}$ against *C. gigas* (Zhang et al., 2012). Genes with support from interproscan, blast comparisons with *C. gigas*, or from RNAseq data were kept in the final list of 64,086 putative genes. 64,086 genes encompassing 86,049 exons remain. In this improved reference genome, we identify 172 Cytochrome P450 genes, 146 ABC transporters, 168 Mitochondria-eating genes, 55 thioredoxins, 109 Hsp70 genes, 282 rhodopsins, 92 chitin metabolism genes, and 242 von Willebrand genes. This new assembly identifies even greater numbers of these highly proliferated gene families in *M. nervosa* than the previous version of the reference sequence even as the number of duplicated BUSCO orthologs decreased by 50%. Hence, we do not have any indication that duplicate gene content is decreasing as assembly quality improves.

A recent high quality reference assembly with greater contiguity (N50 of 2Mb) has recently become available for *Potamilus streckersonii* (Smith, 2021) that can be used for comparison with other Unionidae species. Using a reciprocal best hit BLASTn (Altschul et al., 1990) at an E-value of 10^{-10}, we identify 17862 1:1 orthologs between *M. nervosa* and *Potamilus streckersonii*. We used these to create a syntenic map of scaffolds between these two species where sequence divergence is too great for nucleotide alignment. While rearrangements may have occurred during divergence of these two species, the syntenic map can be used to anchor scaffolds identified as subjects of strong selective sweeps.

Genome alignment and Population Genetics

We aligned DNA for 13 samples and the Reference specimen using bwa aln (Li and Durbin, 2009). Sequencing depth was calculated using samtools depth. We identified SNPs using gatk haplotype caller under default parameters in a diploid model. We used SNPs to calculate π, θ_W (Watterson, 1975), and Tajima's D (Tajima, 1989) in 10 kb windows with a 1 kb slide. Polymorphic TE insertions were identified using paired end read mapping and blast hits to RepBase as well as scaffolding with RepDeNovo contigs. We estimated π, θ_W and

Tajima's D for each scaffold over 10kb in length using a window size of 10kb and a 1kb slide, excluding windows with less than 75% coverage.

Bayesian Changepoint statistics identify the most likely changepoints with posterior probabilities, and fits means to the blocks identified in the data using MCMC. Because *bcp* uses autocorellation to extract information about change points, only non-overlapping windows were used as input (10 kb regions with a 10 kb slide). Chagepoints with greater than 75% support were kept, and means of Tajima's D and π. Blocks within 20,000 bp were joined together to smooth stochastic switching. We used a threshold of $\pi < 0.0015$ for strong, recent selective sweeps, identifying 72 Mb of sequence show signals of extremely strong, extremely recent selective sweeps. Some 155 Mb on 1342 contigs is identified at a less stringent threshold of $\pi < 0.0020$.

Population Genetic Simulations and Modeling

We used msprime v1.1.1 (Baumdicker et al., 2022) and SLiM v3.7 (Messer, 2013; Haller and Messer, 2019; Haller et al., 2019) to perform coalescent simulations and forward simulations of population genetic diversity. We modeled genetic diversity π and Tajima's D under the following scenarios:

1. population stasis
2. population expansion
3. bottlenecks
4. strong selective sweeps
5. moderate selective sweeps
6. reduction in recombination

Coalescent Simulations For Demographic Changes

Using msprime v1.1.1 (Baumdicker et al., 2022) to perform coalescent simulations under different demographic changes. Tajima's D for *M. nervosa* is negative, with an average of -1.16. Genome wide shifts in Tajima's D toward negative values are a signal of population expansion. Populations of *M. nervosa* are common throughout the eastern United states. We simulate recent and ancient population expansion. We used a baseline $N_e = 300000$, consistent with estimates of population size from $\pi = 0.0054$ and $\mu = 5 \times 10^{-9}$. Data were parsed in 10kb windows. Coalescent and forward population genetic simulations require non-overlapping generations, which can be applied without severe loss of information (Baumdicker et al., 2022; Messer, 2013; Haller and Messer, 2019; Haller et al., 2019).

A pattern of 10-fold recent population growth alone does not shift Tajima's D sufficiently to match empirical observations (Figure S5). Populations have insufficient time to accumulate new mutations that contribute to rare alleles after expansion. However, a 12X

population expansion beginning roughly 5000 generations ago can produce a shift in diversity π consistent with empirical observations. Both a sudden expansion at a rate of 25% growth over 10 generations or prolonged population growth of 0.05% per year over 5000 years can produce comparable shifts in Tajima's D to approximately -1.0, with genetic diversity consistent with observations. Population growth beginning 1000 generations in the past produces a lesser shift in Tajima's D to 0.5. At 10,000 generations, the shift in Tajima's D is greater than observations in the data, at nearly -1.4.

The timing of this ancient population expansion will depend on generation time. *M. nervosa* has a minimum age of reproduction at 4 years old, but typically reaches sexual maturity at age 10 (Woody and Holland-Bartels, 1993). This minimum of 4 years can be used to establish the absolute limits for shortest timescales of adaptation in recent time of 25 generations since the damming of the Tennessee River. The maximum lifespan is 54 years, though 40 year old individuals may be more typical (Haag, 2012). Using the geometric mean of minimum sexual maturity and maximum lifespan, we obtain 15 years, but excluding outliers the geometric mean generation time would be 20 years. These simulations would be consistent with a range expansion in prehistoric time likely 75,000 years-100,000 years ago, depending on generation times of 15-20 years. This ancient population expansion is consistent with a species range expansion at the end of the last glaciation as greater habitat became available after glaciers receded as seen in other species (Keogh et al., 2021; Elderkin et al., 2007). Differences in generation time might yield different scaling from time in generations in population genetic simulations to historical timescales. Pervasive selection is also known to affect diversity at linked sites (Sella et al., 2009) and could contribute to observed shifts away from equilibrium as well.

Historical and archaeological data are incompatible with substantial modern bottlenecks. *Megalonaias* from different regions show little differentiation across geographic regions (Pfeiffer et al., 2018). Nevertheless, we still simulated a discrete 10 generation bottleneck representing a 5-fold population reduction (Figure S5). Genetic diversity is largely unchanged, as finite samples taken from a larger population reflect equilibrium genetic diversity originally present. Such results contrast with the effects of ancient bottlenecks which alter Tajima's D as new mutations appear in populations over many generations and contribute to an excess of lower frequency alleles prior to reaching equilibrium. Neither ancient nor modern bottlenecks could produce reductions in diversity similar to those observed in candidate regions for selective sweeps. While bottlenecks do increase the variance in π, we do not observe reductions below $\pi = 0.0025$ in high recombination ($r = 5 \times 10^{-8}$) or low recombination ($r = 5 \times 10^{-9}$) regions. Thus, even implausible cases of strong bottlenecks that escaped historical observation cannot explain the sweep-like signals observed in natural populations.

Forward Simulations for Selection

We then used SLiM v. 3.7 to perform forward simulations with tree sequence recording, then used recapitation to add neutral mutations onto simulated trees. We used simulation output to evaluate the impacts of very strong selection on π and Tajima's D. We simulated a 10

Mb region for a population of 300,000 individuals setting recombination at $r = 5 \times 10^{-8}$. For simplicity and runtime we do not simulate full population growth scenarios implemented in faster coalescent simulations. We first modeled population diversity for this scenario without selection to establish baseline diversity levels (Figure). We then added a single new mutation favored by selection at 5 Mb. We used a fully dominant mutation ($h = 1.0$) with very strong selection s=100, s=10, and s=0.1 under the SLiM fitness model of 1, $(1 + hs)$, $(1 + s)$ (Table S4). Note that these selection coefficients correspond to s=0.99 and s=0.91 in relative fitness models using 1, $(1 - hs)$ and $(1 - s)$. SLiM does not allow for simulation of strictly lethal selective regimes, but these selective coefficients of $s = 10$ are consistent with up to 100-fold greater survival for individuals carrying an adaptive allele. Given potentially lethal effects of pollution, pesticide use, host fish extinction, and other challenges facing Unionidae, extreme selective regimes are plausible and relevant for this species.

With strong selection (s=10, s=100) genetic diversity drops 10-fold for 1-2 Mb in the central region around the selected locus in less than 20 generations, with more pronounced effects at 50 generations (Figure S6-S10). Using simulated data as input for BCP, we can model the posterior mean and standard deviation for simulations. These posterior distributions are used as models to establish goodness of fit tests compared with empirical data from scaffold 22 (a 700kb contiguous scaffold from the largest sweep, Figure 4). Models with $s = 100$ and $r = 5 \times 10^{-8}$ are significantly better fit than $s = 10$, $s = 0.1$, or $r = 5 \times 10^{-9}$ ($\chi^2 > 40$ $df = 1$ or $df = 2$). We attempted to simulate selection at multiple sites, placing 3 selected sites 10kb apart. In 1000 replicate simulations, none could produce sweeps on all 3 loci, due to interference. Hence, multilocus selection is not a plausible explanation for the data.

In simulations, genetic diversity adjacent to the sweep may be depressed to roughly half background levels in the 5 Mb region around the sweep. However, diversity returns to normal values near the edges of the 10 Mb region as recombination breaks linked haploytpes around the selected locus (Figure S6-S10). At 15 generations, the reduction in diversity is observable but more modest (Figure S8, Figure S11). Selection on standing variation at frequencies above $p = 0.01$ would establish deterministic sweeps even more quickly (Hermisson and Pennings, 2005), but is not easy to simulate (as discussed thoroughly in the SLiM Manual). Under recombination reduction to $r = 1 \times 10^{-8}$ or $r = 5 \times 10^{-9}$, diversity is still reduced around the selected locus, but background genetic diversity is severely depressed across the entire 10 Mb region, inconsistent with empirical data (e.g. Figure S13). Hence, we do not believe that pervasive recombination suppression is compatible with data available for *M. nervosa*.

Under lower selection coefficients ($s = 0.1$) and normal recombination ($r = 5 \times 10^{-8}$), the selected allele was lost due to stochastic fluctuations affecting a singleton allele in the first 4 out of 5 initial simulation attempts, consistent with theory (Hermisson and Pennings, 2005). When a deterministic sweep did establish, these simulations began to show sweep-like signals around 300 generations, but not by 200 generations. Genetic diversity approaches zero near the surrounding locus, but produced a classic v-shaped pattern as recombination breaks apart linkage during the sweep (Figure S18-S20). With reduced recombination ($r = 5 \times 10^{-9}$) 19

out of 20 simulations resulted in loss of the allele. These simulations with mild selection do not produce the extensive megabase sized tracts with reduced in genetic diversity that we observe in empirical data. Hence, selective regimes with $s = 0.1$ are not sufficient to produce the extreme population genomic signals we have observed, even over long timescales.

