Science
AAAS

Supplementary Materials for

A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes

Emily R. Hager *et al*.

Corresponding author: Hopi E. Hoekstra, hoekstra@oeb.harvard.edu

Science **377**, 399 (2022)
DOI: 10.1126/science.abg0718

The PDF file includes:

Materials and Methods
Figs. S1 to S21
Tables S1 to S6
References

Other Supplementary Material for this manuscript includes the following:

MDAR Reproducibility Checklist
Data S1 to S8

MATERIALS AND METHODS

Sampling and measuring forest and prairie mice

Field sampling

To test for phenotypic differences between forest and prairie ecotypes, we captured deer mice (*Peromyscus maniculatus*) from two populations separated by approximately 510 kilometers (km): 32 adult mice in sagebrush steppe in eastern Oregon ('prairie' ecotype) and 44 mice (39 adult) in temperate rainforest in western Oregon ('forest' ecotype); of these, we used 20 mice per population to establish laboratory colonies (see below). In addition, we sampled 136 mice (98 adult) from 22 sites focused in a 50 km transect across the Cascade mountain range, between the forest and prairie populations, with 1-19 mice captured at each site. We included both subadult and adult mice for genetic analysis but analyzed phenotypes from adults only. Trapping took place in September-October 2015 and August-September 2016 using Sherman live traps. With a few exceptions, the wild-caught mice were ultimately accessioned in Harvard's Museum of Comparative Zoology Mammal Department (see Data S1).

Establishing laboratory colonies

To establish laboratory colonies, we transported and quarantined 40 wild-caught mice (20 male, 20 female) at Harvard University. From these founders, we maintained two separate laboratory colonies, representing the forest ecotype, *P. m. rubidus* (4 original productive breeding pairs) and the prairie ecotype, *P. m. gambelii* (7 original productive breeding pairs). After quarantine, we housed mice at 23°C on a 16-hour:8-hour light:dark cycle in standard mouse cages (Allentown Inc, Allentown, NJ, USA) with corncob bedding (The Andersons, Inc, Maumee, OH, USA), cotton nestlet (Ancare, Bellmore, NY, USA), Enviro-Dri (Shepherd Specialty Papers, Watertown, TN, USA), and either a red tube or a red hut (BioServ, Flemington, NJ, USA). We provided animals with *ad libitum* water and mouse chow (LabDiet Prolab Isopro RMH 3000 5P75). We used the HAN rotation breeding scheme (*56*) to maintain outbred colonies.

Morphological measurements

For each individual, we took standard morphological measurements while the animal was alive (for colony-founding individuals) or immediately following euthanasia. We measured total length (nose to tail tip), hindfoot length, ear length, and tail length, and weight. We calculated body length as the difference between total and tail length. Laboratory-born mice were measured as adults, between 60 and 70 days old.

Pigmentation measurements

To measure coat color, we used a FLAME UV-VIS spectrometer with a pulsed xenon light source, a 400 um reflectance probe, and OceanView software (Ocean Optics) to measure 3-5 reflectance spectra from each of 3 body regions (dorsal stripe, flank, and ventrum). We used a custom R script (*55*) to obtain brightness, hue, and saturation values in the range 400-700 nanometer (nm) with 1 nm bin width, using the segment classification approach (*57*) with formulae as described for CLR v 1.05 (*58*). For every trait, we calculated the median value for each body region and individual. We measured forest and prairie mice shortly after euthanasia (for wild-caught mice) or while frozen (for laboratory-born specimen). For the transect analysis, all wild-caught mice were fixed in formalin, stored in ethanol, and then air-dried before measuring the dorsal and flank regions with the spectrophotometer.

Statistical analysis of phenotypes

To test for differences between the forest and prairie subspecies, we used t-tests for all 14 morphological and pigment traits (body length, weight, tail length, ear length, and hindfoot length; and brightness, hue and saturation in the dorsal, flank and ventral regions). We corrected all results for multiple testing using the Bonferroni-Holm method and performed statistical analysis using R v 3.6 (*59*).

Permissions and approval

We trapped mice under Oregon Department of Fish and Wildlife Scientific Taking Permits #107-15 and 127-16, with approval from Siuslaw, Deschutes, and Willamette National Forests and the Bureau of Land Management. We imported colony founders to Massachusetts on Division of Fisheries and Wildlife Importation Permit #043.15IMP. All experiments were approved by Harvard's IACUC (Protocol 11-05).

Forward genetic mapping

F2 intercross design

We conducted a reciprocal cross between laboratory-raised 'forest' *P. m. rubidus* and 'prairie' *P. m. gambelii* to generate first-generation (F1) hybrids, and then intercrossed sibling F1 hybrids to generate second-generation (F2) animals. Based on the measurements of laboratory-reared mice, we used R/qtl (*60*) to estimate that *n* = 500 hybrids would provide 80% power to detected loci explaining ~5% of the variance in tail length; we therefore tested 555 F2 hybrids in total (forest female x prairie male, *n* = 203 F2s from 12 F1 breeding pairs; prairie female x forest male, *n* = 352 F2s from 13 F1 breeding pairs). All 555 F2 hybrids descended from four individuals (the cross founders: one sibling male and female of each subspecies) and were measured for morphological traits as described above before genotyping.

Ancestry assignment in F2 hybrids

To genotype F2 hybrids, we generated low-coverage Double Digest Restriction Associated DNA (ddRAD) sequencing libraries (*61*) and then assigned hybrid genotypes with the Multiplexed Shotgun Genotyping (MSG) pipeline (*62*), using high-confidence fixed variants identified from whole genome sequencing of the cross founders.

ddRAD sequencing and joint genotyping of F2 hybrids

Briefly, we extracted DNA from liver tissue using the Autogenprep 965 (Autogen), digested the DNA with restriction enzymes MluCI and NlaIII (New England Biolabs), and then ligated adapter sequences (MluCI overhang: individually barcoded; NlaIII overhang: common, biotinylated). We pooled samples in groups of no more than 48 samples and used a Pippin Prep (Sage Science) to select fragments between 216 and 276 bp. We enriched for DNA with both adapters using streptavidin beads (Dynabeads, Invitrogen) and then performed 9 cycles of PCR amplification with Phusion Taq polymerase, adding pool-specific adapter sequences. After performing quality control with TapeStation (Agilent) and measuring concentration with Qubit 3.0 (ThermoFisher) and qPCR, we combined all pools into a single library and sequenced 125 basepair (bp) paired-end reads across four lanes of Illumina HiSeq v4. To avoid clustering problems during sequencing that can be caused by shared overhang sequence across ddRAD reads, we included a previously-generated RNAseq library as a diversity spike-in (10% of each lane). We performed ddRAD sequencing for all 555 F2 hybrids as well as the 4 cross founders and 49 F1s. Sequencing was performed at The Bauer Core Facility at Harvard University and computations on the Harvard Research Computing cluster.

We demultiplexed reads based on individual barcodes and then mapped sequencing reads to the *P. maniculatus bairdii* reference genome (NCBI accession: GCA_003704035.3) using BWA-MEM, with –p to indicate interleaved paired-end fastq input, and –M to mark short split hits as secondary for compatibility with Picard. The median read depth was 635,012 mapped reads per individual and we excluded the F2 hybrids ($n = 8$) with read depth less than 75,000 total mapped reads. From the mapped bam files, we ran *HaplotypeCaller* (GATK3.8) with the default heterozygosity prior (-hets = 0.001) and –ERC GVCF to produce per sample gVCFs. Then, we ran *GenotypeGVCFs* (GATK3.8) to jointly genotype the samples. We performed hard filtering of single nucleotide polymorphisms (SNPs) based on GATK best practices (filtering variants with QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0) using *VariantFiltration* (GATK3.8).

Identifying fixed SNPs between the forest and prairie cross founders

To identify fixed variants between the forest and prairie cross founders, we performed whole-genome re-sequencing of the 4 laboratory-born cross founders (2 forest *P. m. rubidus* and 2 prairie *P. m. gambelii*), as well as 3 of the 4 wild-caught parents of these founders (both parents of the forest founders and the female parent of the prairie founders). We extracted DNA from ~20mg of liver tissue and generated sequencing libraries using a PCR-free KAPA HTP kit. Following enzymatic fragmentation, we used size selection to enrich for a 450 bp insert size and ligated Illumina adapters. We sequenced the resulting libraries using 150 bp paired-end sequencing on an Illumina NovaSeq S4 flowcell.

Following demultiplexing, we mapped sequencing reads as described above and marked optical and sequencing duplicates using *MarkDuplicates* (Picard), with OPTICAL_DUPLICATE_PIXEL_DISTANCE=2500 to account for artifacts generated from the patterned flowcell found in the NovaSeq S4. To call variant sites, we first used *HaplotypeCaller* (GATK4.1) on each sample as above. Then, intermediate haplotype files for all individuals were consolidated into a GenomicsDB structure using *GenomicsDBImport* (GATK4.1), which was used to create variant + invariant cohort-level vcfs for each chromosome with *GenotypeGVCFs* (GATK4.1). We performed hard filtering of SNPs with parameters as described above, and hard filtering of INDELs with QD < 2.0, FS > 200.0, ReadPosRankSum < -20.0, SOR > 3.0. We also filtered invariant sites with QUAL ≥ 20 using bcftools. Finally, we masked all variant calls with read depth < 5.

To create a set of fixed variants between the prairie and forest founders, we combined the ddRAD and whole genome re-sequencing data using *CombineVariants* (GATK3.8). We then performed SNP filtering independently for each set of reciprocal cross founders. We used *SelectVariants* (GATK3.8) to select SNPs that showed fixed differences between the forest and prairie founders (number of fixed SNPs = 1,022,051 for prairie female x forest male; 984,603 for forest female x prairie male). Since we expected the F1s to be heterozygous at each fixed SNP, we excluded any SNP for which the ratio of the number of reference to alternate calls across all F1 reads fell within the top or bottom 10th percentile of a binomial distribution (p = 0.5, n=total number of F1 reads at a given SNP). Finally, we removed SNPs for which a cross founder was called homozygous for one allele, but the founder's parent was homozygous for the opposite allele. Post-filtration, we obtained a total of 793,928 fixed SNPs for prairie female x forest male and 784,259 fixed SNPs for forest female x prairie male.

Ancestry assignment of F2 hybrids with the multiplexed shotgun genotyping pipeline

We used the hidden Markov model implemented in the Multiplexed Shotgun Genotyping (MSG) pipeline (*62*) to assign ancestry in F2 hybrids. In brief, we used samtools *mpileup* to extract the appropriate set of fixed SNPs from the mapped bam files of F2 hybrids, requiring that SNPs were at least 500 bp apart to ensure they were from independent reads. Then, we ran the fit-HMM step of MSG (fit-hmm.R) on the filtered mpileups with the settings: deltapar1=0.1, deltapar2=0.1, rfac=1; priors=0.25,0.5,0.25; theta=1; one_site_per_contig=1; recRate=25. We combined the fit-HMM genotype probabilities across all F2 hybrids using combine.py (https://github.com/JaneliaSciComp/msg/), which interpolates missing genotypes, resulting in a total of 978,808 markers. Next, we thinned the marker density, retaining every 100th marker and keeping neighboring markers for which at least one F2 hybrid had genotype conditional probabilities differing by 0.1 using pull_thin (https://github.com/dstern/pull_thin) with difffac=0.1. After thinning, we had a total of 109,356 markers. Finally, we created an R/qtl object using read.cross.msg.1.5.R (https://github.com/dstern/read_cross_msg/), where we directly imported genotype probabilities from fit-HMM as the R/qtl genotype probabilities and added the phenotype data. We dropped markers with missing genotypes for more than 25 individuals, resulting in 108,575 markers used for mapping.

Quantitative Trait Locus mapping

To identify genomic regions associated with morphological and pigment variation after testing for correlations among traits (Figure S17), we performed quantitative trait locus (QTL) mapping using the extended Haley-Knott method in R/qtl v. 1.45.11; ('ehk' method in the *scanone* function; (*60*, *63*)). We used permutation tests to obtain thresholds for genome-wide significance, with 1000 permutations for the autosomes and a separate significance threshold for the X chromosome (18,600 permutations), and estimated confidence intervals using the 95% Bayes credible interval (*bayesint* function in R/qtl). To estimate the additive and dominance effect sizes and the percent of the total phenotypic variance associated with each locus, we fit a linear model including the peak markers for all significant loci and any covariates using the *fitqtl* function in R/qtl (method = 'hk', including additive effects of each locus). Because tail and hindfoot length were significantly correlated with body length (tail-body: r = 0.26, 95% confidence interval (CI) = 0.18-0.33; hindfoot-body: r = 0.18, 95% CI = 0.09-0.26), we included body length as an additive covariate for these traits. For all traits, we also tested whether there was a statistically significant association with 3 additional possible covariates: age, cross direction, and sex, using Pearson's correlations for the continuous variable (age) and two-sided Welch's t tests for the categorical variables (cross direction and sex). Where such associations existed (with age, cross direction, and sex for foot length, age and sex for tail length, and cross direction only for pigmentation traits), we ran additional QTL models including these as additive covariates. As including these extra covariates did not substantially alter the results (Table S2), we report the results from the simpler models here.

Recombination breakpoint analysis

We determined locations of recombination breakpoints in the F2 hybrids from their genotype probabilities in the R/qtl object. We localized breakpoints to the midpoint of where an individual's homozygous genotype probability transitioned from below 0.05 to above 0.95 or from above 0.95 to below 0.05. For consecutive breakpoints involving transitions between the same two genotypes, if the breakpoints localized within 500 kb of each other, we excluded the breakpoints from our analysis as such closely neighboring breakpoints were likely due to short tracts of spurious ancestry calls.

Identifying an inversion on chromosome 15

In our QTL analyses, we identified a region of chromosome 15 that showed limited recombination over a large physical distance (~40 Mb). To investigate this region further, and in particular, to determine whether the lack of recombination was due to a chromosomal rearrangement, we first used whole-genome re-sequencing data in a larger sample of wild-caught mice. Using these data, we tested whether this region was more differentiated between the forest and prairie populations compared to the rest of the genome, consistent with a structural variant inhibiting gene flow in the wild. We then used long-read sequencing to determine the nature of the rearrangement.

Estimating F_{ST} between forest and prairie populations

To assess genome-wide levels of nucleotide variation, we performed whole genome re-sequencing for 30 wild-caught mice (forest $n = 15$; prairie $n = 15$). We generated sequencing libraries, called variants, and performed filtering together with the cross founders (see above: "Identifying fixed SNPs between the forest and prairie cross founders"), except that for 4 mice (2 forest, 2 prairie) we generated sequencing libraries using the PCR-based KAPA HTP kit.
We estimated F_{ST} using the program *angsd*, which can take genotype uncertainty into account instead of relying on called genotypes (*64*). First, input bam files were used to generate site allele frequency likelihood files (SAFs) with the following command: "angsd -gl 2 -doSaf 1 -minMapQ 30 -minQ 20 -C 50 -baq 1". We next estimated individual and pairwise site frequency spectra (SFS) using *realSFS fst index* and setting the parameter "-nSites" to 5e8. Finally, we calculated F_{ST} using *realSFS* with the individual and pairwise SFS as priors. Global F_{ST} and sliding window F_{ST} were estimated using "*realSFS fst stats*" and "*realSFS fst stats2 -win 10000 -step 1000*" respectively.

Linkage disequilibrium in forest and prairie populations

We next investigated linkage disequilibrium across chromosome 15 in the forest and prairie mice ($n = 30$). Using the wild-caught re-sequencing data, we filtered chromosome 15 SNPs to include only biallelic SNPs with < 5% of samples missing genotypes and minor allele frequency > 0.1. We then thinned SNPs to ≤ 1 SNP per 100 kb, resulting in a total of 786 SNPs. We used vcftools --geno-r2 to compute r^2 between each pair of SNPs, using genotypes to calculate correlations to accommodate unphased data.

Long-read sequencing

To better characterize the putative chromosome 15 rearrangement, we performed PacBio long-read sequencing (*65*) on 2 individuals: one forest mouse homozygous for the structural variant and one prairie mouse homozygous for the reference allele. First, we extracted high-molecular weight (HMW) DNA from 200 uL fresh blood using the MagAttract HMW DNA mini kit (Qiagen), following the Whole Blood protocol (Qiagen), using wide-bore pipette tips to prevent shearing at the elution step. We quantified the resulting DNA using a Genomic DNA ScreenTape on the Tapestation 4200 (Agilent). Library preparations and sequencing were performed at the University of Washington's PacBio Sequencing Core. In brief, libraries were prepared with the SMRTbell Express Template Prep Kit 2.0 (PacBio). We performed a size selection of 30 kb for the forest sample using the BluePippin (Sage Science), and we did not perform any size selection for the prairie sample since total library mass was below 500 nanograms. Then, we sequenced each on a Sequel II SMRTcell 8M (PacBio), the forest sample with a 15-hour movie and the prairie sample with a 30-hour movie. The unique molecular yield was 131.3 gigabases (Gb) for the forest sample and 134.7 Gb for the prairie sample, with the longest subread N50 of 37,943 bp and 36,619 bp, respectively.

For each sample, we generated *de-novo* assemblies using the program *canu* (*66*). Given the high levels of heterozygosity, we specified the parameters corOutCoverage=200 and correctedErrorRate=0.15 to allow some read mismatch and therefore combine haplotypes, achieving a haploid (rather than diploid) assembly. Contig N50s for the de-novo forest and prairie assemblies were 1.37 Mb and 1.22 Mb, respectively. We then used the program *mummer* to align each de-novo assembly to the *P. maniculatus bairdii* reference genome (*67*). We implemented the *nucmer* command with default parameters to accommodate potential rearrangements between the draft assembly and reference genome.

We noted that the resulting alignments indicate that many genomic regions surrounding the identified breakpoint map to multiple regions throughout the reference chromosome 15 assembly (Figure S5), likely indicating the presence of repetitive, multi-mapping sequence. This feature likely prevented the identification of the breakpoint from short-read sequencing alone.

After determining that chromosome 15 harbors a large chromosomal inversion, we used the program *sniffles* (*68*) to search for additional structural variants on the chromosome in an unbiased manner by calling variants from long-read mapping. From each movie, we converted the subreads bam to fastq using bam2fastx (PacBio). We then aligned fastq files to the *P. maniculatus bairdii* reference genome using the program ngmlr with the pacbio preset parameter "-x pacbio". Next, we converted output sam files to bam format with *samtools view* and added readgroups with *AddOrReplaceReadGroups* (Picard) for downstream compatibility. For each individual, we called variants using the program *sniffles* with the parameter "-d 5000". We then merged these raw variant calls with the program *SURVIVOR* using the parameters "1000 1 1 -1 -1 -1". This merged callset was then used to re-genotype each individual with *sniffles*, so as to obtain a genotype for each individual at every site. We merged the final callset again using *SURVIVOR* and considered any large structural variants (>100 kb) that were fixed differences between the sequenced forest and prairie mice as a candidate set, which we verified using the contig alignment.

Predicting centromere location on chr15

We mapped a 344-bp satellite sequence (NCBI accession: KX555290.1) (known to localize to *P. maniculatus* centromeres (*69*)) to the *P. maniculatus* reference genome and the forest and prairie PacBio genome assemblies using blastn (blast v2.6.0), filtering for alignments >85% identity. We then determined the chr15 centromere location in the reference genome (converting alignment positions in the PacBio genomes to their corresponding or closest reference genome coordinates).

Determining the frequency of the inverted haplotype

To genotype the 30 wild-caught individuals for the inversion, we first examined patterns of relatedness and heterozygosity across chromosome 15. Specifically, we calculated heterozygosity for each individual, within the 'affected' (inversion, 0-41 Mb) and 'unaffected' (no inversion, 50-79 Mb) chromosome 15 regions using "*vcftools –het*". Next, we used *plink –pca* to perform PCA on all biallelic SNPs from both regions. Three distinct clusters in PC1 suggested 3 genotypes: homozygous for the inversion, heterozygous, and homozygous for the reference allele. These genotypes were consistent with observations of decreased heterozygosity in homozygous inversion mice and increased heterozygosity in heterozygous mice (Figure S18).

Gene content analysis within the inversion

To identify genes within the inversion with known variants affecting either tail length or pigmentation in mammals, we first used the *P. m. bairdii* genome annotation (Pman2.1_chr_NCBI.corrected.merged-with-Apollo.Aug19.sorted.gff3) to identify genes located within the chromosome 15 inversion (*n* = 296 annotated genes; *n* = 149 protein-coding genes). We next used the Batch Query function of the Mouse Genome Informatics database (MGI; http://www.informatics.jax.org/) to identify phenotypes reported for mice with mutations in any gene within the inversion. We filtered these for tail length and pigment related phenotypes (terms: 'abnormal coat/hair pigmentation', 'diluted coat color', 'long tail', 'short tail', 'elongated vertebral body', 'increased caudal vertebrae number', 'decreased caudal vertebrae number', 'kinked tail', 'abnormal tail morphology', 'abnormal vertebrae morphology', 'abnormal caudal vertebrae morphology', 'increased length of long bones', and 'decreased length of long bones'). We determined which genes had non-synonymous mutations fixed between the inversion and reference haplotypes using the *PopGenome* package in R (*70*). We imported SNPs from the whole-genome re-sequencing vcf of the wild-caught forest (*n* = 13) and prairie (*n* = 15) mice into *PopGenome*, excluding the 2 forest mice carrying at least one reference allele. We then used the functions *set.synnonsyn* and *MKT* with the annotation file Pman2.1_chr_NCBI.corrected.merged-with-Apollo.Aug19.sorted.gff3 to identify non-synonymous mutations fixed between the inversion and reference allele samples. To explore the functional effects of these non-synonymous mutations, we used the program PROVEAN (*71*), which predicts whether an amino acid substitution is deleterious based on protein sequence alignments.

Distribution of the chromosome 15 inversion and morphological traits across a transect

Cascades transect sampling

To test for a relationship among phenotypic variation, environmental variation, and allele frequency change in the wild, we sampled additional mice (*n* = 136) between the forest and prairie populations. Specifically, we focused sampling in a 50-km region running east-west across the Cascades mountain range for several reasons: first, the Cascades represent a sharp habitat transition from wet coastal forest to dry interior forest, and also form a contact zone for many other species and subspecies (*72*), and second, our initial (2015) samples indicated a sharp phenotypic change at this location, which recapitulated most of the difference between the forest and prairie mice. In addition to sampling across the Cascades, we also included museum specimens (*n* = 12) from a site intermediate between the Cascades samples and the eastern-most prairie site. For these museum specimens, we were able to obtain comparable data for tail length and inversion genotype (see below), although not for pigmentation or whole-genome ancestry.

Habitat characterization

To determine whether phenotypic and/or allele frequency change was associated with environmental variation, we used QGIS v. 3.4 (*73*) to evaluate vegetation type and soil color across the full transect. To assess soil color, we used the publicly available STATSGO2 data from USGS (*74*). To estimate local soil characteristics at each site, we found the total area within a defined radius of the trapping location (i.e. 0.5, 1, or 2 km) that belonged to each official USDA soil series and generated a weighted average of the Munsell soil color, value and chroma. We used a custom python script (with results verified by manual review) to gather Munsell color characterizations for the top-most layer of each soil series from the USDA Official Soil Series Descriptions (*75*). If the top-most layer did not have Munsell color information, we used the values from the top layer that did. We used the Munsell characteristics for moist soil if the description listed the series as ‘usually moist’, udic or aquic moisture regimes, or dry less than 90 consecutive days in summer; otherwise, we used the dry soil characteristics (i.e. if the series was described as ‘usually dry’, aridic moisture regime, or dry more than 80 days in the summer). For analysis, we converted Munsell hues to degrees with hue 5R at 0°; thus, 5YR = 36°, 7.5YR = 45°, 10YR = 54°, and 2.5Y = 63°. Finer-scale soil survey (i.e. SSURGO) data were not available for the central Cascades transect, but where such surveys were available, the estimated soil characteristics we obtained from the two datasets matched. Because the results were similar across all three radii (Figure S7), we chose to focus on the 1-km data. For vegetation data, we made use of detailed habitat models from the Oregon Biodiversity Information Center (Oregon Spatial Data Library, https://spatialdata.oregonexplorer.info/) (*76*) to calculate the proportion of area within the same three radii around the sampled sites occupied by each habitat type (Table S5). These results also did not depend heavily on the chosen radius (Figure S6). To display maps, we used ggmap (*77*).

Genotyping the inversion and estimating ancestry in Cascades mice

Whole genome re-sequencing and joint genotyping

We performed low-coverage, whole-genome re-sequencing of the mice (*n* = 136) from the 22 sites in the central Cascades transect. We extracted genomic DNA from liver tissue using proteinase K digestion followed by the Maxwell RSC (Promega) DNA extraction. For library preparation, we used the Nextera XT kit with Illumina adapters, performing the reactions at 1/4 volume. We quantified libraries with the TapeStation (Agilent), pooled the samples into a single library, and quantified the pooled library by qPCR. We then sequenced the library on a full flow cell of Illumina NovaSeq SP with paired-end sequencing of 150 bp reads.

After de-multiplexing reads based on the Illumina barcodes, we mapped the reads from the fastq files to the *P. maniculatus bairdii* reference genome following the protocol described above. The median read depth for these samples was 15,347,998 mapped reads per sample, which corresponds to ~1.5X sequencing coverage across the genome. From the mapped bam files, we created cohort level vcfs as described above.

Determining inversion genotypes

To determine chromosome 15 genotypes, we created a set of fixed SNPs to differentiate the inversion and reference haplotypes. We used *SelectVariants* (GATK3.8) to select fixed SNPs in the affected region between the wild-caught mice homozygous for the inversion (*n* = 13 forest) and those homozygous for the reference allele (*n* = 15 prairie, *n* = 1 forest). We filtered this set of SNPs by requiring that the one forest mouse heterozygous for the inversion must be heterozygous at the selected SNPs, resulting in a set of 37,242 SNPs fixed between the inversion and reference haplotypes. We then performed the fit-HMM step of MSG for the Cascades mice using this set of SNPs for the affected region, as above.

For the additional museum specimens (*n* = 12), we extracted genomic DNA from liver tissue using proteinase K digestion followed by the Maxwell RSC. Then, we used four custom Taqman SNP genotyping assays (Life Technologies) to genotype diagnostic SNPs between the inverted and reference haplotypes (Table S6). All genotyping reactions were performed with 1-10 ng of genomic DNA, using the following cycling parameters: 95 °C for 10 minutes followed by 40 cycles of 95 °C for 15 s, 60 °C for 1 minute. For all individuals, the four assays gave the same genotype results.

Ancestry estimates

As the forest and prairie populations had low genetic differentiation across the genome, we used ngsAdmix (*78*) to estimate ancestry proportions for each Cascades transect individual (*n* = 136), and for the previously sequenced wild-caught forest and prairie individuals (*n* = 30). To identify a set of SNPs for ngsAdmix, we selected biallelic SNPs from all autosomes that had at least 50% of all samples (forest, prairie, and Cascades mice) with non-missing genotypes using *SelectVariants* (GATK3.8) and thinned SNPs to be at least 1 kb apart to avoid SNPs in strong linkage disequilibrium (Figure S19). We excluded the affected region so that whole-genome ancestry estimates were not influenced by the inversion genotypes, resulting in a total of 472,692 SNPs. To run ngsAdmix, we created a beagle file using a custom R script to convert GATK PLs to genotype likelihoods. Then, we ran ngsAdmix with minMAF=0.1 and *k*=1-6. We performed 20 ngsAdmix runs per k, each with a different random seed. Using CLUMPAK (*79*, *80*), we determined that *k*=2 was the best number of clusters (Figure S10A). Thus, we ran ngsAdmix with minMAF=0.1 and *k*=2 for estimating ancestry coefficients. We determined confidence bounds for the ancestry estimates by performing 100 iterations of subsampling the input SNP set to 25% and re-running ngsAdmix (Figure S10B).

In addition, we used ngsAdmix to confirm MSG genotypes for the inversion. From the set of thinned SNPs as described above, we had 7,090 SNPs in the affected region of chromosome 15. Using this set of SNPs, we ran ngsAdmix as described above, and the ngsAdmix ancestry estimates confirmed MSG genotypes for the inversion (Figure S10F). Finally, we determined the allele frequency difference between the forest and prairie focal populations for all SNPs used in ngsAdmix.

Fitting clines to genotypes and phenotypes

First, to test whether mouse pigmentation was associated with soil color and value, we calculated the correlation between mouse hue and mean soil hue at the site of capture, between mouse brightness and mean soil value, and between mouse saturation and mean soil chroma (Figure S20).

To determine spatial changes in morphological and genetic variation, we next used R to fit sigmoid clines for each trait separately using both the full dataset (i.e. wild-caught forest, prairie, and Cascades mice, with additional museum specimens for tail length and the inversion) and, separately, data restricted to the Cascades. We first converted the spatial coordinates of each site into an east-west transect distance, with the central point (distance = 0 km) at the highest elevation. We then used the *distRhumb* function from the package geosphere v. 1.5.7 (*81*) to calculate east-west distances between this central point and each other sampled site.

We next used the package HZAR v. 0.2.5 (*82*) to fit clines for tail length, dorsal and flank hue, chromosome 15 inversion genotype, and whole-genome ancestry. For phenotype values, we fit 5 cline models, which varied only according to whether and how exponential tails were fit (tails = 'none', 'left', 'right', 'mirror', and 'both'); exponential tails indicate stepped clines (*83*). For genotype data, we fit 10 cline models, which varied by how exponential tails were fit (tails = 'none', 'left', 'right', 'mirror', and 'both') and by the scaling of minimum and maximum allele frequencies (scaling = 'fixed' for minimum and maximum fixed as minimum and maximum of observed mean data, 'free' for minimum and maximum as free parameters). We selected the best model for each trait and genotype using AICc values. We bounded the values for center and width such that the center could not be more than 10 km outside the sampled region, and the width no more than the total transect distance + 20 km (*82*). For phenotypic data, we also bounded the variance in the center of the cline ('varH') to be no more than 1.5 times the total variance in the dataset, which helped runs to converge. For each model, we ran 3 independent chains (varying the random seed) with 3 runs each, with chain length 1e6, thin 1e3, and burnin 1e5 for phenotype data, and chain length 1e5, thin 100, and burnin 1e4 for genotype data (*55*). We assessed chain mixing by visually inspecting the trace and using the potential scale reduction factor (*gelman.diag* in the coda package, (*84*)).

Finally, to test for coincidence of the genome-wide ancestry and chromosome 15 inversion clines, we used the likelihood profile method (*85*). For each variable to compare (i.e. whole genome ancestry proportion and inversion genotype), we constructed likelihood profiles by finding the likelihood of the best-fit cline model with the center fixed at each of 21 values between -30 and 30 km, the region where all best-fit cline centers were located. We fit exponential tails according to the best-fit model from the initial analysis (i.e. tails = 'right' for ancestry, and 'left' for chromosome 15; scale = 'fixed') and used the full, not transect-restricted, dataset. We used a chi-square test with test statistic equal to the sum of the maximum likelihood values for each trait (MLsum) minus the maximum value of the summed likelihood profiles

(MLcomp) to test for cline coincidence of particular clines, with degrees of freedom equal to the number of compared traits minus one. If all the clines are co-located, MLsum and MLcomp should be approximately equal, while if not, MLcomp will be significantly less than MLsum.

Linking genotype and phenotype in Cascades mice

To test whether genotype at the chromosome 15 inversion was associated with phenotype in the mice from the 50-km transect across the Cascade Mountains, we used linear mixed effects models (Figure S21). Specifically, we used tail length, dorsal hue, and flank hue as response variables, with fixed effects of genotype at the chromosome 15 inversion and the estimated proportion of genome-wide forest ancestry excluding chromosome 15 from ngsAdmix, and a random effect of site of capture. We compared these 'full' models to models that included ancestry proportion and capture site, and to models that included site alone, using AIC values to select the best model. In the Cascades, sample sizes for the three chromosome 15 genotypes were highly variable: in particular, while there were 59 forest homozygotes and 34 heterozygotes at the inversion, we caught only 5 mice in this region that were homozygous for the prairie allele. To aid interpretability, we therefore coded genotype as a factor with the homozygous forest genotype as the baseline and separate effects for the heterozygote and homozygous prairie genotypes rather than separately coding the additive and dominance effects of the locus.

Absolute allele frequency differences across the genome

To determine whether the inversion frequency change across the transect was consistent with divergent selection, we compared it with blocks of similar linkage disequilibrium (LD) across the genome. We binned the genome into 200-bp non-overlapping windows, as mean r^2 is approximately 0.1 at this distance (Figure S19). Although the mean r^2 within the inversion is higher (~0.2), the rapid decay of linkage disequilibrium in the forest and prairie populations makes a window size of mean r^2=0.2 infeasible (Figure S19). For each 200-bp window, we considered only SNPs with minor allele frequency (MAF) > 0.1 and reported the SNP with highest absolute allele frequency difference between the wild-caught forest (n = 15) and prairie (n = 15) mice, with a few exceptions. In particular, we noticed three other high-LD blocks that will be subject to further investigation. We collapsed these three blocks of high-LD, as well as the inversion on chromosome 15, to single values representing the most frequent maximum absolute allele frequency difference of 200-bp windows within each block. We then compared the absolute allele frequency difference of the inversion (0.9) to the maximum absolute allele frequency difference of all similar linkage disequilibrium blocks genome-wide.

Chromosome 15 maximum-likelihood trees

To assess forest-prairie divergence at the inversion, we constructed maximum likelihood trees for the affected (0-41 Mb) and unaffected (41-79 Mb) regions of chromosome 15 for the wild-caught forest (n = 14) and prairie (n = 15) mice, excluding a single forest mouse heterozygous for the inversion. From the whole-genome re-sequencing vcf, we thinned SNPs to a maximum of 1 SNP per 100-bp, converted the vcf to a PHYLIP matrix using vcf2phylip.py (*86*) (https://github.com/edgardomortiz/vcf2phylip) and removed invariant sites using ascbias.py (https://github.com/btmartin721/raxml_ascbias). This resulted in 76,163 and 50,909 SNPs for the affected and unaffected regions, respectively. We built trees with RAxML v8.2.12 (*87*) using the ASC_GTRCAT model, with the conditional likelihood method, -asc-corr=lewis, to correct for the ascertainment bias due to using SNPs (*88*). We ran 100 bootstraps, with "-f a" to perform rapid bootstrap analysis. We visualized trees in iTOL (*89*) and collapsed branches with bootstrap support < 75%.