In light of these simulations, we suggest that scenarios of very strong selection are sufficient and necessary to explain reductions in genetic diversity observed in the natural population of *M. nervosa*.

References

Ahlstedt, S. A. and McDonough, T. A. (1992). Quantitative evaluation of commercial mussel populations in the Tennessee river portion of Wheeler Reservoir, Alabama. In *Conservation and management of freshwater mussels. Proceedings of a UMRCC symposium*, pages 12–14.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410.

Aminetzach, Y. T., Macpherson, J. M., and Petrov, D. A. (2005). Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*. *Science*, 309(5735):764–767.

Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, 6(1):11.

Baumdicker, F., Bisschop, G., Goldstein, D., Gower, G., Ragsdale, A. P., Tsambos, G., Zhu, S., Eldon, B., Ellerman, E. C., Galloway, J. G., et al. (2022). Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, 220(3):iyab229.

Bennetzen, J. L. (2000). Transposable element contributions to plant gene and genome evolution. *Plant Molecular Biology*, 42(1):251–269.

Bolotov, I. N., Kondakov, A. V., Vikhrev, I. V., Aksenova, O. V., Bespalaya, Y. V., Gofarov, M. Y., Kolosova, Y. S., Konopleva, E. S., Spitsyn, V. M., Tanmuangpak, K., et al. (2017). Ancient river inference explains exceptional oriental freshwater mussel radiations. *Scientific Reports*, 7(1):1–14.

Breton, S., Beaupre, H. D., Stewart, D. T., Hoeh, W. R., and Blier, P. U. (2007). The unusual system of doubly uniparental inheritance of mtDNA: isnÕt one enough? *Trends in genetics*, 23(9):465–474.

Cahn, A. R. (1936). *The Molluscan Fauna of the Clinch River Below Norris Dam Upon the Completion of that Structure*. Tennessee Valley Authority.

Campbell, D. C., Serb, J. M., Buhay, J. E., Roe, K. J., Minton, R. L., and Lydeard, C. (2005). Phylogeny of North American amblemines (*Bivalvia, Unionoida*): prodigious polyphyly proves pervasive across genera. *Invertebrate Biology*, 124(2):131–164.

Chen, S. and Corces, V. G. (2001). The *gypsy* insulator of *Drosophila* affects chromatin structure in a directional manner. *Genetics*, 159(4):1649–1658.

Chu, C., Nielsen, R., and Wu, Y. (2016). REPdenovo: inferring *de novo* repeat motifs from short sequence reads. *PloS One*, 11(3):e0150719.

Conant, G. C. and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nature Reviews Genetics*, 9(12):938–950.

Cosby, R. L., Chang, N.-C., and Feschotte, C. (2019). Host–transposon interactions: conflict, cooperation, and cooption. *Genes & Development*, 33(17-18):1098–1116.

Cridland, J. M., Macdonald, S. J., Long, A. D., and Thornton, K. R. (2013). Abundance and distribution of transposable elements in two *Drosophila* QTL mapping resources. *Molecular Biology and Evolution*, 30(10):2311–2327.

Crouch, N. M., Edie, S. M., Collins, K. S., Bieler, R., and Jablonski, D. (2021). Calibrating phylogenies assuming bifurcation or budding alters inferred macroevolutionary dynamics in a densely sampled phylogeny of bivalve families. *Proceedings of the Royal Society B*, 288(1964):20212178.

Darwin, C. (1878). Transplantation of shells. *Nature:International Weekly Journal of Science*, (448).

Des Marais, D. L. and Rausher, M. D. (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature*, 454(7205):762–765.

Dubin, M. J., Scheid, O. M., and Becker, C. (2018). Transposons: a blessing curse. *Current Opinion in Plant Biology*, 42:23–29.

Elderkin, C., Christian, A., Vaughn, C., Metcalfe-Smith, J., and Berg, D. (2007). Population genetics of the freshwater mussel, *Amblema plicata* (say 1817)(Bivalvia: Unionidae): evidence of high dispersal and post-glacial colonization. *Conservation Genetics*, 8(2):355–372.

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, 29(1):51–63.

Emerson, J., Cardoso-Moreira, M., Borevitz, J. O., and Long, M. (2008). Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science*, 320(5883):1629–1631.

Erdman, C., Emerson, J. W., et al. (2007). *bcp*: an R package for performing a Bayesian analysis of change point problems. *Journal of Statistical Software*, 23(3):1–13.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics*, 9(5):397.

Gao, X., Hou, Y., Ebina, H., Levin, H. L., and Voytas, D. F. (2008). Chromodomains direct integration of retrotransposons to heterochromatin. *Genome research*, 18(3):359–369.

Garner, J. and McGregor, S. (2001). Current status of freshwater mussels (Unionidae, Margaritiferidae) in the Muscle Shoals area of Tennessee River in Alabama (Muscle Shoals revisited again). *American Malacological Bulletin*, 16(1-2):155–170.

Gause, M., Morcillo, P., and Dorsett, D. (2001). Insulation of enhancer-promoter communication by a *gypsy* transposon insert in the *Drosophila cut* gene: cooperation between suppressor of hairy-wing and modifier of mdg4 proteins. *Molecular and Cellular Biology*, 21(14):4807–4817.

Goldman, N. and Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, 11(5):725–736.

Haag, W. R. (2012). *North American freshwater mussels: natural history, ecology, and conservation*. Cambridge University Press.

Haag, W. R. (2013). The role of fecundity and reproductive effort in defining life-history strategies of North American freshwater mussels. *Biological Reviews*, 88(3):745–766.

Haag, W. R. and Williams, J. D. (2014). Biodiversity on the brink: an assessment of conservation strategies for North American freshwater mussels. *Hydrobiologia*, 735(1):45–60.

Haggerty, T. M., Garner, J. T., and Rogers, R. L. (2005). Reproductive phenology in *Megalonaias nervosa (Bivalvia: Unionidae)* in Wheeler Reservoir, Tennessee River, Alabama, USA. *Hydrobiologia*, 539(1):131–136.

Haller, B. C., Galloway, J., Kelleher, J., Messer, P. W., and Ralph, P. L. (2019). Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Molecular ecology resources*, 19(2):552–566.

Haller, B. C. and Messer, P. W. (2019). SLiM 3: forward genetic simulations beyond the Wright-Fisher model. *Molecular Biology and Evolution*, 36(3):632–637.

Han, M. V., Demuth, J. P., McGrath, C. L., Casola, C., and Hahn, M. W. (2009). Adaptive evolution of young gene duplicates in mammals. *Genome Research*, 19(5):859–867.

Hartl, D. L. (2020). *A Primer of Population Genetics and Genomics*. Oxford University Press.

Hermisson, J. and Pennings, P. S. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4):2335–2352.

Hu, Z., Song, H., Feng, J., Zhou, C., Yang, M.-J., Shi, P., Yu, Z.-L., Li, Y.-R., Guo, Y.-J., Li, H.-Z., and Zhang, T. (2022). Massive heat shock protein 70 genes expansion and transcriptional signatures uncover hard clam adaptations to heat and hypoxia. *Frontiers in Marine Science*, 9.

Isom, B. G., Yokley, P., and Gooch, C. H. (1973). Mussels of the Elk River Basin in Alabama and Tennessee-1965-1967. *American Midland Naturalist*, pages 437–442.

Jones, C. D. and Begun, D. J. (2005). Parallel evolution of chimeric fusion genes. *Proceedings of the National Academy of Sciences*, 102(32):11373–11378.

Kaplan, N. L., Hudson, R. R., and Langley, C. H. (1989). The" hitchhiking effect" revisited. *Genetics*, 123(4):887–899.

Kent, W. J. (2002). BLAT the BLAST-like alignment tool. *Genome Research*, 12(4):656–664.

Keogh, S. M., Johnson, N. A., Williams, J. D., Randklev, C. R., and Simons, A. M. (2021). Gulf coast vicariance shapes phylogeographic history of a north american freshwater mussel species complex. *Journal of Biogeography*, 48(5):1138–1152.

Kongim, B., Sutcharit, C., and Panha, S. (2015). Cytotaxonomy of unionid freshwater mussels (unionoida, unionidae) from northeastern thailand with description of a new species. *ZooKeys*, (514):93.

Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. (2017). Timetree: a resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, 34(7):1812–1819.

Lake, P. S., Palmer, M. A., Biro, P., Cole, J., Covich, A. P., Dahm, C., Gibert, J., Goedkoop, W., Martens, K., and Verhoeven, J. (2000). Global change and the biodiversity of freshwater ecosystems: Impacts on linkages between above-sediment and sediment biota: All forms of anthropogenic disturbance—changes in land use, biogeochemical processes, or biotic addition or loss—not only damage the biota of freshwater sediments but also disrupt the linkages between above-sediment and sediment-dwelling biota. *BioScience*, 50(12):1099–1107.

Langley, C. H., Stevens, K., Cardeno, C., Lee, Y. C. G., Schrider, D. R., Pool, J. E., Langley, S. A., Suarez, C., Corbett-Detig, R. B., Kolaczkowski, B., et al. (2012). Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics*, 192(2):533–598.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14):1754–1760.

Liu, H.-P., Mitton, J. B., and Wu, S.-K. (1996). Paternal mitochondrial DNA differentiation far exceeds maternal mitochondrial DNA and allozyme differentiation in the freshwater mussel, *Anodonta grandis grandis*. *Evolution*, 50(2):952–957.

Long, M. and Langley, C. H. (1993). Natural selection and the origin of jingwei, a chimeric processed functional gene in drosophila. *Science*, 260(5104):91–95.

Lynch, M. (2007). *The Origins of Genome Architecture*, page 494. Sinauer Associates, Sunderland, Mass.

Mackay, T. F., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., Casillas, S., Han, Y., Magwire, M. M., Cridland, J. M., et al. (2012). The *Drosophila melanogaster* genetic reference panel. *Nature*, 482(7384):173–178.

MacManes, M. D. (2018). The Oyster River Protocol: a multi-assembler and kmer approach for *de novo* transcriptome assembly. *PeerJ*, 6:e5428.

Matuszewski, S., Hildebrandt, M. E., Achaz, G., and Jensen, J. D. (2018). Coalescent processes with skewed offspring distributions and nonequilibrium demography. *Genetics*, 208(1):323–338.