Nucleotide diversity across chromosome 15

To investigate genetic diversity across the affected and unaffected chromosome 15 regions, we used the *PopGenome* package in R (*70*). We imported biallelic SNPs from the whole-genome re-sequencing vcf of the wild-caught forest (n = 13) and prairie (n = 15) mice into *PopGenome*, excluding the 2 forest mice carrying at least one reference allele. We computed statistics in 10-kb windows, with step size of 10 kb. We calculated D_{xy} using the *diversity.stats.between* function and pairwise nucleotide diversity separately for the forest and prairie mice using the *diversity.stats* function.

Demographic model inference

Estimation of 2D site frequency spectra (2DSFS)

To estimate the genome-wide 2DSFS from the wild-caught forest and prairie mice, we used ANGSD (*64*). First, we generated site allele frequency (SAF) files using the command *angsd -gl 1* with the options "-skipTriallelic 1 -sb_pval 0.01 -hwe_pval 0.01 -minMapQ 30 -minQ 20 -minInd 12 -setMinDepthInd 5 -doMajorMinor 1 -doMaf 1 -doSnpStat 1 -doGeno 3 -doPost 2 -doHWE 1 -doCounts 1", using a germline mutation rate of 5.3e-9 (*90*). We also excluded regions with low mappability (e.g. repetitive content), putative inversions, and genic regions +-10 kb by providing a set of callable regions with the option "-rf". Next, we directly estimated the 2DSFS from these files using *realSFS* and setting the number of sites (-nSites) to 5e8, in order to reduce computation time and memory requirements while still sampling a large portion of the genome. In the resulting 2DSFS, we moved sites that were fully fixed in both forest and prairie populations (30, 30) to the monomorphic (0, 0) category for compatibility with *fastsimcoal2*.

Model selection with *fastsimcoal2*

We performed all demographic inference with *fastsimcoal2* version 26 (*91*) and the following options "—multiSFS -d --numBatches 1 -C 1 --maxlhood –numloops 100 --numsims 100000 --logprecision 18 --brentol 0.0001". We initially tested a set of two-population models that encompassed scenarios from no gene flow to multiple admixture events, and constant population size to variable size through time. After an initial model comparison involving 50 independent runs per model, we found that histories with no gene flow and/or constant population size consistently returned much lower likelihoods than all other scenarios. Therefore, we refined our set to 14 competing demographic models that varied in (a) the number of distinct migration rates to describe forest-prairie gene flow, (b) the timing of the divergence between forest and prairie populations, (c) the number and nature of size changes in each population's history, and (d) the size of the ancestral population prior to forest-prairie divergence (Figure S9A). For each of these 14 models, we performed 100 independent parameter runs. We used the top 5 independent maximum likelihood estimates for each model to examine overall model fit and chose the model with the highest likelihood distribution and highest single likelihood estimate: 2Pop2Size2MigRateAncChange (Figure S9B).

To further assess the suitability of our chosen demographic model, we simulated the top model and 7 competing models that also exhibited high likelihood distributions (Figure S9B). Specifically, using the parameter estimates provided by the top scoring run of each model, we performed 100 neutral forward simulations per model of a 500-kb genome using SLiM (*92*). With these simulated data, we calculated F_{ST} and θ_π using *scikit-allel v1.3.2* (https://scikit-allel.readthedocs.io/en/stable/) and LD decay using *PopLDdecay* (*93*). We visually compared these statistics to those calculated from the empirical data to ensure that the top scoring demographic model adequately captured the empirical data (Figure S9C-F). We found that the top scoring model was equivalent to or better than all competing models in capturing F_{ST}, θ_π, and *LD* decay.

Power analysis

To assess our power to infer an accurate demographic history with *fastsimcoal2* when the true history is known, we simulated our top scoring model (2Pop2Size2MigRateAncChange) and reapplied the model fitting described above to the simulated dataset. For this inference, we used the same set of 100 SLiM simulations of 500-kb genomes for the statistical comparisons described above. First, we thinned the simulated data to 1-kb windows separated by 1-kb intervals to restrict the input to unlinked sites. Then, we extracted a 2DSFS from the data using *easySFS.py* (https://github.com/isaacovercast/easySFS) with the options "-a --unfolded". Finally, we modified the count of monomorphic sites (0,0) to reflect the number of invariant sites represented by the thinned dataset.

We next used this simulated 2DSFS to perform 100 independent parameter searches per model for the same 14 competing models used on the empirical data. We examined the top 5 runs per model and observed that two models – 2Pop2Size2MigRateAncChange and 2Pop2Size2MigRate – exhibited the most likely parameter estimates, with highly overlapping distributions that outperformed all other models. Given that the former model corresponded to the true demographic history of the simulated input data and the latter model differed only in ancestral population dynamics, we interpreted this result as support for our power to infer the true demographic history from the empirical data and to distinguish amongst competing histories.

Estimation of confidence intervals for top demographic model

To quantify uncertainty in our demographic history and incorporate this uncertainty into downstream analyses, we applied a non-parametric bootstrapping approach to obtain confidence intervals for all parameters of our top-fitting model. To do this, we generated 100 bootstrap replicates of the empirical SFS using the ANGSD command *realSFS* with the option "-bootstrap 100". For each replicate 2DSFS, we performed 100 independent runs under the top model, 2Pop2Size2MigRateAncChange, using the same options as described above. We extracted the top scoring run for each replicate 2DSFS and used this set of 100 runs to generate 95% confidence intervals for each parameter in the 2Pop2Size2MigRateAncChange model.

SLiM simulations

All simulations were performed using SLiM v3.6 (*92*). For all simulations, unless noted otherwise, we used a mutation rate of 5.3e-9, and a recombination rate of 5e-9 (before scaling), which we calculated from the F2 intercross. To reduce computational time, we scaled parameters by a factor of 100, with population sizes and times divided by 100 and mutation and migration rates multiplied by 100. Recombination rate was scaled according to $r_{scaled} = (1 - (1 - 2r)^n)/2$ where *n* is the scaling factor, in our case, equal to 100 (*92*).

Neutral simulations of *fastsimcoal2* models

To explore the fit of particular *fastsimcoal2* models (see methods section "Model selection with fastsimcoal2"), we performed forward simulations of neutral mutations (*55*). For each of the 8 *fastsimcoal2* models, we performed 100 simulations of a 500-kb genome in SLiM, using the parameter estimates from the *fastsimcoal2* model, with scaling of parameters as described above. We implemented tree-sequencing to reduce run time, and at the final generation, we selected 15 forest and 15 prairie samples for the output tree-sequencing file. Next, we used *pyslim* v0.6 (https://github.com/tskit-dev/pyslim) and *msprime* v0.7.4 (https://github.com/tskit-dev/msprime) to add neutral mutations. We first recapitated the trees so that every site coalesced, using the ancestral population size estimate for N_e. We then added neutral mutations with the *SlimTreeSequence* (pyslim) and *mutate* (msprime) functions with a mutation rate of 5.3e-9 and saved the output as a vcf, which we used for assessing *fastsimcoal2* model selection.

Simulations of the selection history of the inversion

To test whether the inversion evolved neutrally, we simulated the evolution of the inversion under the best-fit *fastsimcoal2* model (*55*). We used the parameter estimates from the best-fit *fastsimcoal2* model, with scaling of parameters as described above. As we were interested in estimating the selection regime necessary for a locus to obtain 90% frequency in the forest and 0% frequency in the prairie, we simulated the inversion as a single locus, evolving under a range of selection coefficients. We used a grid of forest (s_{forest}) and prairie ($s_{prairie}$) selection coefficients for the inversion, with values (before scaling) s_{forest} = 0, 1e-6, 1e-5, 1e-4, 5e-4, 1e-3, 2.5e-3, 5e-3, 7.5e-4, 1e-2, and $s_{prairie}$ = 0, -1e-5, -1e-4, -1e-3, -1e-2. In the simulations, we scaled the selection coefficients, multiplying by the same scaling factor of 100, to keep *Ns* consistent (*92*). These selection coefficients span the four likely selection scenarios: the inversion is evolving under drift ($s_{forest} = 0$, $s_{prairie} = 0$), positive selection in the forest ($s_{forest} > 0$, $s_{prairie} = 0$), negative selection in the prairie ($s_{forest} = 0$, $s_{prairie} < 0$) or divergent selection ($s_{forest} > 0$, $s_{prairie} < 0$). Because the inversion showed additive effects for tail length and coat color in the F2 hybrids, we assumed the inversion was semi-dominant, with heterozygotes experiencing intermediate selection coefficients in forest and prairie. We introduced the inversion as a single mutation into the forest population at times varying from 15 thousand (k) to 9 million (m) generations ago, to allow the inversion enough time to reach high frequency when evolving under drift. When the inversion was introduced prior to the forest-prairie split (2.2m generations ago), the selection coefficient for the inversion was equal to the forest selection coefficient until the forest-prairie split. For each combination of forest and prairie selection coefficients and time of introduction of the inversion, we ran 100 simulations. We then compared the probability that the inversion obtained high frequency (>80%) in the forest population and low frequency (<10%) in the prairie population across the combinations of forest and prairie selection coefficients (Figure S11).

To test whether the selection results were sensitive to uncertainty in the best-fit *fastsimcoal2* parameter estimates, we performed the same selection simulations as described above, using the bootstrapped intervals for the *fastsimcoal2* parameter estimates instead of the top parameter estimates from the model. We grouped the *fastsimcoal2* parameters into three parameter types:

population size parameters, time parameters and migration rate parameters. For each type of parameter, we set the parameters of that type to the 2.5th or 97.5th percentile from the bootstrapped intervals for those parameters; we then ran the selection simulations for 8 scenarios representing all possible combinations of 2.5th and 97.5th percentile values for the 3 parameter types. For these simulations, we introduced the inversion 5k generations after the forest-prairie split time, which ensured that the inversion experienced both forest and prairie selection coefficients immediately. We then compared the probability that the inversion obtained high frequency (>80%) in the forest population and low frequency (<10%) in the prairie population across the combinations of forest and prairie selection coefficients (Figure S12).

Approximate Bayesian computation estimates of inversion selection coefficient

We used approximate Bayesian computation (ABC) to estimate posterior probability distributions for the inversion's selection coefficients in the forest and prairie populations. We ran forward simulations under the best-fit *fastsimcoal2* model in SLiM, scaling parameters as described above. Since we found that the inversion most likely evolved under divergent selection, we drew the positive forest and negative prairie inversion selection coefficients independently from a $\log_{10}$-uniform distribution from -2 to -6 (corresponding to forest selection coefficients from +1e-6 to +0.01 and prairie selection coefficients from -0.01 to -1e-6). We introduced the inversion as a single mutation into the forest population at 4 timepoints: 15k, 150k, 1.5m or 2.2m (time of forest-prairie split) generations ago. For each timepoint, we ran 50,000 simulations and recorded the frequency of the inversion in the forest and prairie populations at the end of each simulation. Using the *abc* package in R (*94*), we set a threshold of 0.5%, keeping the 250 simulations with inversion frequencies closest to the empirical observation of 90% frequency in forest and 0% frequency in prairie. We then used the *loclinear* approach (with *hcorr* = TRUE to correct for heteroscedasticity) to estimate posterior probability distributions for the forest and prairie selection coefficients; the *loclinear* approach weights parameter values based on how close the simulation was to the empirical observations. We report the mean values with 95% confidence intervals from the posterior distributions (Figure S13).

Simulations and ABC estimate of the age of the inversion

To estimate the age of the inversion, we simulated in SLiM a 1-Mb genome, with the inversion as a 500-kb region, evolving under the best-fit *fastsimcoal2* model in SLiM, with parameters scaled as described above (*55*). To simulate the suppression of recombination within the inversion, we introduced the inversion as a point mutation, and when this mutation was found in heterozygote individuals, we used a recombination callback to suppress recombination across the 500-kb region. Outside of the inversion and for all individuals homozygous for the inversion or reference haplotypes, recombination occurred normally with rate 5e-9. The inversion was introduced as a single copy into the forest population, with selection coefficients of 1.8e-3 in the forest and -3e-4 in the prairie and assuming semi-dominance; these selection coefficients were the ABC estimates for the most recent time of introduction of the inversion (t = 15k generations ago), thus allowing the inversion to reach high frequency in the forest population across timepoints. We introduced the inversion at timepoints drawn from a uniform distribution from 0 to 750k generations ago, and ran 25,000 simulations, each with a 10k generation burn-in period (before scaling time parameters). We performed tree-sequencing to reduce computational time, selecting 15 forest and 15 prairie individuals at the end of each simulation. For simulations in which the inversion was neither fixed nor lost across all individuals (*n* = 3,215 simulations), we added neutral mutations at a rate of 5.3e-9 to the tree-sequencing file using *pyslim* and *msprime*. From the output vcfs, we next computed D_{xy} and F_{ST} in the 500-kb inversion region, between forest individuals homozygous for the inversion and prairie individuals homozygous for the reference allele, using *scikit-allel*. Finally, we used the *abc* package in R to estimate the age of the inversion. Based on empirical values for D_{xy} and F_{ST} as calculated in PopGenome (see "Nucleotide diversity across chromosome 15" section) between the inversion and reference alleles, we set a threshold of 10%, keeping a total of 321 simulations with D_{xy} and F_{ST} closest to the empirical values. We then used the *loclinear* approach (with *hcorr* = TRUE to correct for heteroscedasticity) to create a posterior probability distribution of the age of the inversion based on D_{xy} and F_{ST} summary statistics.

Simulations of fitness effects due to suppression of recombination within the inversion

Two-beneficial-locus model without deleterious mutations

To explore the fitness effects from suppression of recombination within the inversion, we first simulated a model with two beneficial loci, A and B, in the forest population (*55*). Under the best-fit *fastsimcoal2* model, we introduced A and B into the forest population on a single haplotype at 250k generations ago (ABC estimate for the age of the inversion). The haplotype was either a standard, freely recombining haplotype, or an inversion (with complete suppression of recombination in heterozygotes); this allowed us to test whether the beneficial loci were more likely to spread on an inversion or a standard haplotype initially carrying the two beneficial loci. A and B were modeled as beneficial in the forest population, with selection coefficients s_A+s_B = 3e-4 (ABC estimate for the inversion's selection coefficient in forest) and as deleterious in the prairie population, with selection coefficients s_{pA}+s_{pB} = -1e-2, assuming semi-dominance for both loci. On the standard haplotype, we simulated the following distances between A and B: 100 bp, 1 kb, 10 kb, 100 kb, 1 Mb, 10 Mb (since the inversion is 41 Mb). On the inversion, since A and B were completely linked, we did not vary the distances between A and B when simulating the inversion. This allowed us to compare the distances between two beneficial loci for which an inversion confers an advantage from suppressing recombination between A and B. We also varied the ratio of the forest selection coefficients for A and B (s_A/s_B = 1.0, 0.5, 0.1, or 0.01) to explore how the relative strengths of selection on two beneficial loci changes the dynamics of their evolution on a standard or inverted haplotype. We ran 500,000 simulations for all combinations of A and B distances (100 bp, 1 kb, 10 kb, 100 kb, 1 Mb, 10 Mb) and ratio of selection coefficients (1.0, 0.5, 0.1, 0.01), and we ran 500,000 simulations for the inversion. At the final generation for each simulation, we recorded the frequencies of A and B in the forest population.

Using the two-beneficial-locus simulation results, we characterized the spread of A and B when they were introduced into the forest population on the standard haplotype versus the inversion. Across all simulations for each scenario, we calculated the probability that A (or B) was lost in the forest population, the frequency of A when A was not lost (or B when B was not lost) in the forest population (which represents the migration-selection equilibrium frequencies), and the mean fitness of the forest population (Figure S14). The mean fitness was calculated as the mean frequency of A (over all simulations, including simulations where A was lost) multiplied by the selection coefficient of A plus the mean frequency of B (over all simulations, including simulations where B was lost) multiplied by the selection coefficient of B in the forest population, where a fitness of zero represents the initial forest population's fitness when both A and B are absent (Figure S14). We then used permutation tests to assess whether the mean fitness of the forest population obtained when the inversion was introduced significantly differed from the mean forest fitness when the standard haplotype was introduced. To create a null distribution of inversion-standard forest fitness differences for each scenario of A and B distances and selection coefficients, we randomly set inversion versus standard haplotype assignments to each simulation, sampling without replacement, and computed the difference in mean forest fitness between the inversion versus standard haplotype simulations. We performed 1,000 permutations and then compared the true inversion-standard difference in mean forest fitness from our simulations to the null distribution to obtain the probability of obtaining that fitness difference from chance alone.

Finally, we computed the theoretical predictions for an inversion's fitness gain in the forest population from linking together multiple adaptive loci and compared these predictions to our simulation results. In a two-beneficial-locus model (assuming $s > m$ for both loci), the mean equilibrium fitness of a locally adapting population is defined by the following equation (*5*):

$$\frac{1}{2}\left(\alpha + \beta - \rho - 4m + \sqrt{(\alpha + \beta + \rho)^2 - 8m\rho}\right)$$

where m = migration rate into the locally adapting population, ρ = recombination rate between the two beneficial loci, and α, β = selection coefficients for two beneficial loci (*5*). When there is no recombination between beneficial loci, such as with an inversion, the equilibrium fitness for the locally adapting population is $\alpha + \beta - 2m$. Thus, the fitness gain obtained from an inversion over a standard haplotype is:

$$\frac{1}{2}\left(\alpha + \beta + \rho - \sqrt{(\alpha + \beta + \rho)^2 - 8m\rho}\right)$$

assuming that $\alpha > m, \beta > m$, and the inversion completely suppresses recombination between loci. For the variables in this equation, we set m = the final prairie-to-forest migration rate estimated as 6.26e-6 from the *fastsimcoal2* model, $\alpha = \beta$ = 1.5e-4 (0.5 multiplied by the inversion's forest selection coefficient), and ρ = 5e-9 multiplied by the distance between A and B. To compare these predicted fitness gains with the simulation results, we subset the simulation results to simulations for which A and B were not lost in the forest, since the theoretical equation assumes that neither A nor B is lost in the locally adapting population (Figure S15).

Two-beneficial-locus model with deleterious mutations

In addition to linking together beneficial mutations, inversions can also carry deleterious mutations. To explore the effects of hitchhiking deleterious mutations, we performed additional simulations of the two-beneficial-locus model, while including deleterious mutations (*55*). We simulated the same two-beneficial-locus model under the best-fit *fastsimcoal2* model as described above, with the following differences: (1) we simulated a 400-kb genome, which is approximately the inversion length (41 Mb) divided by the scalar of 100; (2) we included six 20-kb functional regions, separated by 30-kb non-functional regions, which represent a generous estimate of the density of functional regions (including exons, introns, UTRs) within the inversion (frequently only exons are included as functional content); (3) the two beneficial loci A and B were located 100-kb apart, within functional regions, with ratio of selection coefficients equal to either 1.0 or 0.1; (4) deleterious mutations were introduced into the forest and prairie populations at a mutation rate of 5.3e-9, within the functional regions only, according to four distributions of fitness effects (DFEs) (as described in (*14*, *95*). The DFEs included f_0 with neutral mutations only ($2Ns = 0$), f_1 with neutral and weakly deleterious mutations (50% $2Ns = 0$, 50% $-10 < 2Ns < -1$), f_2 with neutral, weakly and moderately deleterious mutations (33% $2Ns = 0$, 33% $-10 < 2Ns < -1$, 33% $-100 < 2Ns < -10$) and f_3 with neutral, weakly, moderately and strongly deleterious mutations (25% $2Ns = 0$, 25% $-10 < 2Ns < -1$, 25% $-100 < 2Ns < -10$, 25% $-1000 < 2Ns < -100$) where N is the forest-prairie ancestral population size of 4.2e6. All deleterious mutations were assumed to be semi-dominant. Both the inversion and standard haplotypes were introduced into the forest population 250k generations ago, following a burn-in period of 10k generations (before scaling time parameters). The inversion was introduced as a single mutation, which completely suppressed recombination across the 400-kb region with a recombination callback when found as heterozygous. At the end of each simulation, the final frequencies of A and B were recorded. Since we were specifically interested in the spread of A and B in the presence of deleterious mutations on an inversion or standard haplotype, we computed the final forest fitness using only the frequencies of A and B. We ran 100,000 simulations per scenario, and for each scenario, the mean forest fitness was calculated as the mean frequency of A (across all simulations of that scenario) multiplied by the selection coefficient of A plus the mean frequency of B (across all simulations of that scenario) multiplied by the selection coefficient of B. We also reported the probability that A (or B) was lost and the mean frequency of A when A was not lost (or B when B was not lost) (Figure S16).

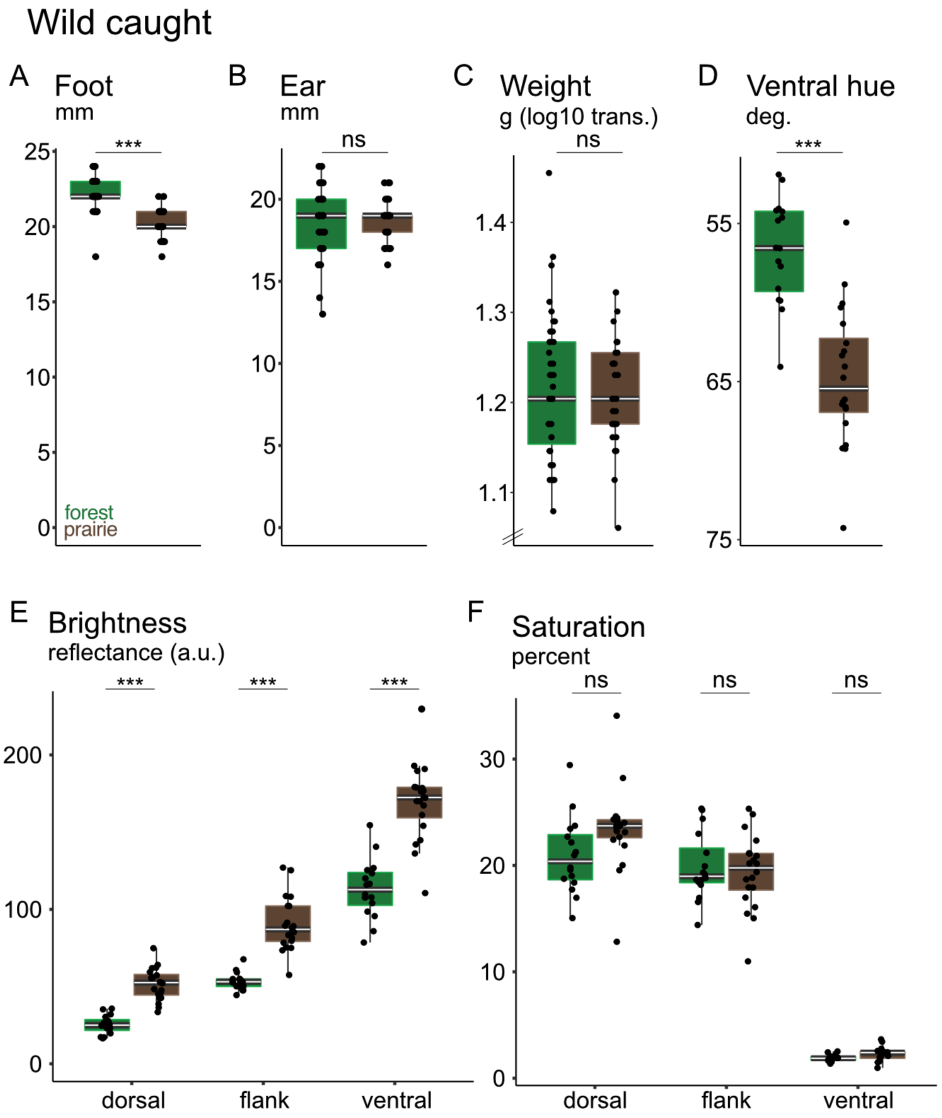

Fig. S1. Additional ecotype-related traits of wild-caught mice. Phenotypes of adult, wild-caught mice from the forest (green, left) and prairie (brown, right) ecotypes. (**A**) Hindfoot length (*n* = 33 forest, 29 prairie), (**B**) ear length (*n* = 33 forest, 29 prairie), (**C**) weight (shown after $\log_{10}$ transformation; *n* = 39 forest, 30 prairie), and (**D**) ventral hue. (**E**) brightness and (**F**) saturation of all three body regions (*n* = 16 forest, 20 prairie for all pigment traits). Symbols: ns = $p > 0.05$; *** = $p < 0.001$ (two-sided Welch's t-tests); deg. = degrees; a.u. = arbitrary units of reflectance. Boxplots indicate the median (center white line), the 25th and 75th percentiles (box extent); whiskers show largest or smallest value within 1.5 times the inter-quartile range. Black dots show individual data points.

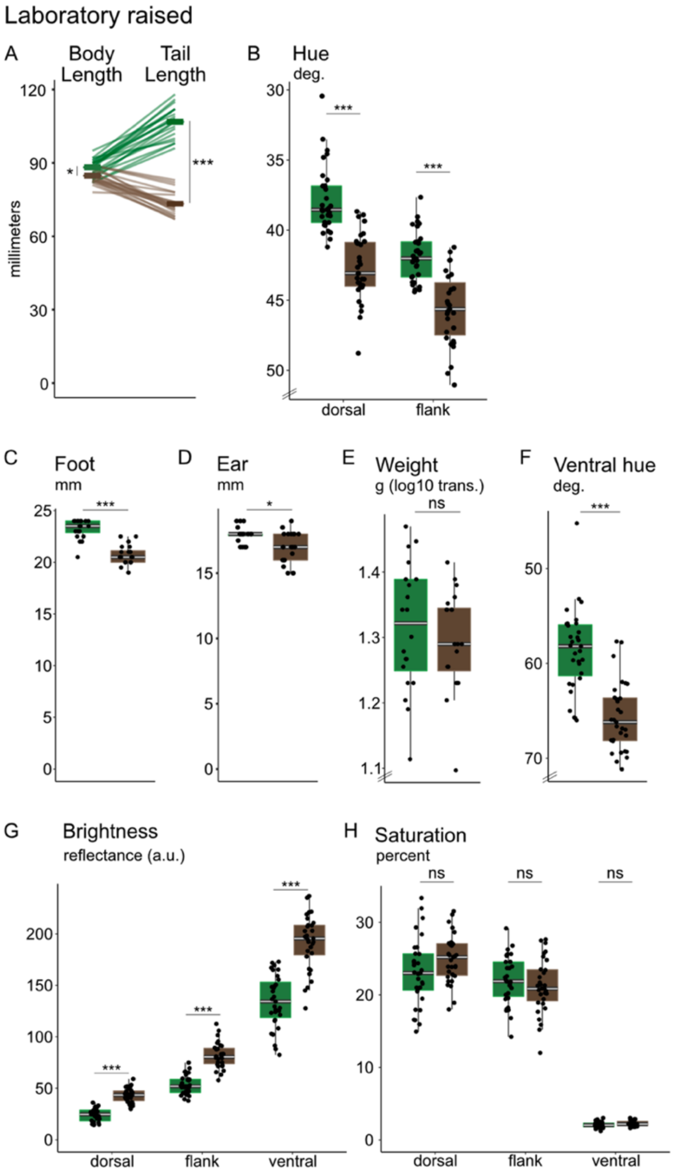

Fig. S2. Phenotypes of laboratory-born mice are consistent with wild-caught specimens. Phenotypes of laboratory-born adult forest (green) and prairie (brown) mice, aged 60-70 days (*n* = 20 for each ecotype, body measurements; *n* = 31 for each ecotype, pigment traits). (**A**) Body length (excluding tail) and tail length. Lines connect measurements for the same individual. Means shown in bold. (**B**) Dorsal and flank hue. (**C**) Hindfoot length, (**D**) Ear length, (**E**) Weight (after $\log_{10}$ transformation), and (**F**) Ventral hue. (**G**) Brightness and (**H**) Saturation for all three body regions. Symbols: ns = $p > 0.05$; * = $p < 0.05$; *** = $p < 0.001$ (two-sided Welch's t-tests); deg. = degrees; a.u. = arbitrary units of reflectance. Boxplots indicate the median (center white line), the 25th and 75th percentiles (box extent); whiskers show largest or smallest value within 1.5 times the inter-quartile range. Black dots show individual data points.

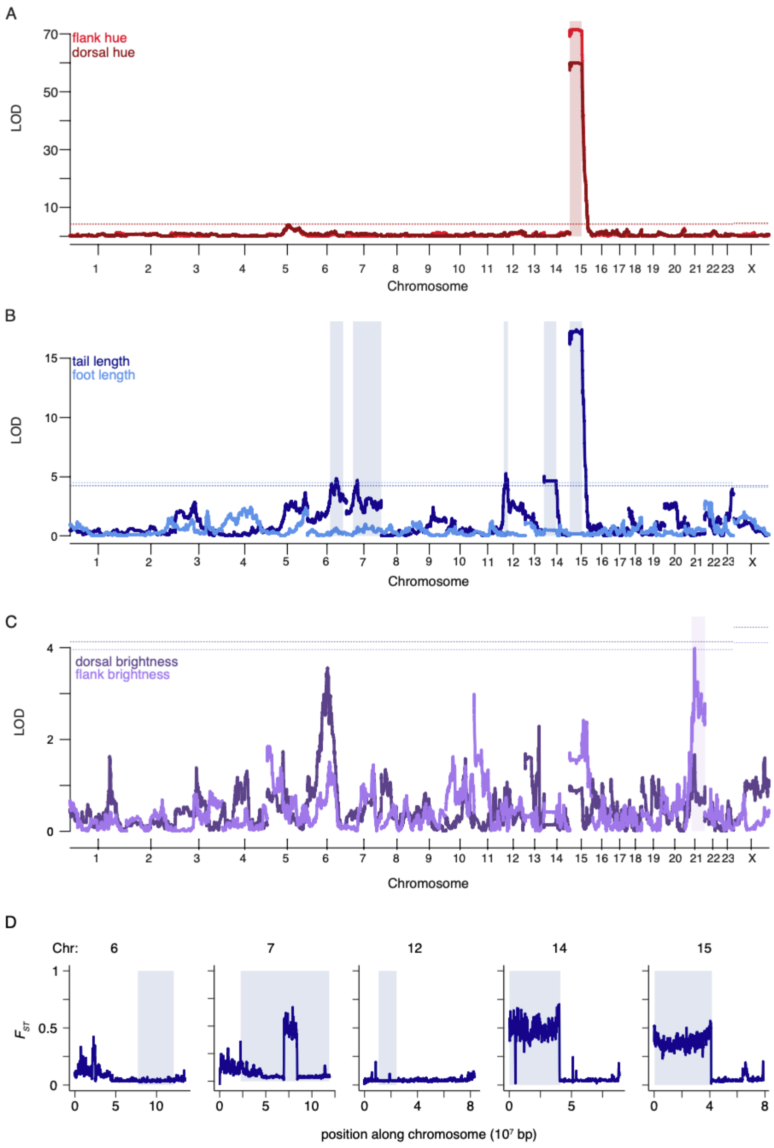

Fig. S3. QTL maps for all traits that differed between wild-caught forest and prairie mice. (**A**) Flank hue (light red) and dorsal hue (dark red). (**B**) Tail length (dark blue) and hindfoot length (light blue), with body length included as an additive covariate. (**C**) Dorsal brightness (dark purple) and flank brightness (light purple). The peak for flank brightness on chromosome 21 is transgressive. LOD = log of the odds score. Physical distance (basepairs) is shown on the x-axis; axis labels indicate the center of each chromosome. Dotted lines indicate the genome-wide significance threshold ($\alpha = 0.05$) based on permutations, and shaded rectangles indicate the 95% Bayes' credible intervals for all chromosomes with significant QTL peaks. $n = 542$ (tail), 455 (foot), 541 (pigment). (**D**) F_{ST} between forest and prairie mice for the five chromosomes with significant tail length QTL. F_{ST} was estimated in 10-kb windows with step size of 1 kb and smoothed with *loess* regression. Shaded rectangles indicate the 95% Bayes' credible intervals for the significant tail length QTL peaks.

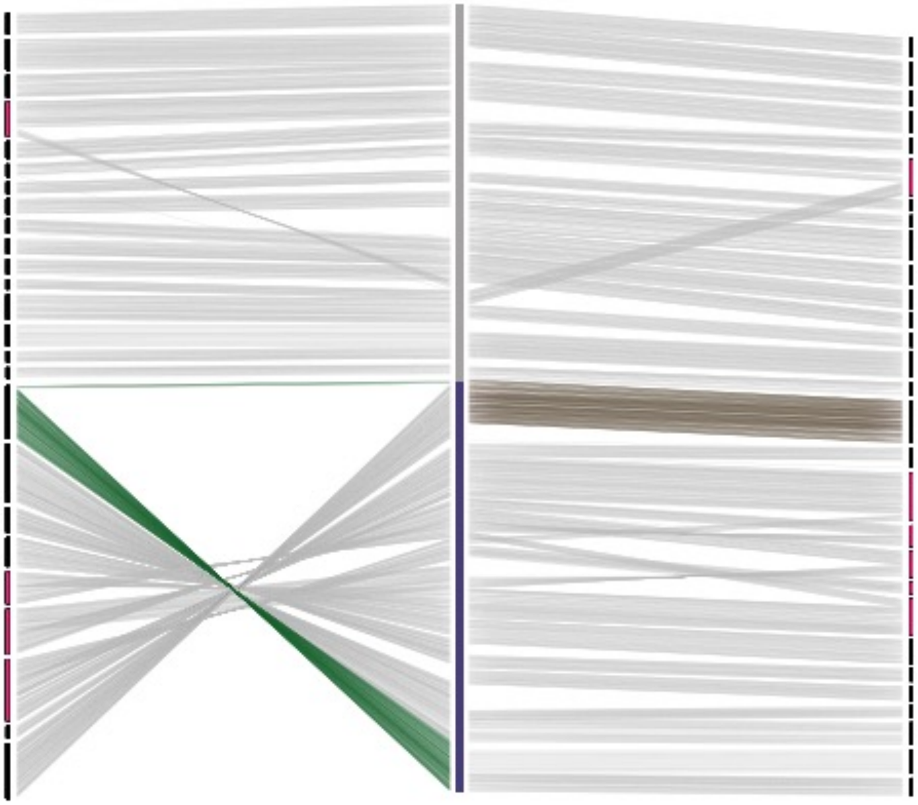

Fig. S4. Alignments of the long-read sequencing-based contigs to the reference genome for chromosome 15. All contigs greater than 1 Mb that align to chromosome 15 are shown for forest (left) and prairie (right) with reference sequence in center (unaffected region = gray; affected region = purple). Contigs are ordered by mapping position to reference genome, except forest contigs in the inversion region that are ordered by inferred position in forest genome. The forest contig that identified the inversion (green) and the prairie contig that spans the inversion breakpoint (brown) are highlighted. Contigs containing forest-prairie rearrangements that are shared relative to the reference genome are indicated (red).