Maynard Smith, J. (1971). What use is sex? *Journal of Theoretical Biology*, 30(2):319–335.

Messer, P. W. (2013). Slim: simulating evolution with selection and linkage. *Genetics*, 194(4):1037–1039.

Modesto, V., Ilarri, M., Souza, A. T., Lopes-Lima, M., Douda, K., Clavero, M., and Sousa, R. (2018). Fish and mussels: importance of fish for freshwater mussel conservation. *Fish and Fisheries*, 19(2):244–259.

Nelson, D. R., Zeldin, D. C., Hoffman, S. M., Maltais, L. J., Wain, H. M., and Nebert, D. W. (2004). Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. *Pharmacogenetics and Genomics*, 14(1):1–18.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.*, 39:197–218.

Ohno, S. (1970). *Evolution by Gene Duplication*. Springer-Verlag, Berlin, New York.

Orr, H. A. (2006). The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation. *Journal of theoretical biology*, 238(2):279–285.

Ortmann, A. (1924). Mussel Shoals. *Science*, 60(1564):565–566.

Patterson, M. A., Mair, R. A., Eckert, N. L., Gatenby, C. M., Brady, T., Jones, J. W., Simmons, B. R., and Devers, J. L. (2018). *Freshwater mussel propagation for restoration*. Cambridge University Press.

Pfeiffer, J. M., Breinholt, J. W., and Page, L. M. (2019). Unioverse: A phylogenomic resource for reconstructing the evolution of freshwater mussels (Bivalvia, Unionoida). *Molecular Phylogenetics and Evolution*, 137:114–126.

Pfeiffer, J. M., Sharpe, A. E., Johnson, N. A., Emery, K. F., and Page, L. M. (2018). Molecular phylogeny of the Nearctic and Mesoamerican freshwater mussel genus *Megalonaias*. *Hydrobiologia*, 811(1):139–151.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Research*, 33(suppl_2):W116–W120.

Ranz, J. M. and Parsch, J. (2012). Newly evolved genes: moving from comparative genomics to functional studies in model systems: how important is genetic novelty for species adaptation and diversification? *Bioessays*, 34(6):477–483.

Régnier, C., Fontaine, B., and Bouchet, P. (2009). Not knowing, not recording, not listing: numerous unnoticed mollusk extinctions. *Conservation Biology*, 23(5):1214–1221.

Renaut, S., Guerra, D., Hoeh, W. R., Stewart, D. T., Bogan, A. E., Ghiselli, F., Milani, L., Passamonti, M., and Breton, S. (2018). Genome survey of the freshwater mussel *Venustaconcha ellipsiformis (Bivalvia: Unionida)* using a hybrid *de novo* assembly approach. *Genome Biology and Evolution*, 10(7):1637–1646.

Rogers, R. L., Bedford, T., Lyons, A. M., and Hartl, D. L. (2010). Adaptive impact of the chimeric gene *Quetzalcoatl* in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, 107(24):10943–10948.

Rogers, R. L., Cridland, J. M., Shao, L., Hu, T. T., Andolfatto, P., and Thornton, K. R. (2015). Tandem duplications and the limits of natural selection in *Drosophila yakuba* and *Drosophila simulans*. *PLoS One*, 10(7):e0132184.

Rogers, R. L., Grizzard, S. L., Titus-McQuillan, J. E., Bockrath, K., Patel, S., Wares, J. P., Garner, J. T., and Moore, C. C. (2021). Gene family amplification facilitates adaptation in freshwater unionid bivalve *Megalonaias nervosa*. *Molecular Ecology*, 30(5):1155–1173.

Rogers, R. L. and Hartl, D. L. (2011). Chimeric genes as a source of rapid evolution in *Drosophila melanogaster*. *Molecular Biology and Evolution*, 29(2):517–529.

Rogers, R. L., Shao, L., Sanjak, J. S., Andolfatto, P., and Thornton, K. R. (2014). Revised annotations, sex-biased expression, and lineage-specific genes in the *Drosophila melanogaster* group. *G3: Genes, Genomes, Genetics*, 4(12):2345–2351.

Rogers, R. L., Shao, L., and Thornton, K. R. (2017). Tandem duplications lead to novel expression patterns through exon shuffling in *Drosophila yakuba*. *PLoS Genetics*, 13(5):e1006795.

Schaack, S., Gilbert, C., and Feschotte, C. (2010). Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends in Ecology & Evolution*, 25(9):537–546.

Schmidt, J. M., Good, R. T., Appleton, B., Sherrard, J., Raymant, G. C., Bogwitz, M. R., Martin, J., Daborn, P. J., Goddard, M. E., Batterham, P., et al. (2010). Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus. *PLoS Genetics*, 6(6):e1000998.

Schrider, D. R. and Hahn, M. W. (2010). Gene copy-number polymorphism in nature. *Proceedings of the Royal Society B: Biological Sciences*, 277(1698):3213–3221.

Sella, G., Petrov, D. A., Przeworski, M., and Andolfatto, P. (2009). Pervasive natural selection in the *Drosophila* genome? *PLoS Genetics*, 5(6):e1000495.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19):3210–3212.

Smith, C. H. (2021). A high-quality reference genome for a parasitic bivalve with doubly uniparental inheritance (Bivalvia: unionida). *Genome Biology and Evolution*, 13(3):evab029.

Stewart, N. B. and Rogers, R. L. (2019). Chromosomal rearrangements as a source of new gene formation in *Drosophila yakuba*. *PLoS Genetics*, 15(9):e1008314.

Strayer, D. L. (1999). Effects of alien species on freshwater mollusks in North America. *Journal of the North American Benthological Society*, 18(1):74–98.

Strayer, D. L., Downing, J. A., Haag, W. R., King, T. L., Layzer, J. B., Newton, T. J., and Nichols, J. S. (2004). Changing perspectives on pearly mussels, North America's most imperiled animals. *BioScience*, 54(5):429–439.

Sun, J., Zhang, Y., Xu, T., Zhang, Y., Mu, H., Zhang, Y., Lan, Y., Fields, C. J., Hui, J. H. L., Zhang, W., et al. (2017). Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nature Ecology & Evolution*, 1(5):1–7.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3):585–595.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22):4673–4680.

Van Etten, M., Lee, K. M., Chang, S.-M., and Baucom, R. S. (2020). Parallel and nonparallel genomic responses contribute to herbicide resistance in *Ipomoea purpurea*, a common agricultural weed. *PLoS Genetics*, 16(2):e1008593.

Watterson, G. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, 7(2):256–276.

Wen, H. B., Cao, Z. M., Hua, D., Xu, P., Ma, X. Y., Jin, W., Yuan, X. H., and Gu, R. B. (2017). The complete maternally and paternally inherited mitochondrial genomes of a freshwater mussel *Potamilus alatus (Bivalvia: Unionidae)*. *PloS One*, 12(1):e0169749.

Williams, J. D., Bogan, A. E., Garner, J. T., et al. (2008). *Freshwater mussels of Alabama and the Mobile basin in Georgia, Mississippi, and Tennessee*. University of Alabama Press.

Williams, J. D., Warren Jr, M. L., Cummings, K. S., Harris, J. L., and Neves, R. J. (1993). Conservation status of freshwater mussels of the United States and Canada. *Fisheries*, 18(9):6–22.

Woodside, M. D. (2004). *Water quality in the lower Tennessee River Basin, Tennessee, Alabama, Kentucky, Mississippi, and Georgia, 1999-2001*, volume 1233. US Geological Survey.

Woody, C. A. and Holland-Bartels, L. (1993). Reproductive characteristics of a population of the washboard mussel *Megalonaias nervosa* (Rafinesque 1820) in the upper Mississippi River. *Journal of Freshwater Ecology*, 8(1):57–66.

Yang, S., Arguello, J. R., Li, X., Ding, Y., Zhou, Q., Chen, Y., Zhang, Y., Zhao, R., Brunet, F., Peng, L., et al. (2008). Repetitive element-mediated recombination as a mechanism for new gene origination in *Drosophila*. *PLoS Genetics*, 4(1):e3.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*, 13(5):555–556.

Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., et al. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, 490(7418):49–54.

Table S1: Specimens of *M. nervosa* from Pickwick Lake Oct 2018

Number	Length (cm)	Depth (cm)	Height (cm)	Ring Count	Gravid	Sequencing Status
1	12.3	5.2	8.4	13	Yes	Sequenced
2	10.0	4.4	7.4	11	Yes	Sequenced
3	15.8	5.7	11.0	11	Yes	Reference Genome
4	10.9	4.7	8.2	10	Yes	Sequenced
5	10.6	4.6	7.9	8	Yes	Sequenced
6	11.4	4.8	8.4	9	Yes	Sequenced
7	9.9	4.1	7.3	7	Yes	Sequenced
8	12.4	4.6	9.0	10	No	Sequenced
9	10.4	3.8	7.1	6	Yes	Sequenced
10	10.4	4.4	7.5	11	No	Sequenced
11	12.1	4.9	8.5	12	Yes	Sequenced
12	10.7	4.2	7.8	10	Yes	Sequenced
13	10.4	4.2	8.1	8	Yes	Sequenced
14	10.9	4.7	8.6	8	Yes	Not sequenced
15	11.9	5.8	11.9	11	Yes	Not sequenced
16	8.5	2.5	5.0	8	No	Sequenced

Table S2: Diversity Metrics

π	mean	0.005447792
	median	0.005031502
	st. dev.	0.002679202
Tajima's D	mean	-1.160924
	median	-1.220574
	st. dev.	0.8012326

Table S3: Demographic simulations

Scenario	growth start	growth rate	duration	Ratio
Ancient Expansion	2000	0.25	10	10X
	5000*	0.25	10	10X
	10000	0.25	10	10X
Continuous Expansion	2000	0.00125	-	12X
	5000*	0.005	-	12X
	10000	0.00025	-	12X
Modern Bottleneck	10	instant	10	5X

*Best fit models for for whole genome background diversity.

Table S4: Simulation parameters

Recombination r	Selection s
5×10^{-8}	0.1
	10
	100*
1×10^{-8}	0.1
	10
	100
5×10^{-9}	0.1
	10
	100

*Best fit model for for scaffold 22 test data.