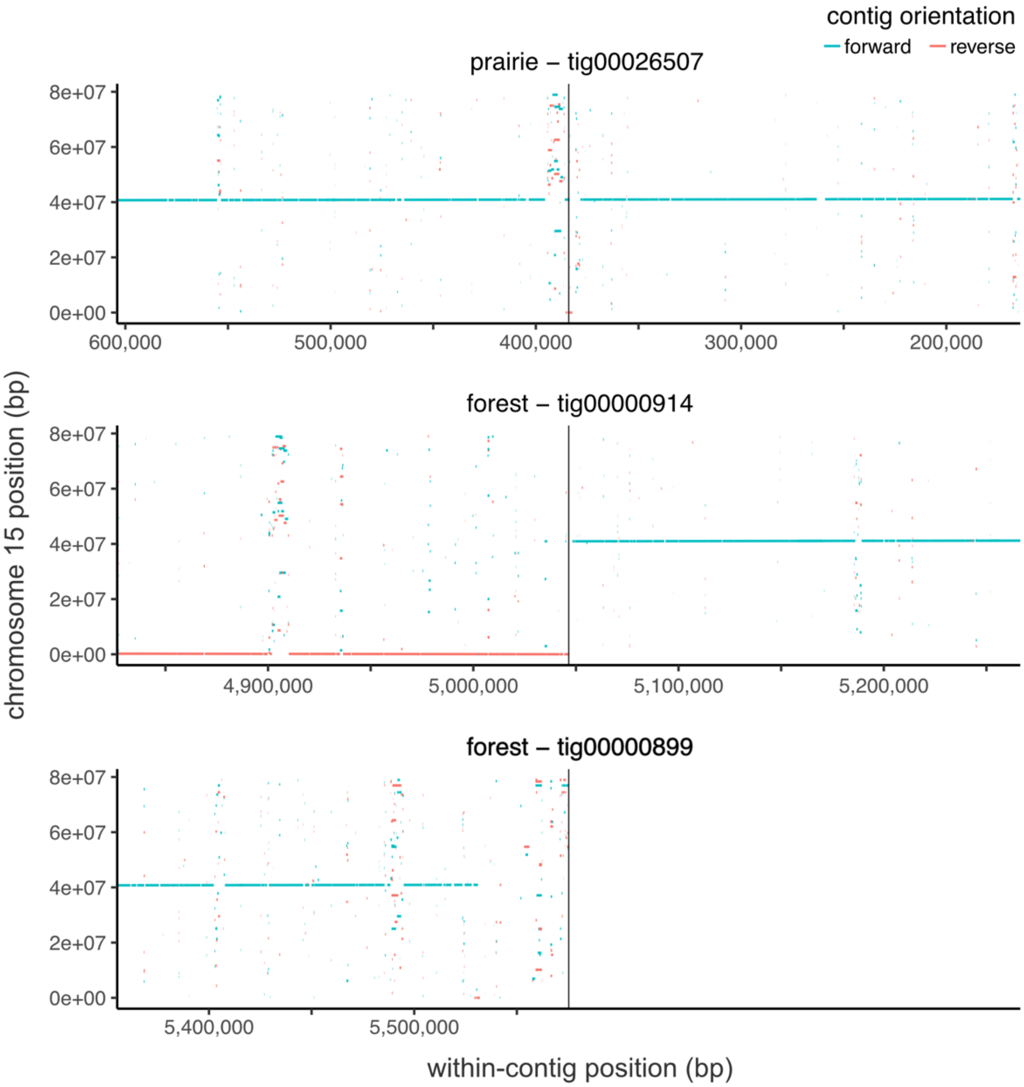

Fig. S5. Contig alignments from forest and prairie sequences encompassing the inversion breakpoint. Alignments of the three long-read sequencing-based contigs relevant for the inversion breakpoint: prairie contig spanning the breakpoint (top), forest contig spanning the breakpoint (middle), and forest contig containing sequence adjacent to the breakpoint (bottom). Contigs are aligned to chromosome 15 from the reference genome (y-axis). Vertical line indicates the identified breakpoint. Colors (blue, red) indicate alignment direction with respect to the reference genome.

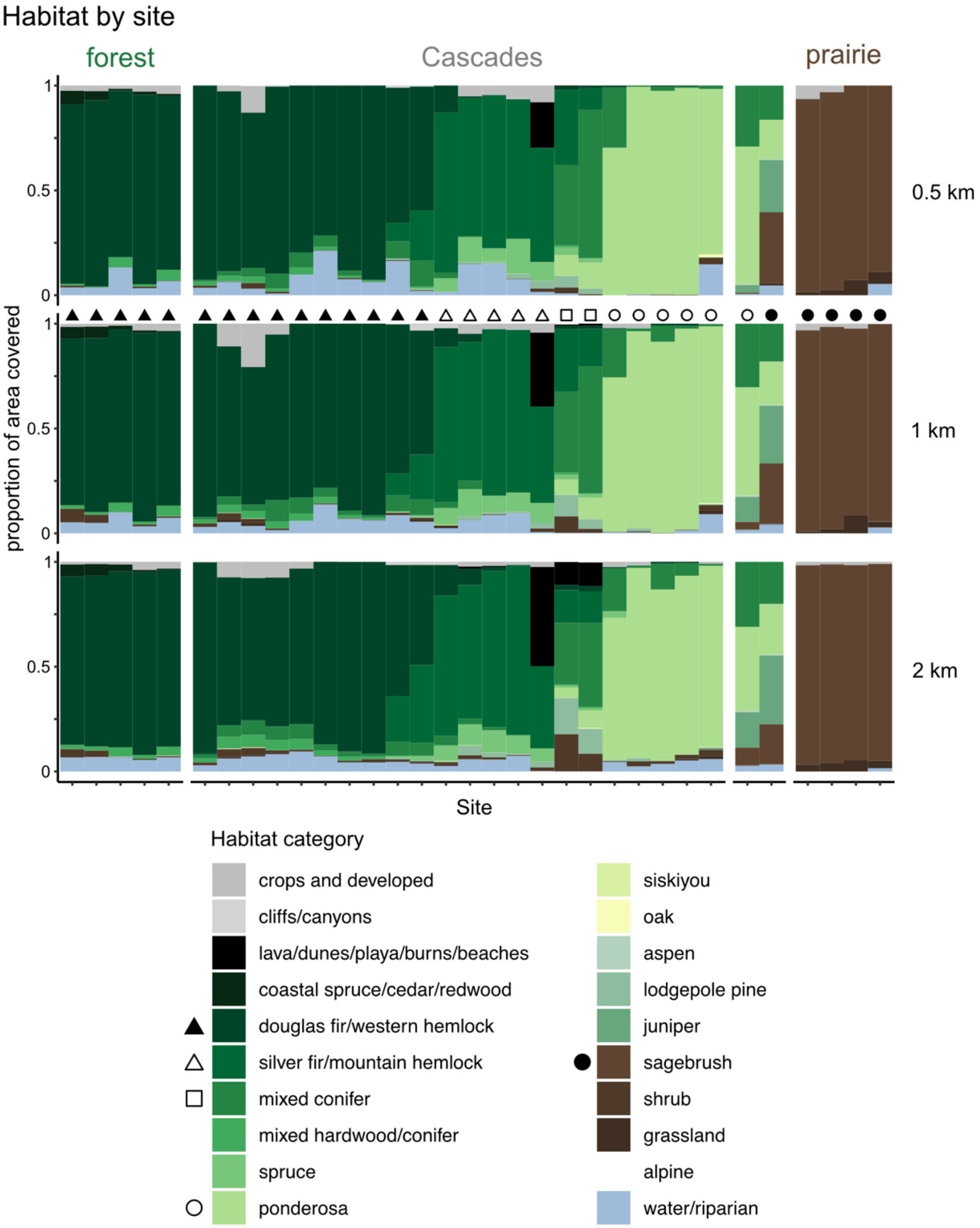

Fig. S6. Habitat categories at sampled sites. The proportion of the area within 0.5 km (top), 1 km (middle) and 2 km (bottom) covered by each habitat category. Data drawn from the habitat map by the Oregon Biodiversity Information Center of the Institute for Natural Resources at Portland State University, and habitat categories were binned across age categories as shown in Table S5. The habitat categories shown in Figure 4 represent the habitat that covers the most area within 1 km of the site (symbols above 1 km plot).

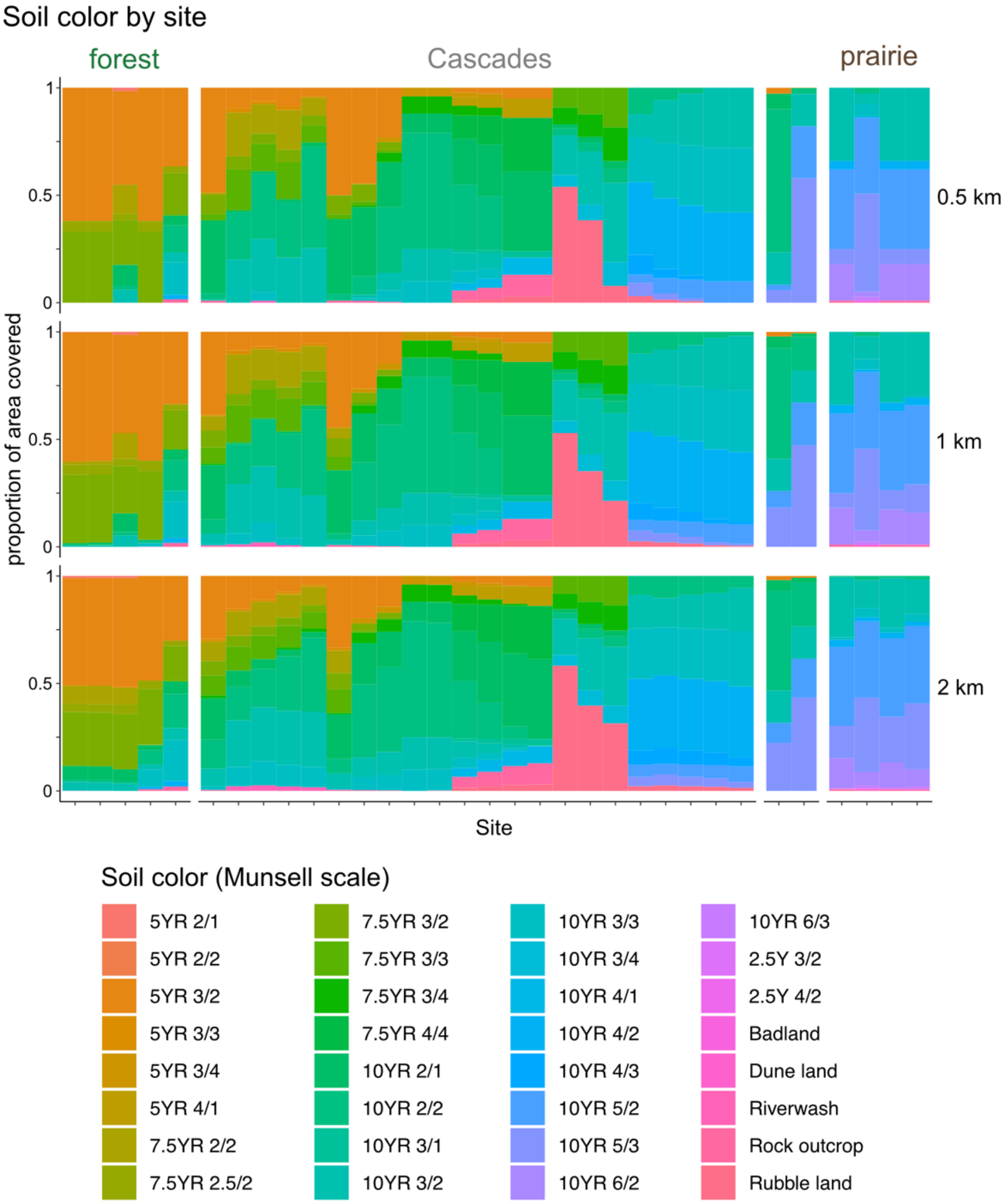

Fig. S7. Soil characteristics at sampled sites. The proportion of top-layer soil with each Munsell scale color (shown as hue value/chroma) within 0.5 km (top), 1 km (middle), or 2 km (bottom) of each sampled site. Sites are ordered by transect distance (i.e., distance east from the central point). Hue values shown in Figure 4 are the weighted average of the Munsell hue for the 1 km radius shown, after excluding regions with no soil series data (i.e., badland, dune land, riverwash, rock outcrop, and rubble land categories).

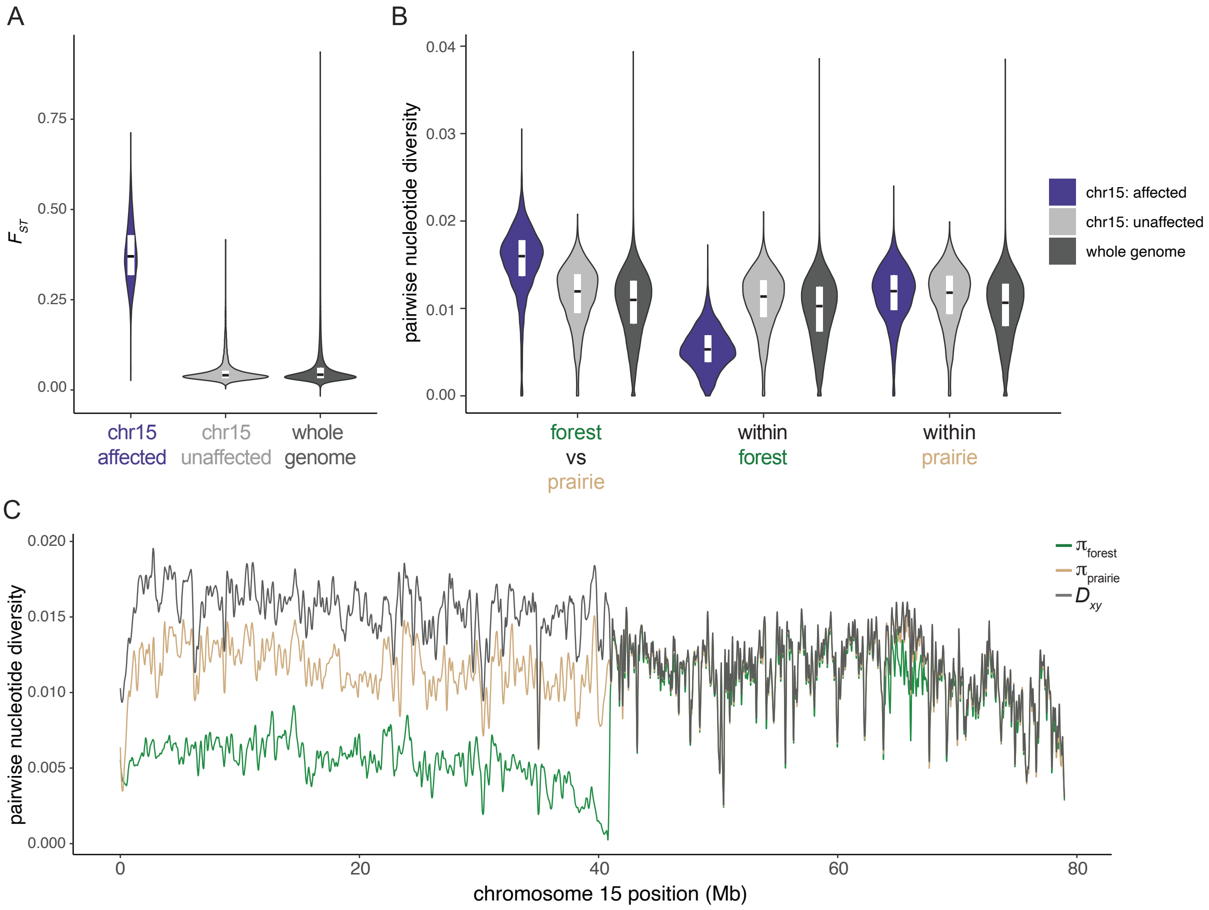

Fig. S8. Genetic differentiation statistics for chromosome 15. (**A**) F_{ST} between wild-caught forest (n = 15) and prairie (n = 15) mice calculated in 10-kb windows with step size of 1 kb in the affected (purple, chr15:0-40 Mb) and unaffected (light gray, chr15:41-79 Mb) regions of chromosome 15, and across the whole genome excluding the affected region of chromosome 15 (dark gray). (**B**) Pairwise nucleotide diversity for forest v. prairie ecotypes (D_{xy}), within forest mice (π_{forest}), or within prairie ($\pi_{prairie}$) shown for the affected (purple, chr15:0-41 Mb) and unaffected (light gray, chr15:41-79 Mb) regions. Nucleotide diversity statistics were computed in 10-kb windows with step size of 10 kb. All prairie mice (n = 15) but only forest mice homozygous for the inversion (n = 13) were included in nucleotide diversity analyses. For all violin plots, white boxes represent first and third quartiles, with median shown as black line. (**C**) Smoothed nucleotide diversity shown across chromosome 15 (green = π_{forest}, tan = $\pi_{prairie}$, gray = D_{xy}).