Table S5: Genes in Selective Sweeps

Apoptosis inhibitor	18	g7640, g37204, g40603, g9485, g65154, g13778, g83829, g15106, g2052, g2174, g103700, g19660, g2931, g23549, g130064, g26500; g77619 (65 kDa Yes-associated protein); g28882 (APAF1-interacting protein)
Stress Tolerance	17	g75456, g75457, g81615, g81619, g97141, g110382, g110383 (Hsp70); g49053, g72756, g85437, g100073 (Early growth response protein); g15855, g96078,g96079 (Carnitine metabolism); g115821, g115825, g115827 (Ectonucleoside triphosphate diphosphohydrolase 7)
Shell Formation	10	g41573 (Chitin synthase 3); g32, g36, g48649 (sodium bicarbonate exchanger); g2794 (Perlucin); g37568 (Anoctamin); g56763, g72205, g87285, g3813 (Calmodulin)
Development	9	g8580 (Bone morphogenetic protein 3);g87703, g73099, g75719, g104521, g120171, g30182 (Homeobox); g65282, g65283 (Wnt)
Detox	8	g62671, g62674 (Cytochrome b5 reductase 4); g59044 (Rieske domain-containing protein); g81499 (Cytochrome P450); g13536 (Dimethylaniline monooxygenase); g43545 (Flavin-containing monooxygenase); g77388 (Peroxiredoxin); g73090 (Protoporphyrinogen oxidase - target of herbicides)
Mitochondria	8	g22770 (Atlastin); g13794, g13795, (ADP,ATP carrier protein 1,); g84734(ATP synthase subunit delta); g15674 (Fumarylacetoacetate hydrolase); g74352, g74355 (Isocitrate dehydrogenase); g105152 (Mitochondrial fission factor), g39152 (Nuclease EXOG, mitochondrial)
Parasite-Host	6	g87422 (Plasminogen); g84442 (Pleckstrin); g100841 (Fibronectin); g83864 (Multimerin-1); g57648, g104196 (von Willebrand factor)
DNA repair	6	g2915, g2920, g2909 (DNA mismatch repair protein); g35257, g35258, g74360 (Werner syndrome helicase)

Table S6: Polymorphic TE Insertions in *M. nervosa*

Polinton	1701
Gypsy	1570
RTE	171
L2	138
Nimb	135
Penelope	130
Daphne	80
Neptune	78
CR1	71
SAT	69
All Other	828
Total	4971

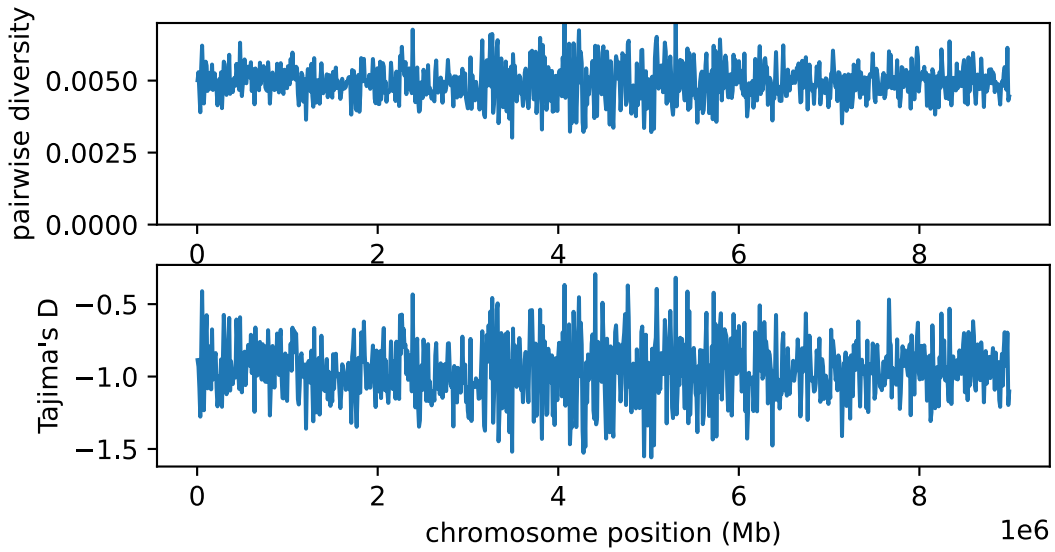

Figure S1: Genetic diversity π and Tajima's D under simulated 12-fold population growth at a rate of 1.0005 per generation over 5000 generations. Simulations show highly recombining regions ($r = 5 \times 10^{-8}$) from 0-3 Mb and 6-9 Mb, and lowly recombining regions ($r = 5 \times 10^{-9}$) from 3-6 Mb.

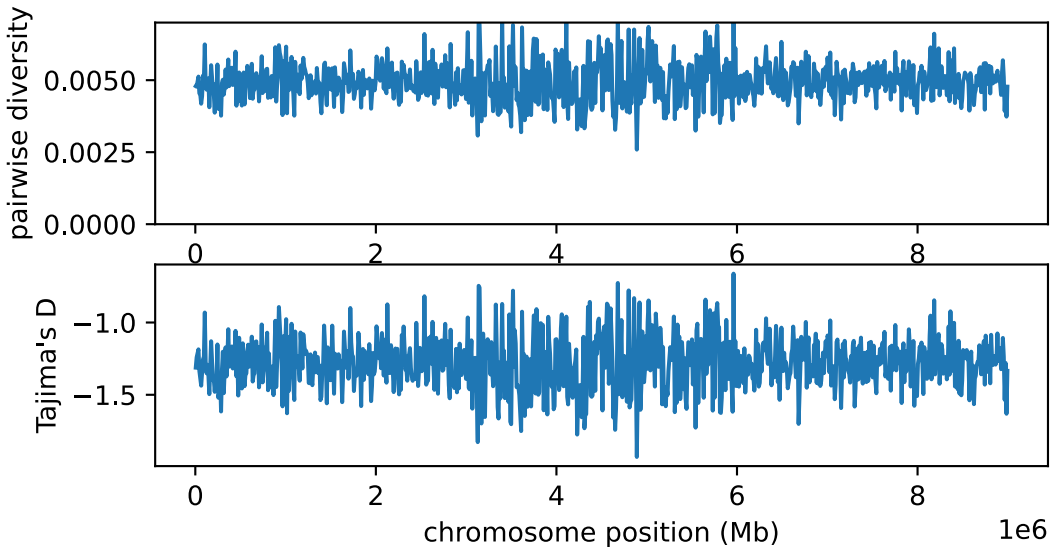

Figure S2: Genetic diversity π and Tajima's D under simulated 12-fold population growth at a rate of 1.00025 per generation over 10000 generations. Simulations show highly recombining regions ($r = 5 \times 10^{-8}$) from 0-3 Mb and 6-9 Mb, and lowly recombining regions ($r = 5 \times 10^{-9}$) from 3-6 Mb.

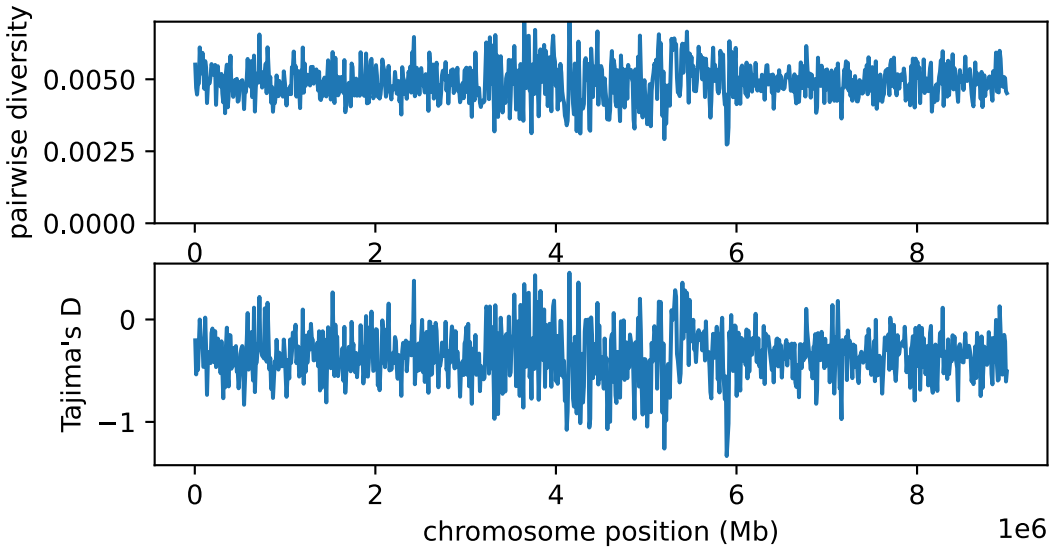

Figure S3: Genetic diversity π and Tajima's D under simulated 12-fold population growth at a rate of 1.0025 per generation over 1000 generations. Simulations show highly recombining regions ($r = 5 \times 10^{-8}$) from 0-3 Mb and 6-9 Mb, and lowly recombining regions ($r = 5 \times 10^{-9}$) from 3-6 Mb.

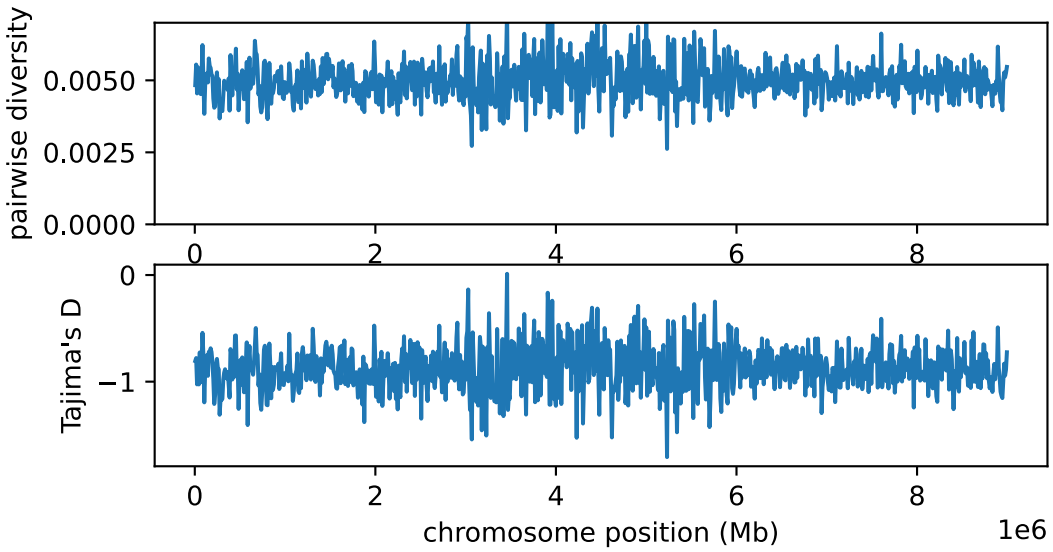

Figure S4: Genetic diversity π and Tajima's D under simulated 12-fold population growth at a rate of 1.0005 per generation over 1000 generations with a 5-fold population bottleneck 10 generations in the past followed by population recovery. Simulations show highly recombining regions ($r = 5 \times 10^{-8}$) from 0-3 Mb and 6-9 Mb, and lowly recombining regions ($r = 5 \times 10^{-9}$) from 3-6 Mb.