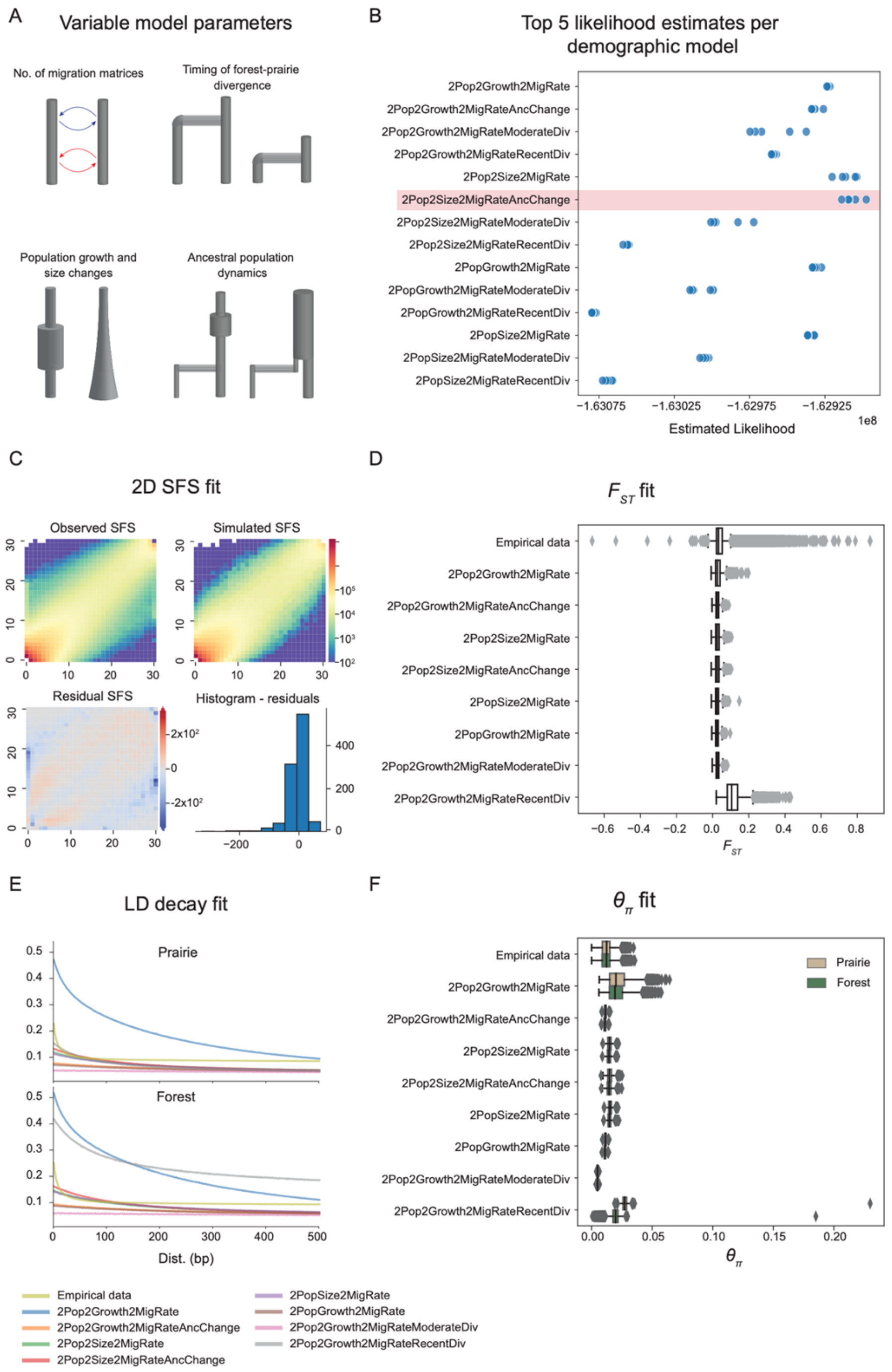

Fig. S9. Demographic model inference. (**A**) Illustrations highlighting parameters that varied among competing demographic models. (**B**) Top five maximum likelihood scores plotted for 14 competing models, pulled from a set of 100 independent parameter searches per model. Highlight indicates model with the highest likelihood. (**C**) Comparison between observed two-dimensional site frequency spectra (2DSFS) for top-scoring model, 2Pop2Size2MigRateAnchChange. (**D-F**) Comparisons of population-genetic parameters between empirical data and simulations of 8 competing demographic models with similar likelihood

distributions. The most likely model, 2Pop2Size2MigRateAncChange, is at least equivalent to, if not better than, all competing models for all parameters. (**D**) Distributions of F_{ST} as calculated from 10-kb genomic windows. (**E**) Linkage disequilibrium (LD) decay, per population. (**F**) Distributions of θ_π per population, as calculated from 10-kb genomic windows.
In (D) and (F), boxes show median, with first and third quartiles; whiskers show largest or smallest value within 1.5 times the inter-quartile range.

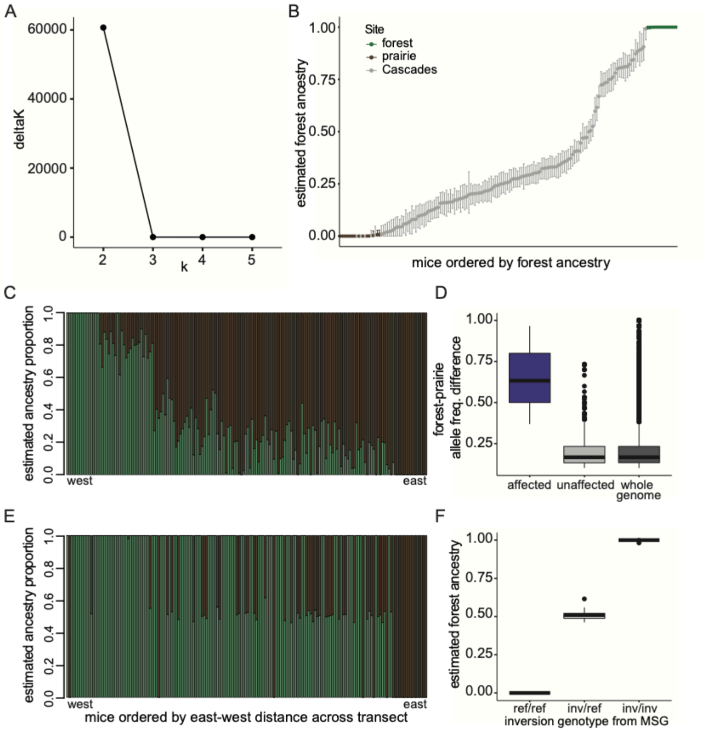

Fig. S10. Genomic ancestry estimates. (**A**) Best K method (CLUMPAK) shows ancestry for all transect mice (*n* = 136) was best assigned to two clusters (*k* = 2). (**B**) Bootstrapped confidence intervals for genome-wide ngsAdmix ancestry estimates. Points show percent of ancestry assigned to cluster 1 (forest ancestry) with the full SNP set, and bars show 95% confidence interval on forest ancestry from bootstrapping. Mice are ordered by their forest ancestry estimates from full SNP set (*n* = 472,692), and colored by site (forest = green, prairie = brown, Cascades = gray). (**C**) Genome-wide ngsAdmix ancestry estimates for forest, Cascades and prairie mice for all autosomes, excluding affected region of chromosome 15. Mice are ordered by their distance along an east-west axis of their capture sites (West = left, East = right). All forest mice (*n* = 15, left) are assigned 100% ancestry in cluster 1, whereas all prairie mice (*n* = 15, right) are assigned 100% ancestry in cluster 2, suggesting that the two ancestry groups correspond to forest (green) and prairie ancestry (brown). Central Cascades mice (*n* = 136) have varying proportions of ancestry assignments (middle). (**D**) Allele frequency difference between forest and prairie populations for the top quartile of differentiated SNPs that were used in ngsAdmix for determining genome-wide ancestry (whole genome, dark gray) and ancestry for the inversion region (affected, purple). ngsAdmix SNPs from the unaffected region of chromosome 15 (unaffected, light gray) are also shown for comparison with the affected region. (**E**) ngsAdmix ancestry estimates for the affected region of chromosome 15 (*n* = 7,090 SNPs) to confirm inversion genotypes, with mice ordered as in (C). (**F**) ngsAdmix estimates for forest ancestry at the affected region agree with inversion genotypes as determined by MSG. In (D) and (F), boxes show median, with first and third quartiles; whiskers show largest or smallest value within 1.5 times the inter-quartile range.

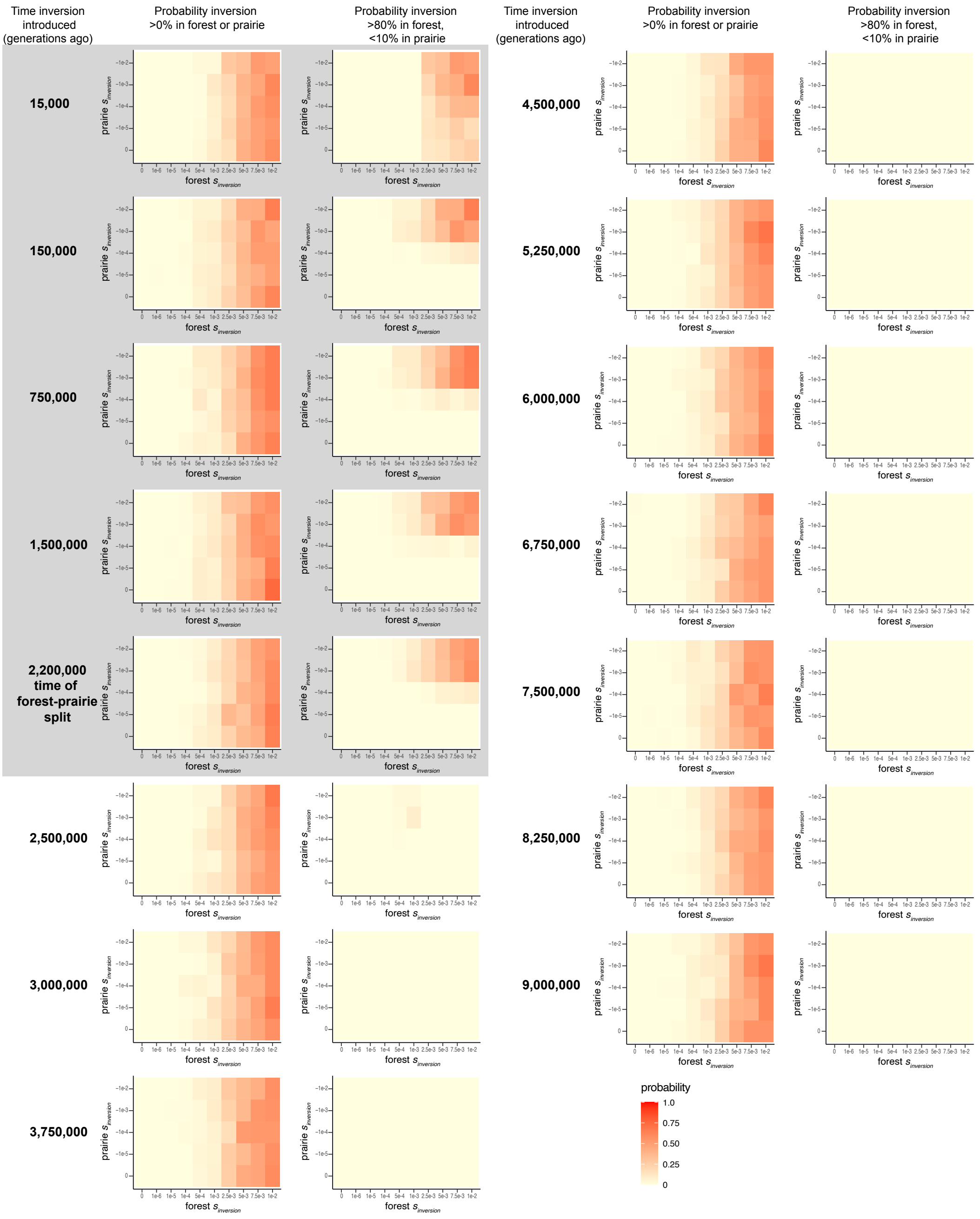

Fig. S11. Simulation results for selection history of the inversion. In SLiM, the inversion was introduced at varying timepoints (15 thousand to 9 million generations ago), with a range of forest (0 to 0.01) and prairie (0 to -0.01) selection coefficients. Heatmaps show the probability of the inversion reaching frequency >0% in the forest and/or prairie population (left) and reaching >80% frequency in the forest and <10% frequency in the prairie (right) for varying positive forest and negative prairie selection coefficients. Gray box highlights simulations for which the inversion was introduced following the forest-prairie split.

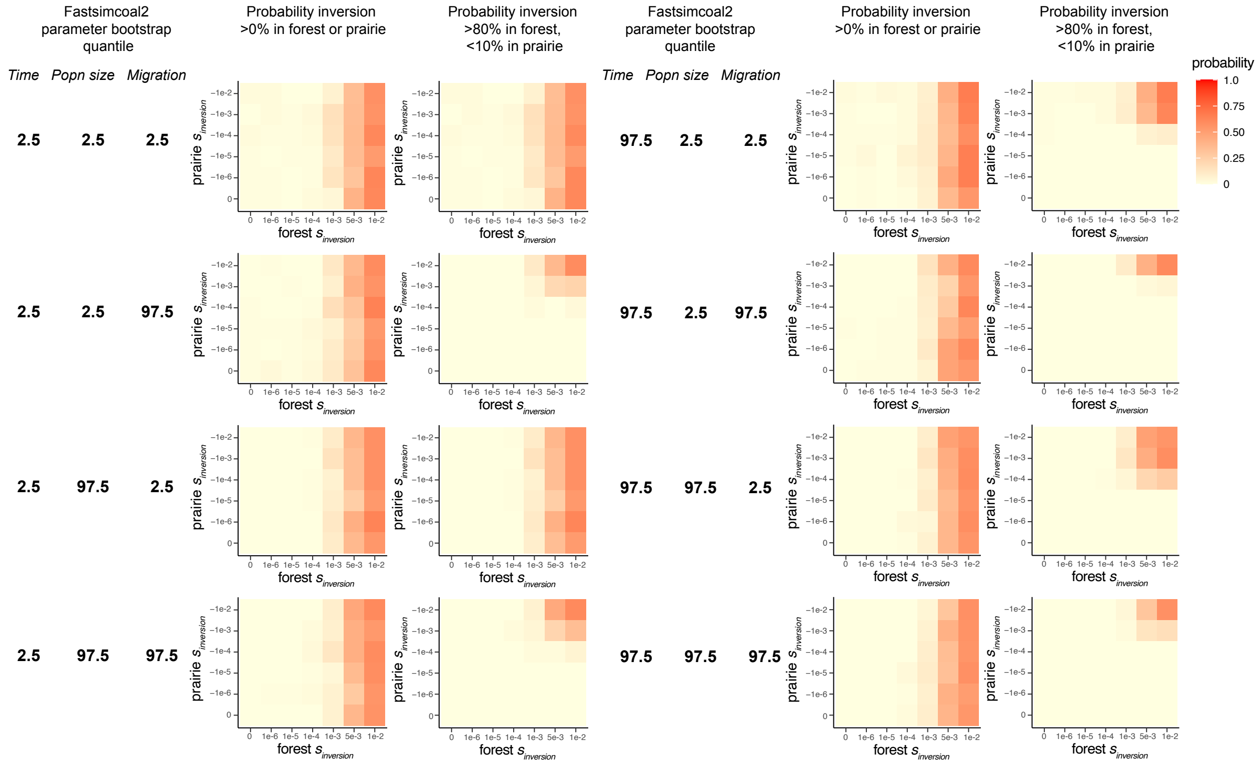

Fig. S12. Effects of uncertainty in *fastsimcoal2* parameter estimates on inversion selection simulations. Selection history of the inversion was simulated under the demographic model using the 2.5th or 97.5th quantile for *fastsimcoal2* estimates of time, population size and migration rate parameters. Heatmaps show the probability of the inversion reaching frequency >0% in the forest and/or prairie population (left) and the probability of the inversion reaching >80% frequency in the forest and <10% frequency in the prairie (right) for varying positive forest and negative prairie selection coefficients.