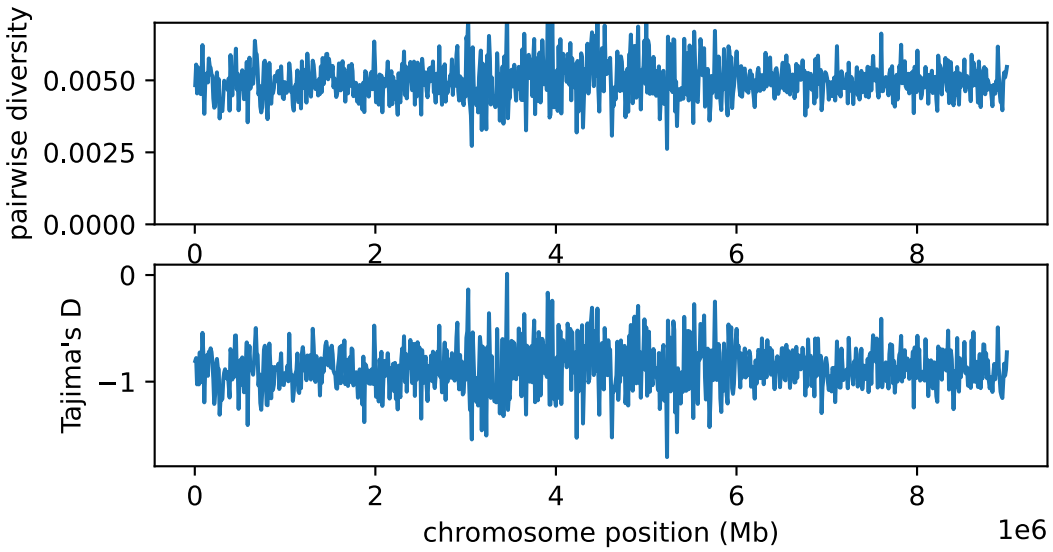

Figure S5: Genetic diversity π and Tajima's D under simulated 12-fold population growth at a rate of 1.005 per generation over 5000 generations with a 5-fold population bottleneck for 10 generations, 20-10 generations in the past followed by population recovery. Simulations show highly recombining regions ($r = 5 \times 10^{-8}$) from 0-3 Mb and 6-9 Mb, and lowly recombining regions ($r = 5 \times 10^{-9}$) from 3-6 Mb.

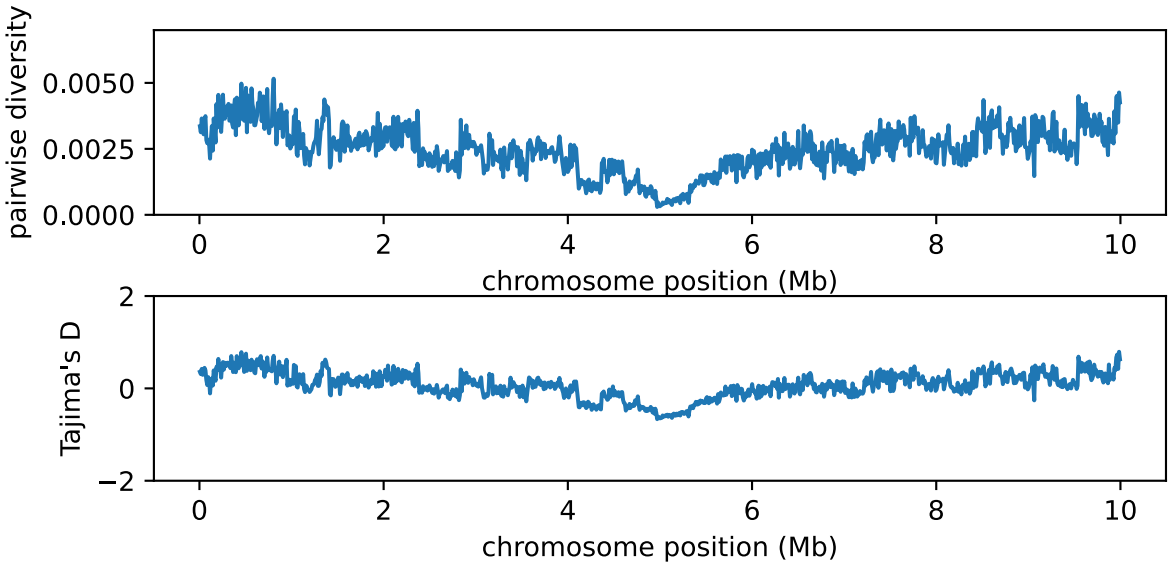

Figure S6: Genetic diversity π and Tajima's D in response to a very strong selective sweep with $s = 100$ after 20 generations. Simulations assume a population size of 300,000 individuals, with $r = 5 \times 10^{-8}$.

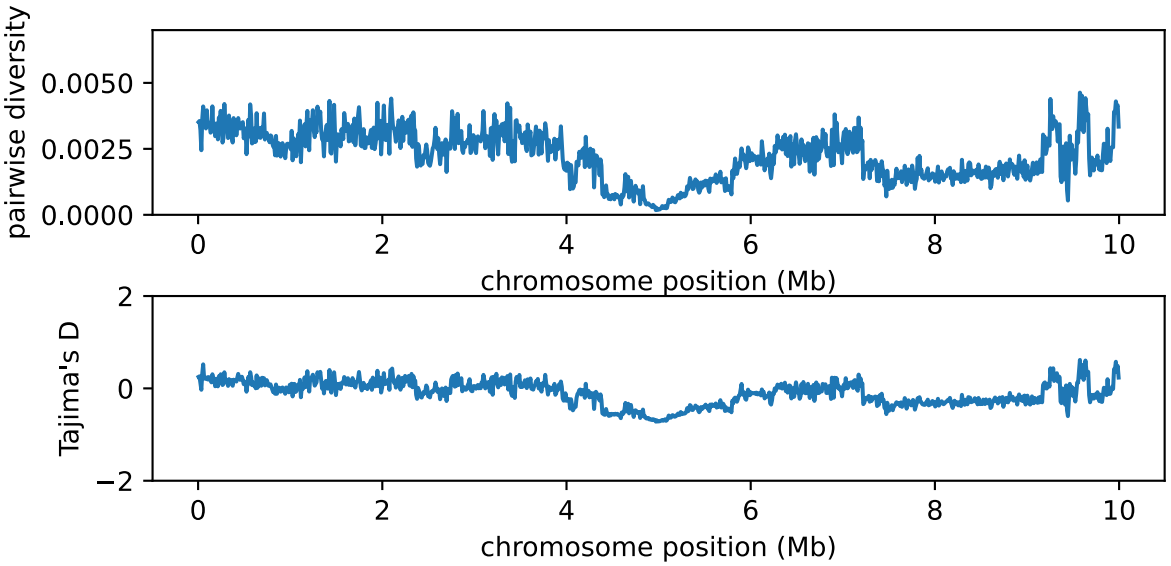

Figure S7: Genetic diversity π and Tajima's D in response to a very strong selective sweep with $s = 100$ after 50 generations. Simulations assume a population size of 300,000 individuals, with $r = 5 \times 10^{-8}$.

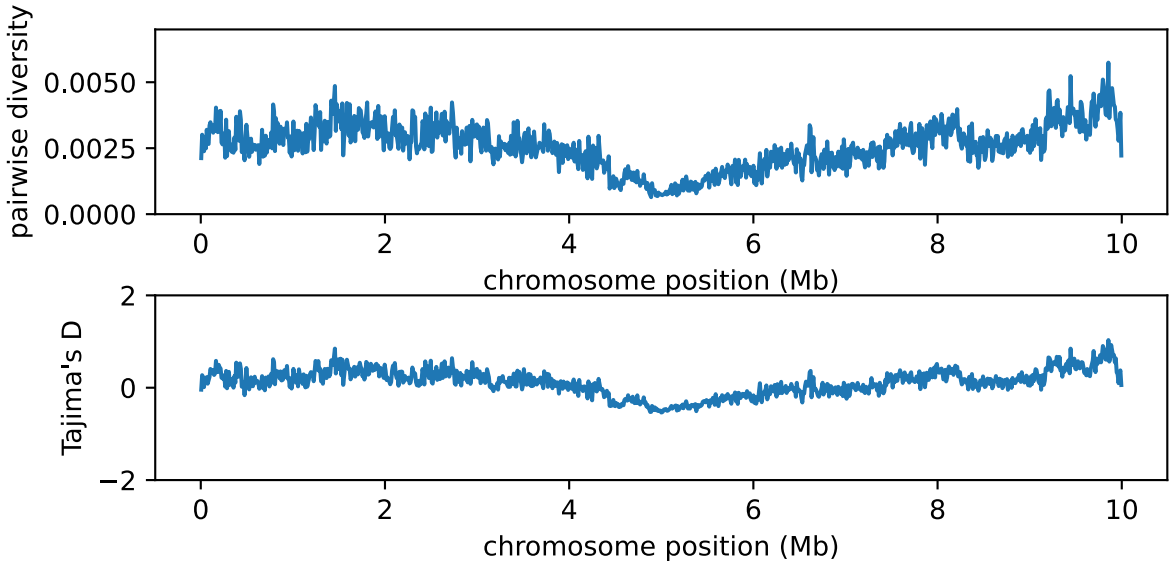

Figure S8: Genetic diversity π and Tajima's D in response to a very strong selective sweep with $s = 100$ after 15 generations. Simulations assume a population size of 300,000 individuals, with $r = 5 \times 10^{-8}$.

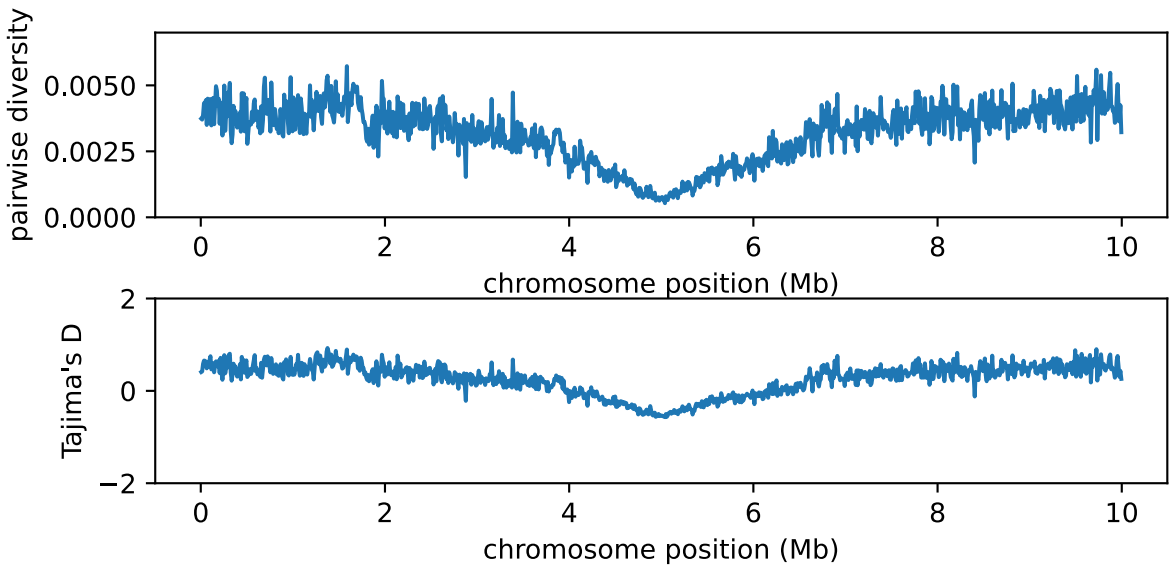

Figure S9: Genetic diversity π and Tajima's D in response to a strong selective sweep with $s = 10$ after 20 generations. Simulations assume a population size of 300,000 individuals, with $r = 5 \times 10^{-8}$. Diversity is reduced close to 0.0 in the 0.5 Mb region surrounding the selected locus. Diversity returns to near background levels towards the edges of the 10 Mb region.

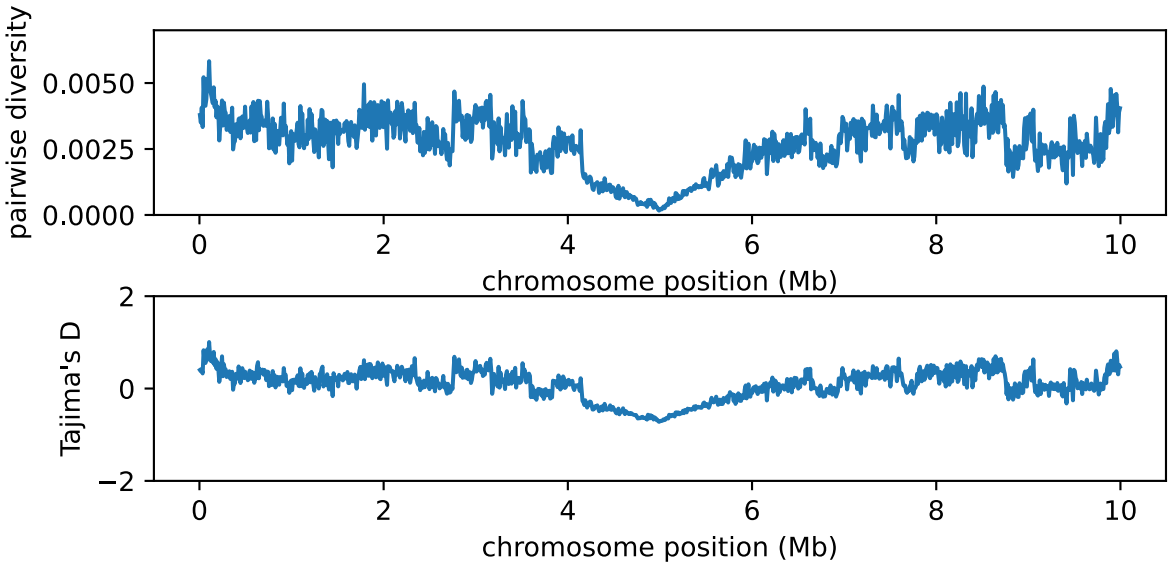

Figure S10: Genetic diversity π and Tajima's D in response to a strong selective sweep with $s = 10$ after 50 generations. Simulations assume a population size of 300,000 individuals, with $r = 5 \times 10^{-8}$. Diversity is reduced close to 0.0 for more than 1 Mb in the region surrounding the selected locus. Diversity returns to near background levels towards the edges of the 10 Mb region.

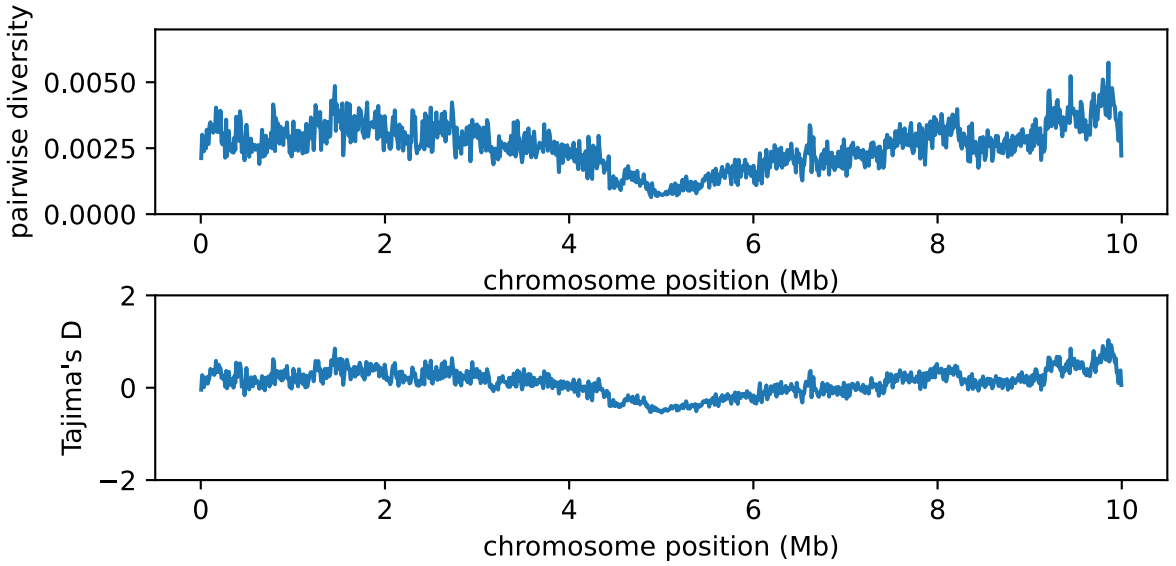

Figure S11: Genetic diversity π and Tajima's D in response to a strong selective sweep with $s = 10$ after 15 generations. Simulations assume a population size of 300,000 individuals, with $r = 5 \times 10^{-8}$.

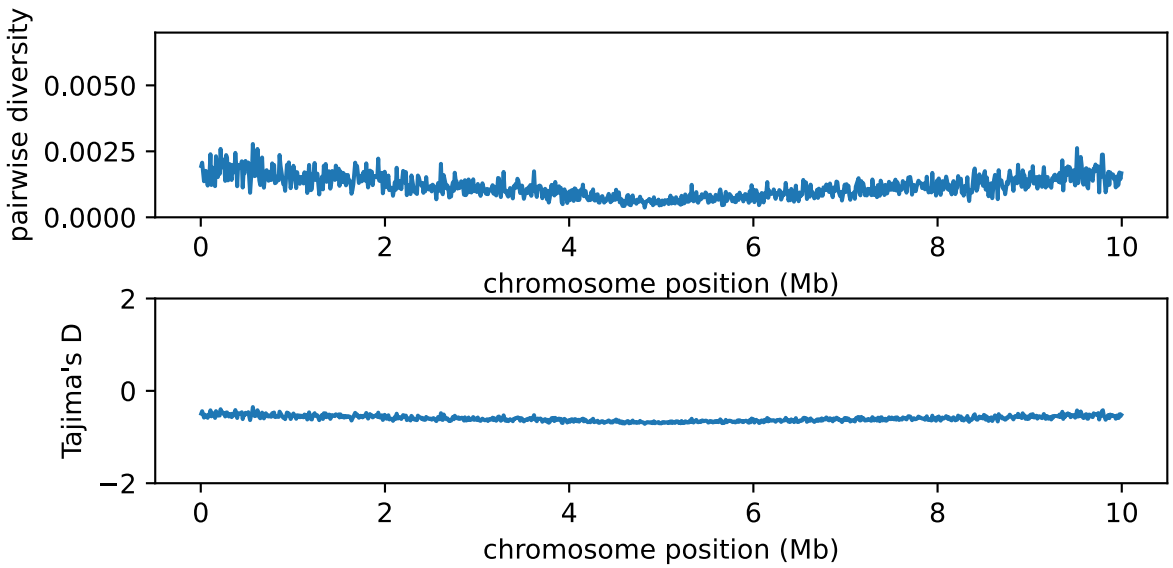

Figure S12: Genetic diversity π and Tajima's D in response to a very strong selective sweep with $s = 100$ after 20 generations in a genome with low recombination. Simulations assume a population size of 300,000 individuals, with $r = 5 \times 10^{-9}$. Under reduced recombination, background diversity is affected for the entire 10 Mb region and does not return to normal levels across the simulated contig, a pattern not observed in empirical data.