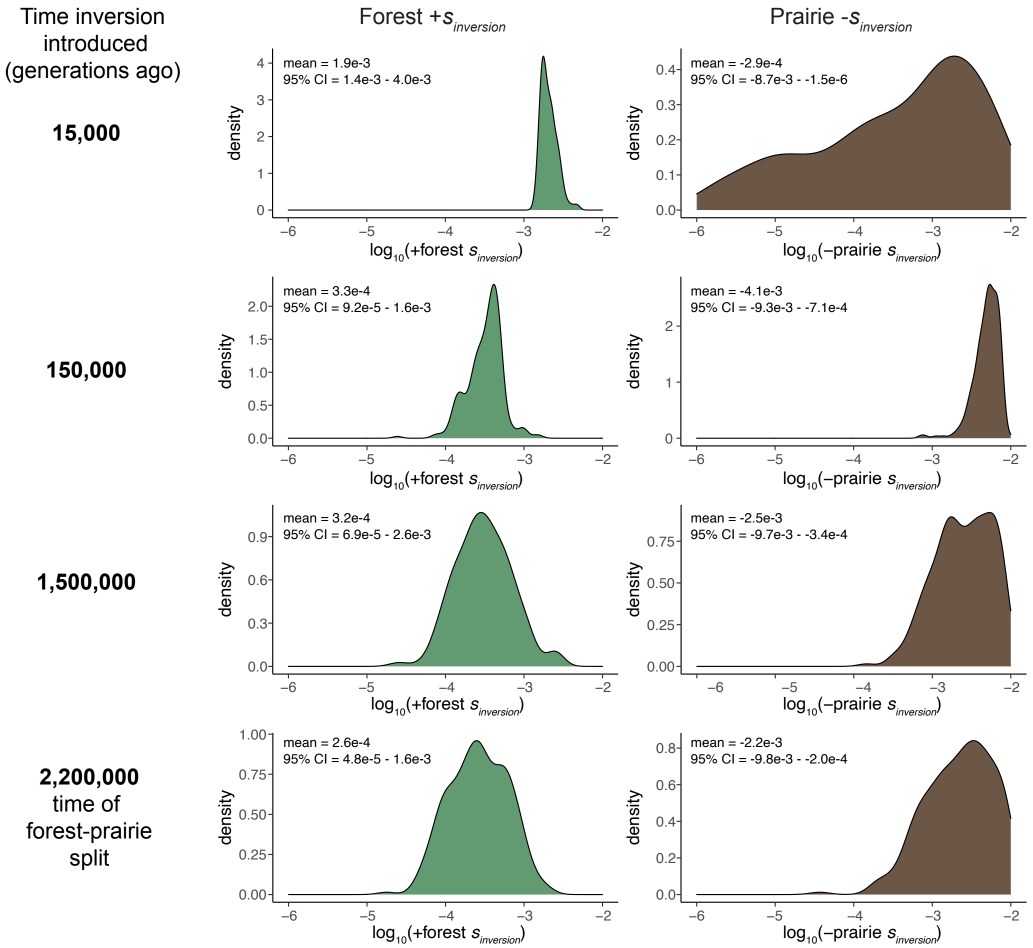

Fig. S13. ABC selection estimates for the inversion. Posterior probability distributions for the inversion's selection coefficient ($s_{inversion}$) in forest (left, green) and prairie (right, brown). Posterior distributions of the inversion's selection coefficient are shown for when the inversion is introduced into the forest population under the demographic model at four timepoints. Note: forest selection coefficients are estimated to be positive, and prairie selection coefficients negative. CI = confidence interval.

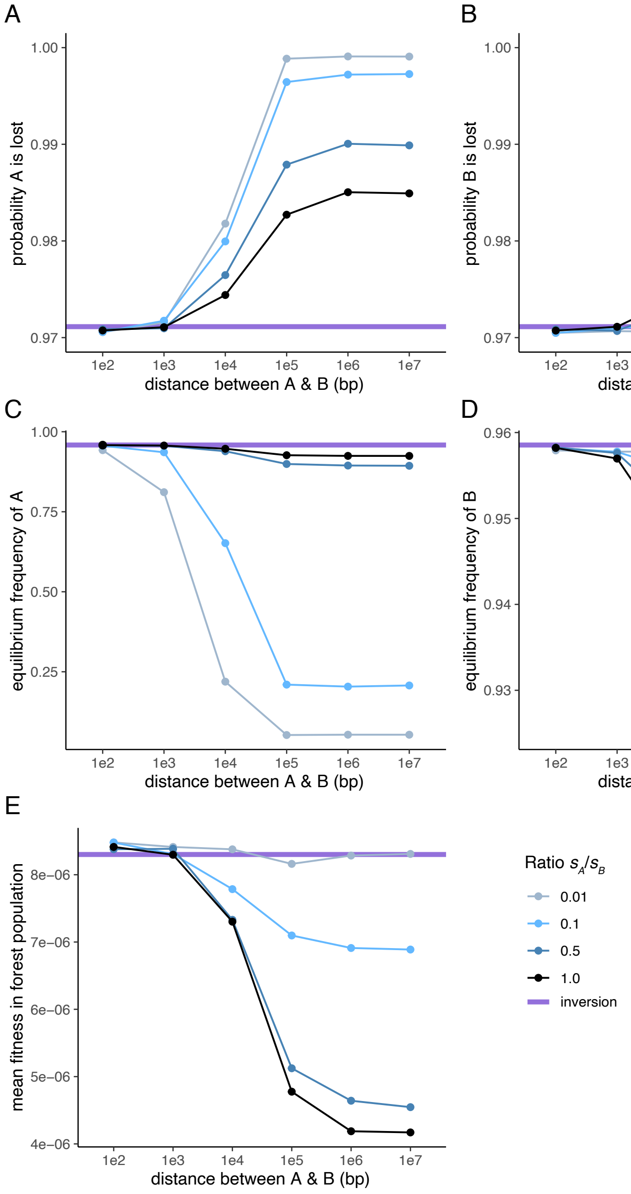

Fig. S14. Simulation results from the two-locus model of recombination effects. A two-locus model was simulated in SLiM under the demographic model, where the two loci (A & B) were introduced into the forest population on an inversion or on a standard haplotype. On the standard haplotype, distances between A and B varied from 100 bp to 10 Mb. Ratio of selection coefficients for A (s_A) and B (s_B) varied with $s_A/s_B = 0.01$ to 1.0 on the standard haplotype, with $s_A+s_B = 3\text{e-}4$ for both the inversion and standard haplotype, with 500,000 simulations per scenario. (**A**) Probability that A was lost in the simulation. (**B**) Probability that B was lost. (**C**) Mean frequency of A in forest when A was not lost. (**D**) Mean frequency of B in forest when B was not lost. (**E**) The mean forest fitness across all simulations. **Horizontal purple line:** simulation results for the inversion (no recombination between A & B).

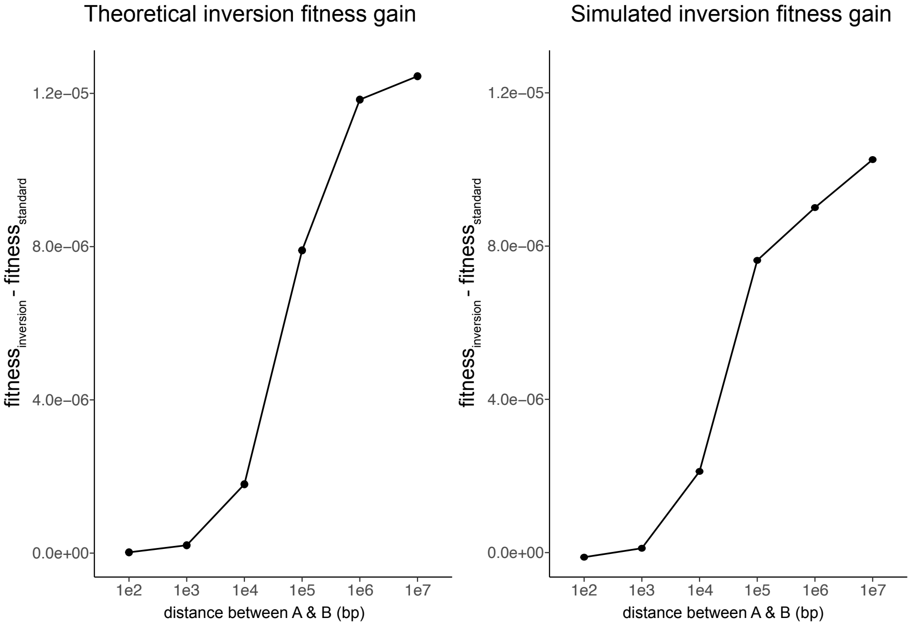

Fig. S15. Theoretical versus simulated fitness gain from an inversion. Left: The theoretical fitness gain from an inversion over a standard haplotype with two adaptive loci computed using estimated migration rates, recombination rates and the inversion selection coefficient, and assuming the two loci had equal selection coefficients. The theoretical fitness gain of an inversion is equal to 2*m* (the migration load of an allele, *m* = migration rate) when the loci are unlinked (e.g. distance between A and B = 10 Mb). **Right:** The simulated fitness gain of the inversion where A and B had equal selection coefficients. Since the theoretical predictions assume A and B are not lost because their selection coefficients are greater than *m*, only simulations where A and B are not lost are shown.

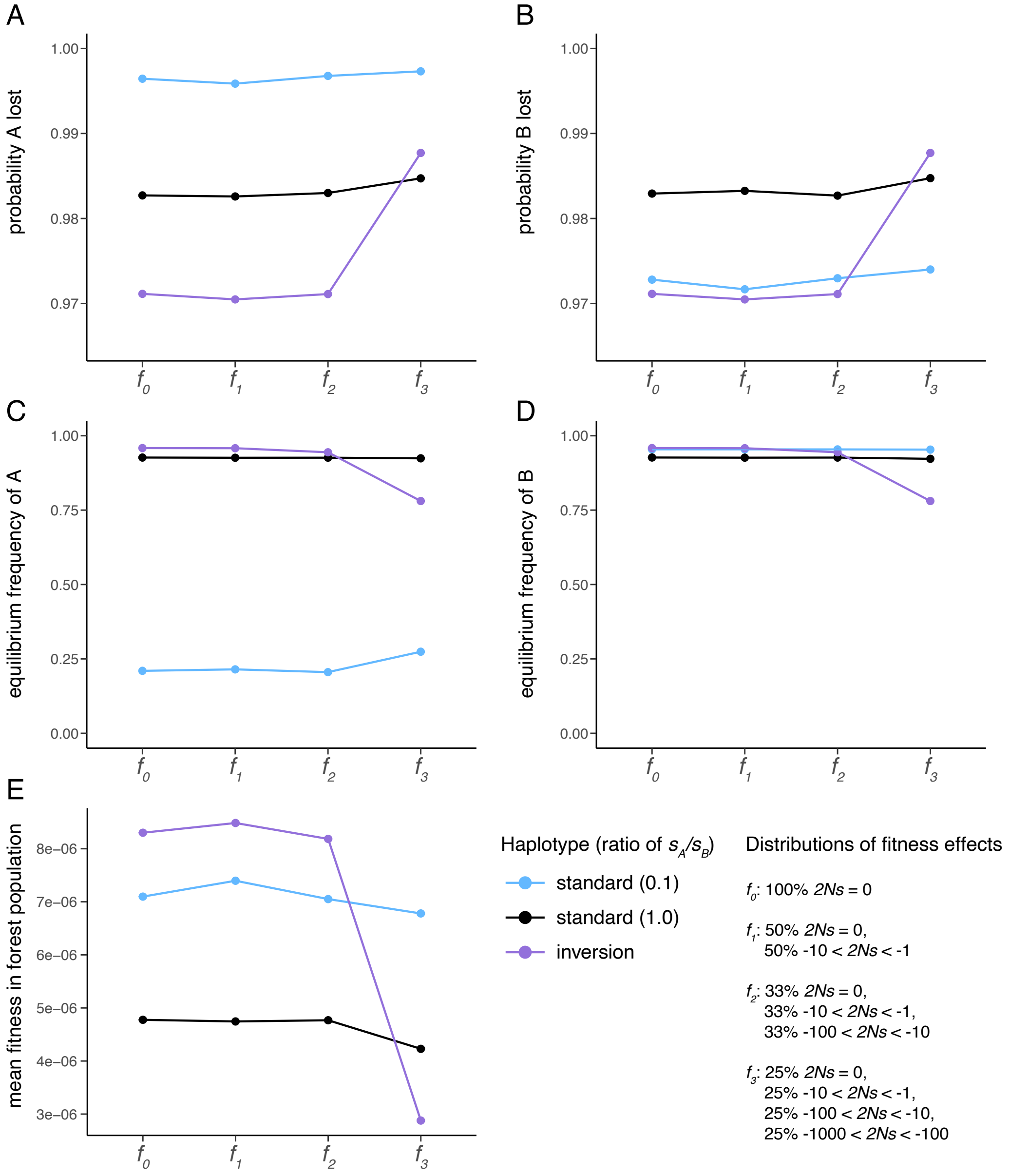

Fig. S16. Simulation results from the two-locus model with deleterious mutational load. A two-locus model was simulated in SLiM under the demographic model, with deleterious mutations added to functional regions according to four different distributions of fitness effects (f_0, f_1, f_2, f_3; see Methods). The two loci (A & B) were located 100-kb apart. On the standard haplotype, the ratio of selection coefficients for A (s_A) & B (s_B) varied with $s_A/s_B = 0.1$ or 1, and $s_A+s_B = 3\text{e-}4$. For the inversion, $s_A+s_B = 3\text{e-}4$. For each scenario, 100,000 simulations were performed. (**A**) Probability that A was lost in the simulation. (**B**) Probability that B was lost. (**C**) Mean frequency of A in forest when A was not lost. (**D**) Mean frequency of B in forest when B was not lost. (**E**) The mean forest fitness across all simulations.

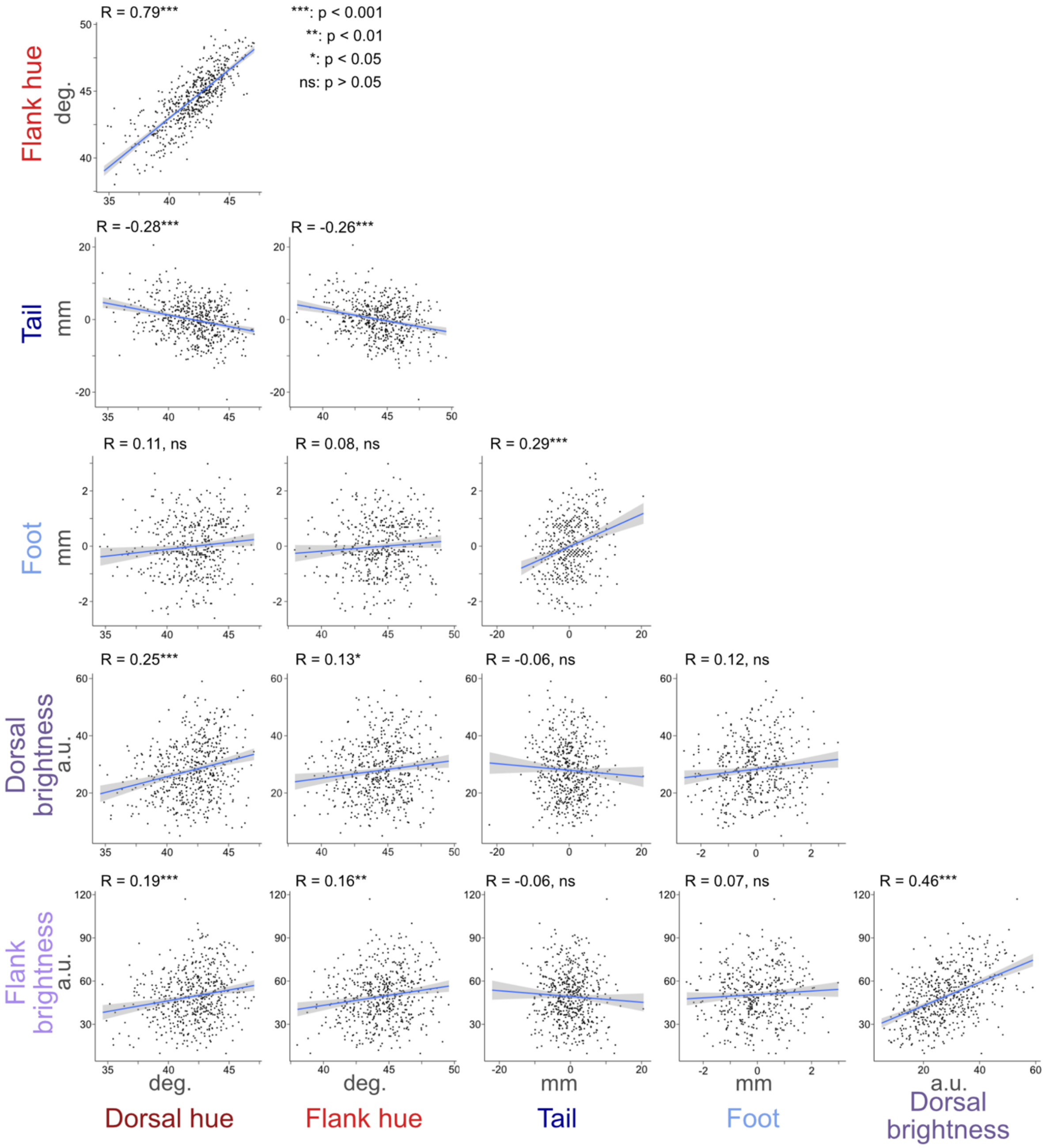

Fig. S17. Correlations among traits in F2 hybrids. Pairwise Pearson's correlations (R = correlation coefficient) among all six traits used for QTL mapping in F2 hybrids. Tail and hindfoot length are shown after taking the residual against body length in the hybrid mice. *N* = 542 (tail), 455 (foot), 541 (pigment). Lines indicate linear smoothing with 95% confidence intervals (gray). Statistics and analysis performed using Pearson's correlations. Abbreviations: deg = degrees, mm = millimeters, a.u. = arbitrary units.