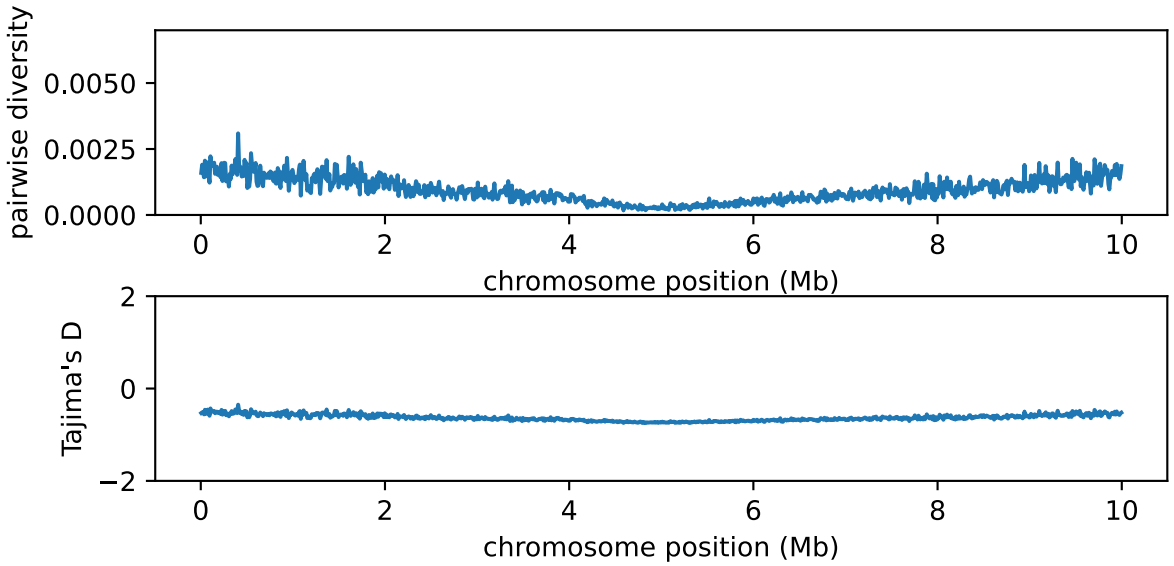

Figure S13: Genetic diversity π and Tajima's D in response to a very strong selective sweep with $s = 100$ after 50 generations. Simulations assume a population size of 300,000 individuals, with $r = 5{\times}10^{-9}$. Under reduced recombination, background diversity is affected for the entire 10 Mb region and does not return to normal levels across the simulated contig, a pattern not observed in empirical data.

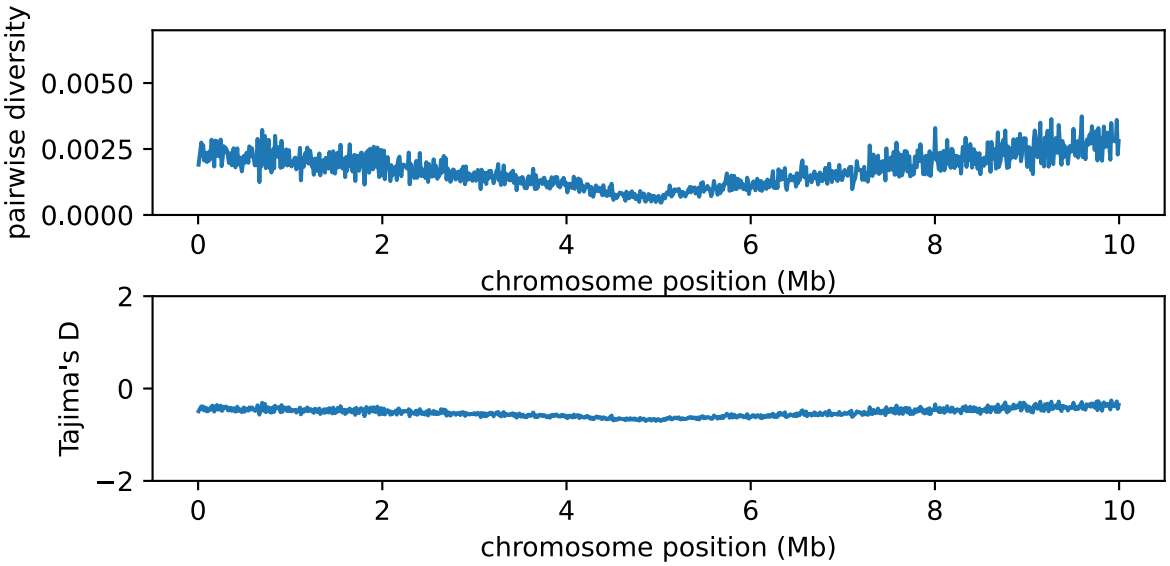

Figure S14: Genetic diversity π and Tajima's D in response to a very strong selective sweep with $s = 100$ after 20 generations under moderately reduced recombination. Simulations assume a population size of 300,000 individuals, with $r = 1 \times 10^{-8}$. Under reduced recombination, background diversity is affected for the entire 10 Mb region and does not return to normal levels across the simulated contig, a pattern not observed in empirical data.

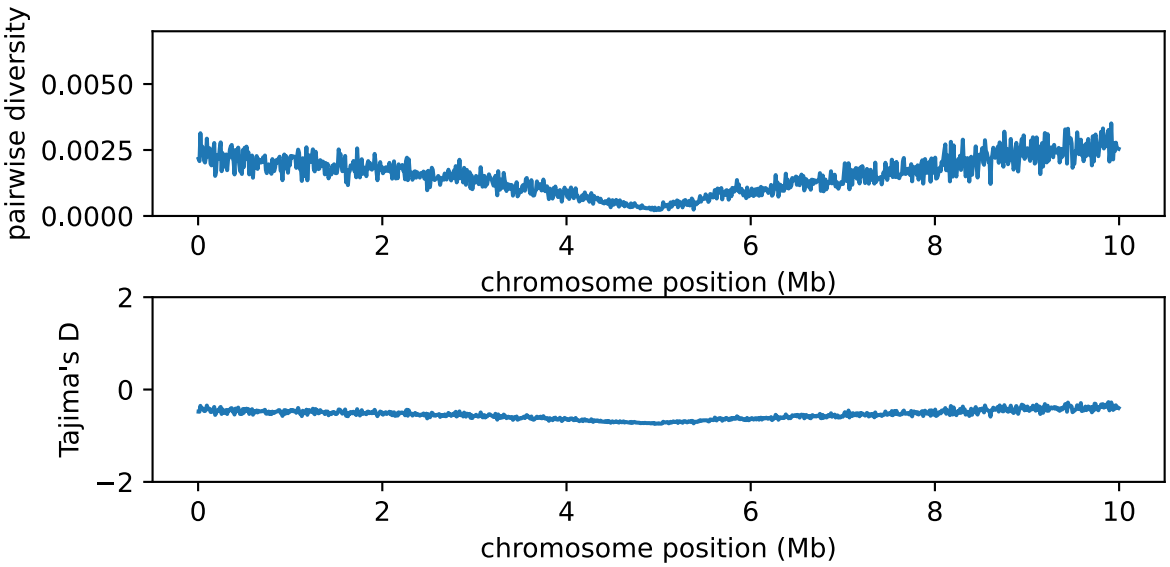

Figure S15: Genetic diversity π and Tajima's D in response to a very strong selective sweep with $s = 100$ after 50 generations under moderately reduced recombination. Simulations assume a population size of 300,000 individuals, with $r = 1 \times 10^{-8}$. Under reduced recombination, background diversity is affected for the entire 10 Mb region and does not return to normal levels across the simulated contig, a pattern not observed in empirical data.

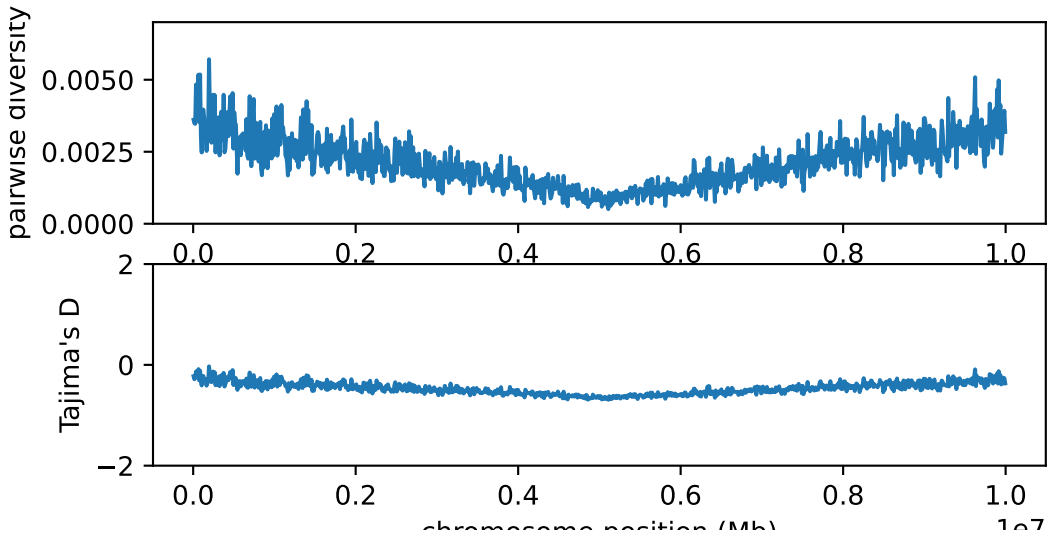

Figure S16: Genetic diversity π and Tajima's D in response to a very strong selective sweep with $s = 10$ after 20 generations under moderately reduced recombination. Simulations assume a population size of 300,000 individuals, with $r = 1 \times 10^{-8}$. Under reduced recombination, the locus of the sweep does not move as close to zero diversity and background diversity is affected for the entire 10 Mb region and does not return to normal levels across the simulated contig, a pattern not observed in empirical data.

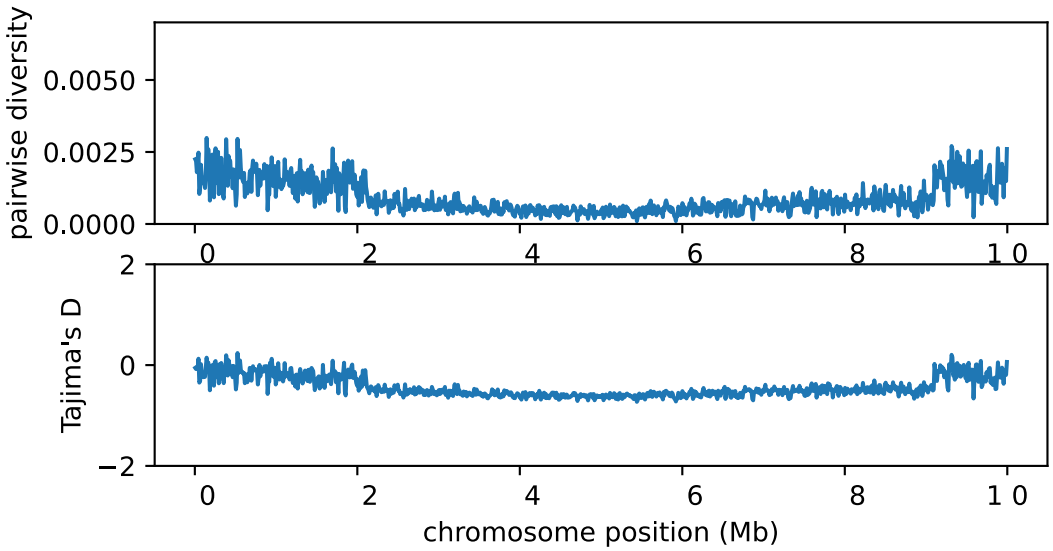

Figure S17: Genetic diversity π and Tajima's D in response to a very strong selective sweep with $s = 10$ after 20 generations under moderately reduced recombination. Simulations assume a population size of 300,000 individuals, with $r = 5 \times 10^{-9}$. Under reduced recombination, the locus of the sweep does not move as close to zero diversity and background diversity is affected for the entire 10 Mb region and does not return to normal levels across the simulated contig, a pattern not observed in empirical data.

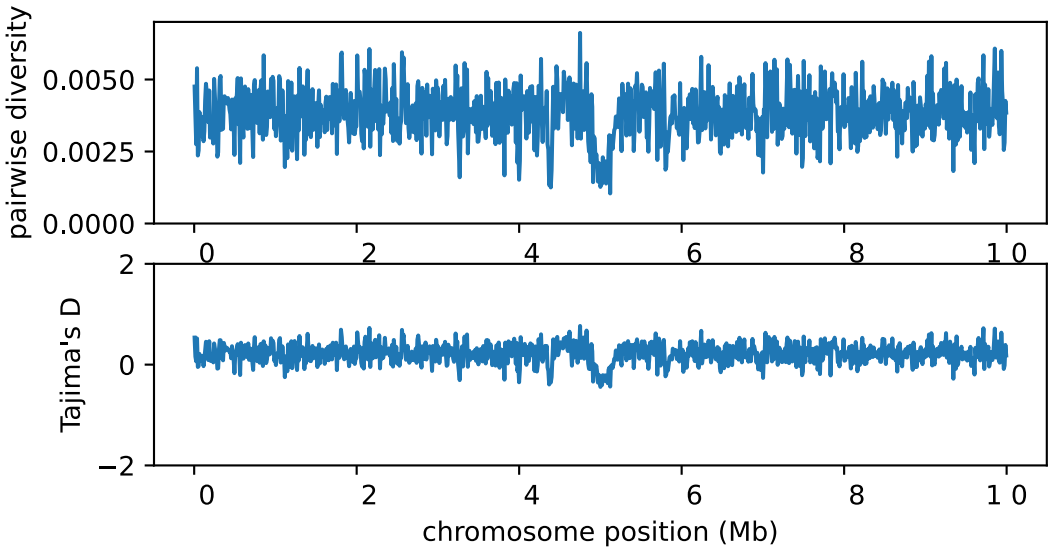

Figure S18: Genetic diversity π and Tajima's D in response to a very strong selective sweep with $s = 0.1$ after 300 generations under normal recombination. Simulations assume a population size of 300,000 individuals, with $r = 5 \times 10^{-8}$. These simulations show the classic v-like pattern for selective sweeps rather than extended blocks with near zero diversity.

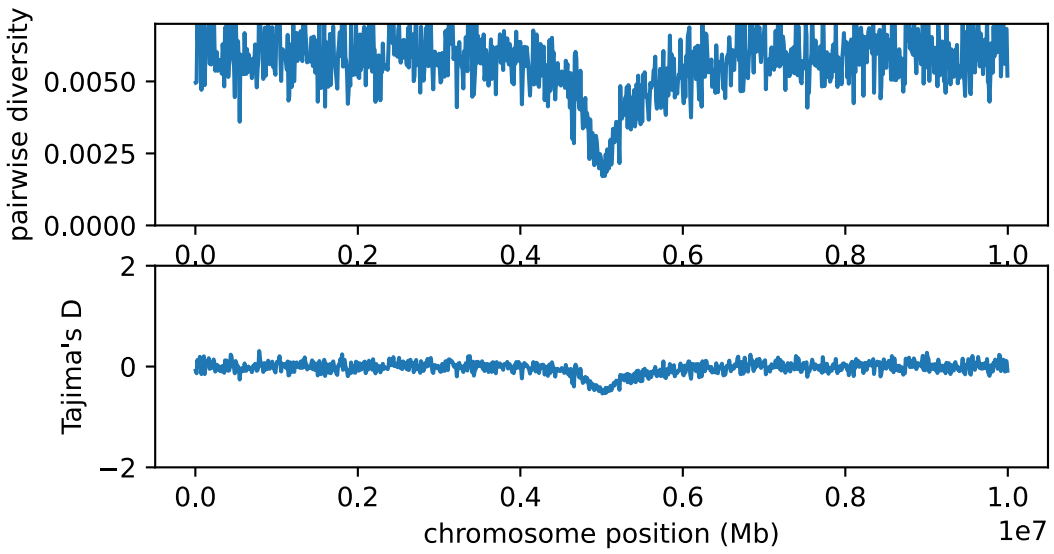

Figure S19: Genetic diversity π and Tajima's D in response to a weak selective sweep with $s = 0.1$ after 300 generations under moderately reduced recombination. Simulations assume a population size of 300,000 individuals, with $r = 1\times10^{-8}$. These simulations show the classic v-like pattern for selective sweeps rather than extended blocks with near zero diversity but with a wider scope than in normal recombination.

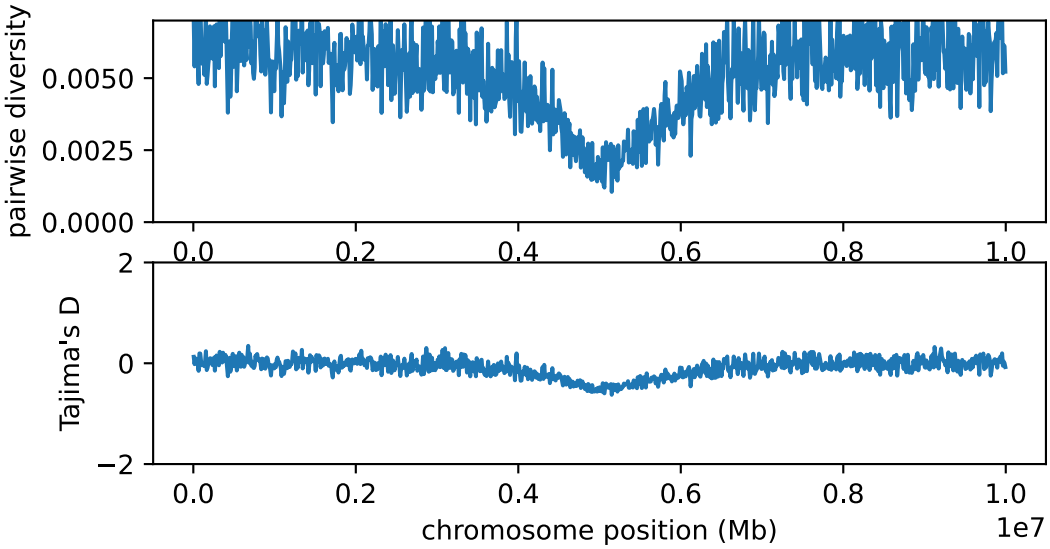

Figure S20: Genetic diversity π and Tajima's D in response to a very strong selective sweep with $s = 100$ after 50 generations under moderately reduced recombination. Simulations assume a population size of 300,000 individuals, with $r = 1 \times 10^{-8}$. These simulations show the classic v-like pattern for selective sweeps rather than extended blocks with near zero diversity, though the pattern is broader than in normal recombination.

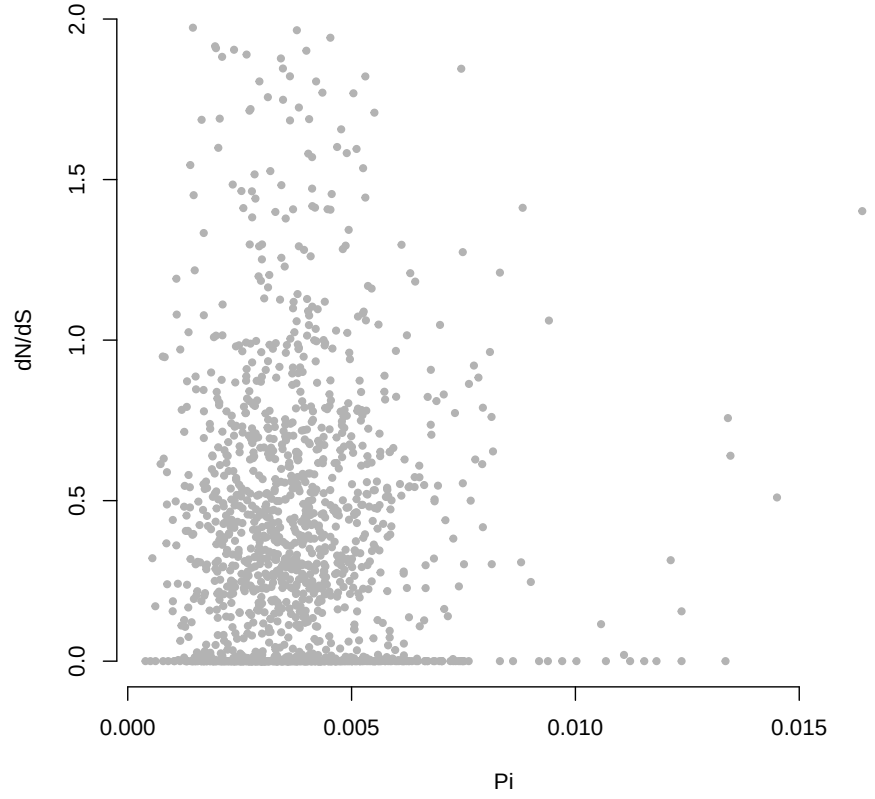

Figure S21: Genetic diversity π and dN/dS across paralogs are not correlated in *M. nervosa*, indicating that these two measures of selection can be applied independently in sequence data ($P = 0.89$, $R^2 = -0.00047$). Axes are truncated at dN/dS=2.0 for purposes of visualization.