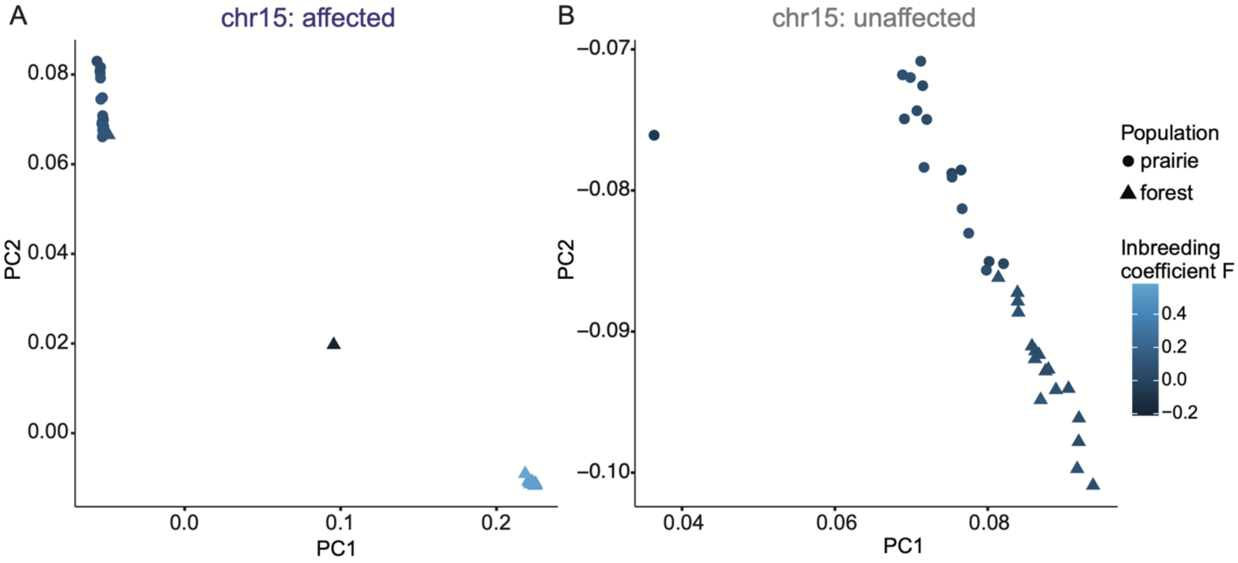

Fig. S18. Genetic principal component analyses and heterozygosity of the affected and unaffected regions of chr15. Genetic principal component analyses (PCA) performed for (A) the affected and (B) the unaffected region of chromosome 15, using wild caught prairie (circles, $n = 15$) and forest (triangles, $n = 15$) mice. Each point represents an individual and is colored by the inbreeding coefficient F. The structure of the affected region portions out individuals into three distinct clusters congruent with heterozygosity and inversion genotypes: homozygous reference individuals (left), a heterozygous individual (middle), and homozygous inversion individuals (right). The unaffected region shows little variation in F, and structures samples primarily by source population (forest vs. prairie).

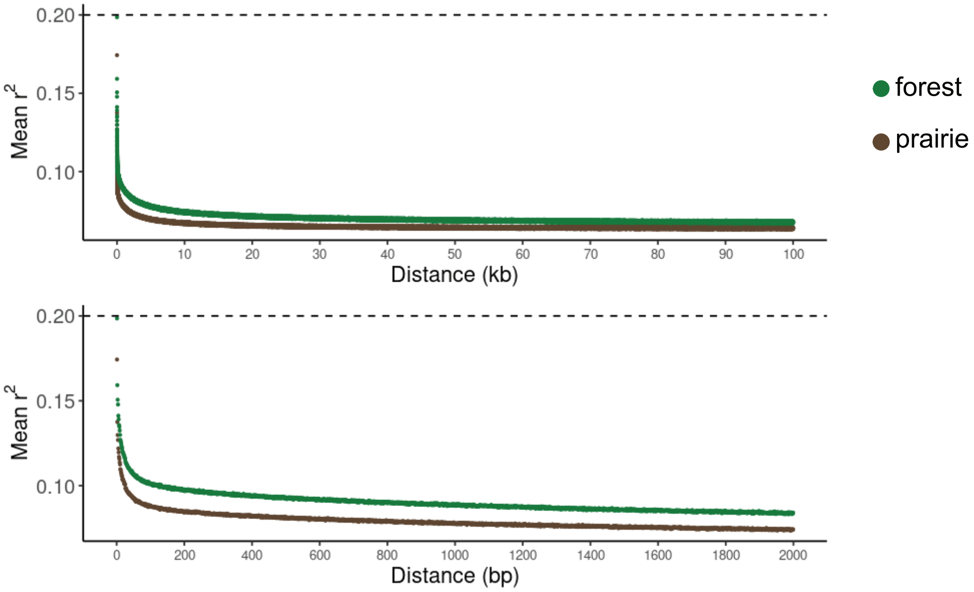

Fig. S19. Linkage disequilibrium decay in wild prairie and forest populations. Mean r^2 between biallelic SNPs, as a function of physical distance in kilobases (top) and basepairs (bottom). Horizontal line at $r^2 = 0.20$ indicates a common threshold below which linkage disequilbrium between variants is often considered to be negligible.

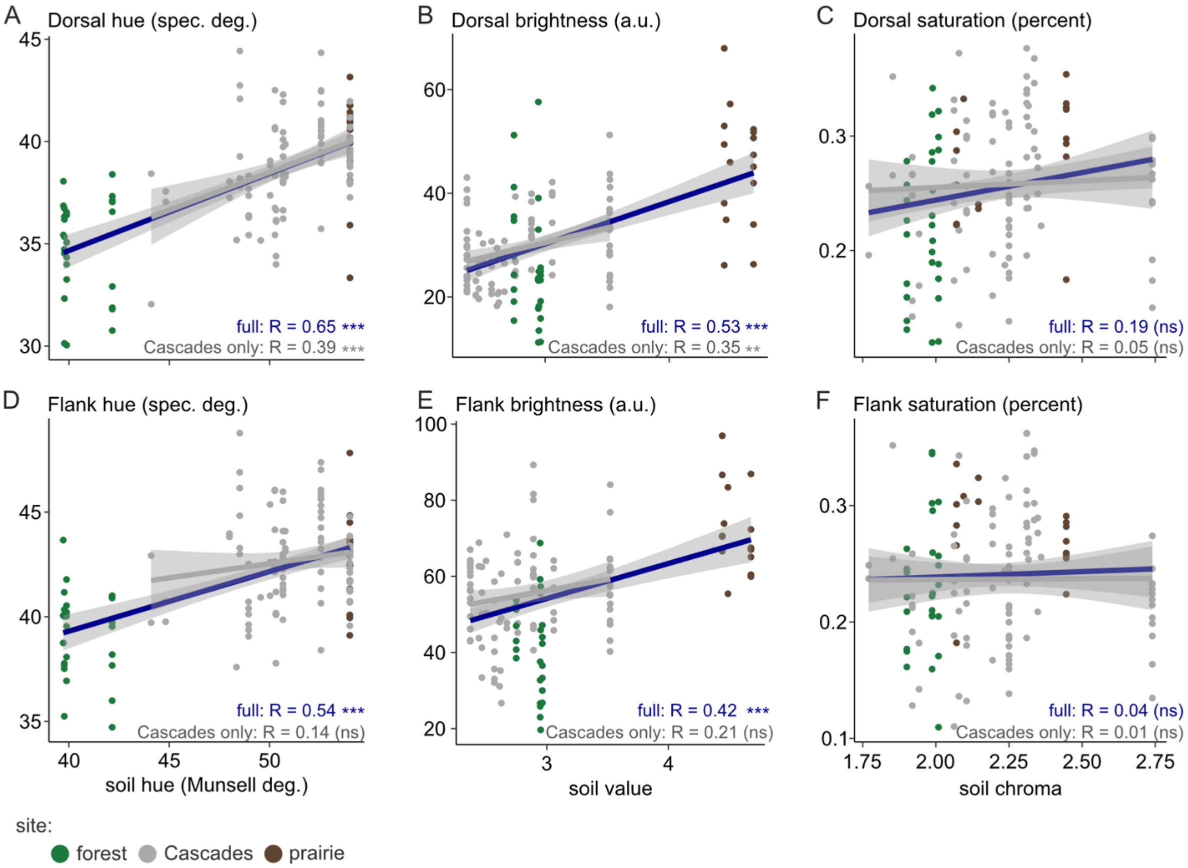

Fig. S20. Correlation between mouse pigmentation and soil characteristics. Pearson's correlations (R = correlation coefficient) between mean soil hue (**A**,**D**), value (**B**,**E**), and chroma (**C**,**F**) and dorsal (**A**-**C**) or flank (**D**-**F**) hue, brightness, and saturation in wild-caught mice. Points are colored by location of capture (green = forest, gray = Cascades, brown = prairie). Correlations are shown both using all data (full = blue, $n = 133$) and using the central Cascades transect only (gray, $n = 90$). Lines indicate linear smoothing with 95% confidence intervals (gray). Statistics and analysis performed using Pearson's correlations. Symbols: ns = $p > 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

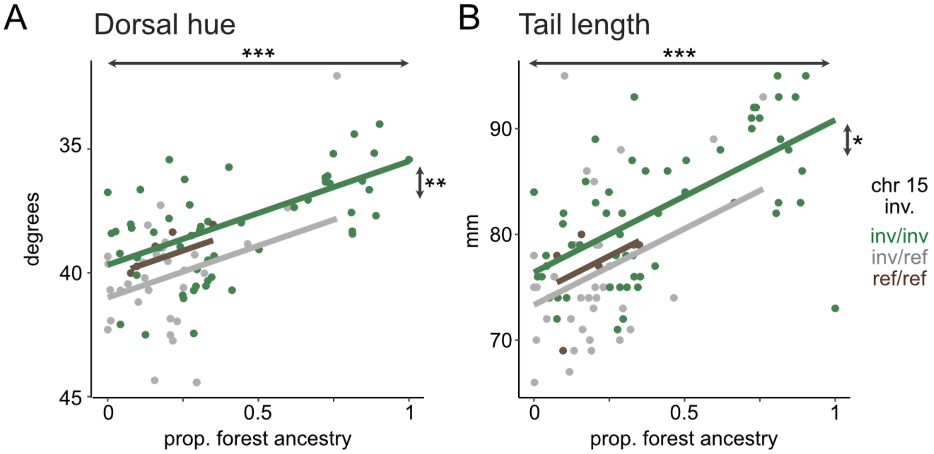

Fig. S21. Association between genotype and phenotype in transect mice. Relationship between estimated forest ancestry proportion (x axis), genotype at the chromosome 15 inversion (green = homozygous for the inverted [forest] allele; gray = heterozygous; brown = homozygous for the reference [prairie] allele), and dorsal hue (**A**) and tail length (**B**) in wild-caught adults from the central Cascades transect. Points represent individual mice (dorsal hue: $n = 90$; tail: $n = 97$), and lines show the results of mixed-effect linear models including capture site as a random effect, inversion genotype as a categorical fixed effect, and proportion forest ancestry genome-wide as a continuous fixed effect. Models including both inversion genotype and genome-wide ancestry were selected over equivalent models with genome-wide ancestry only (tail, inversion + ancestry, model log-likelihood (LL) = -304.4, AIC = 620.9; tail, ancestry only, LL = -307.0, AIC = 622.0; hue, inversion + ancestry, model log-likelihood (LL) = -174.6, AIC = 361.3; hue, ancestry only, LL = -179.0, AIC = 366.0). Symbols: * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

Table S1. Differences in phenotype between forest and prairie mice in the wild and lab. Results of Welch's t-tests comparing forest and prairie phenotypes. Weight: $\log_{10}$-transformed values. s.d. = standard deviation; p (adjusted) = p-value after multiple test correction using the Bonferroni-Holm method; p (initial) = p-value without multiple test correction. Symbols: mm = millimeters, g = grams, a.u. = arbitrary units of reflectance, d = dorsal, f = flank, v = ventral, ns = $p > 0.05$, * = $p < 0.05$, *** = $p < 0.001$ after multiple test correction.

WILD

trait	N (forest)	N (prairie)	forest (mean ± s.d.)	prairie (mean ± s.d.)	t statistic		p (adjusted)	p (initial)
body (mm)	38	32	81.5 ± 7.5	82.9 ± 5.7	0.92	ns	1	0.36
weight (g)	39	30	1.22 ± 0.08	1.21 ± 0.06	-0.3	ns	1	0.77
ear (mm)	33	29	18.6 ± 2.1	18.7 ± 1.3	0.26	ns	1	0.8
tail (mm)	39	32	97.2 ± 7.4	69.9 ± 4.6	-19.01	***	4.63E-27	3.31E-28
hindfoot (mm)	33	29	22.2 ± 1.2	20.2 ± 1.0	-6.95	***	3.37E-08	3.07E-09
d. brightness (a.u.)	16	20	25.3 ± 6.1	51.6 ± 10.4	9.51	***	1.20E-09	9.26E-11
f. brightness (a.u.)	16	20	53.4 ± 5.7	90.4 ± 17.4	8.92	***	4.68E-08	4.68E-09
v. brightness (a.u.)	16	20	113.2 ± 19.3	169.7 ± 24.8	7.69	***	5.46E-08	6.07E-09
d. hue (degrees)	16	20	40.3 ± 1.9	45.7 ± 2.0	8.38	***	1.23E-08	1.03E-09
f. hue (degrees)	16	20	44.7 ± 2	48.9 ± 2.2	5.95	***	7.31E-06	1.04E-06
v. hue (degrees)	16	20	56.7 ± 3.3	64.7 ± 4.4	6.24	***	3.45E-06	4.31E-07
d. saturation (%)	16	20	0.21 ± 0.04	0.23 ± 0.04	2.03	ns	0.25	0.05
f. saturation (%)	16	20	0.20 ± 0.03	0.19 ± 0.03	-0.47	ns	1	0.64
v. saturation (%)	16	20	0.019 ± 0.003	0.023 ± 0.006	2.38	ns	0.14	0.02

LAB

trait	N (forest)	N (prairie)	forest (mean ± s.d.)	prairie (mean ± s.d.)	t statistic		p (adjusted)	p (initial)
body (mm)	20	20	88.2 ± 3.5	84.8 ± 3.1	-3.25	*	0.0148	0.0025
weight (g)	20	20	1.32 ± 0.1	1.3 ± 0.08	-0.72	ns	0.75	0.48
ear (mm)	20	20	18.0 ± 0.6	16.9 ± 1.2	-3.28	*	0.0148	0.0028
tail (mm)	20	20	106.8 ± 6.4	73.3 ± 5.0	-18.42	***	1.17E-18	8.37E-20
hindfoot (mm)	20	20	23.2 ± 0.9	20.8 ± 0.9	-8.32	***	3.92E-09	4.36E-10
d. brightness (a.u.)	31	31	24.1 ± 6.3	43.0 ± 7.0	11.16	***	4.54E-15	3.49E-16
f. brightness (a.u.)	31	31	53.2 ± 9.2	82.0 ± 12.2	10.5	***	9.81E-14	8.17E-15
v. brightness (a.u.)	31	31	134.2 ± 25.7	191.6 ± 26.4	8.68	***	3.71E-11	3.38E-12
d. hue (degrees)	31	31	37.8 ± 2.4	42.7 ± 2.4	8.03	***	4.38E-10	4.38E-11
f. hue (degrees)	31	31	41.9 ± 1.8	45.6 ± 2.5	6.77	***	6.98E-08	9.97E-09
v. hue (degrees)	31	31	58.6 ± 4.3	65.7 ± 3.6	6.97	***	2.62E-08	3.28E-09
d. saturation (%)	31	31	0.23 ± 0.05	0.25 ± 0.03	1.86	ns	0.27	0.07
f. saturation (%)	31	31	0.22 ± 0.03	0.21 ± 0.04	-0.89	ns	0.75	0.38
v. saturation (%)	31	31	0.021 ± 0.004	0.022 ± 0.004	1.68	ns	0.3	0.1

Table S2. QTL effect sizes. Model results using fitqtl for each trait. Abbreviations: chr. = chromosome; LOD = log of the odds score; pos. = position in basepairs (bp); CI = 95% Bayes' credible interval (bounds given in bp, width given in Mb); a = additive effect (estimate ± standard error); d = dominance effect (estimate ± standard error); abs = absolute value; PVE = percent variance explained; * = QTL is transgressive; pgm = cross direction. For tail models, effect sizes were estimated while including all significant loci (left) or in models with each locus alone (right). Note that linear models strongly supported the presence of all five loci ($F_{2,540}$ comparing a full model with all 5 loci vs models with one locus dropped: 10.6 - 49.0). Labels indicate trait of interest, with any additive covariates in parentheses.

Primary models	Trait	**Tail (body)** % variance explained (all significant QTL): 27.3%				
	chr.	6	7	12	14	15
	LOD	4.9	4.7	5.3	5.1	17.4
	pos.	97457639	37729627	17493789	229987	41332882
	CI	(77197532 , 120898287)	(22930804 , 118666602)	(10643762 , 24185533)	(220736 , 41048679)	(406222 , 41452107)
	CI width	43.7	95.7	13.5	40.8	41
Models including all significant QTL // one QTL alone	a	1.17 ± 0.26 // 1.41 ± 0.3	1.28 ± 0.26 // 1.37 ± 0.3	1.6 ± 0.27 // 1.52 ± 0.31	1.59 ± 0.28 // 1.55 ± 0.32	2.71 ± 0.27 // 2.76 ± 0.3
	d	-0.29 ± 0.38 // -0.2 ± 0.43	0.18 ± 0.38 // 0.33 ± 0.44	0.19 ± 0.38 // 0.20 ± 0.44	0.14 ± 0.38 // 0.37 ± 0.44	0.50 ± 0.38 // 0.50 ± 0.41
	abs. d/a	0.25 // 0.14	0.14 // 0.24	0.12 // 0.13	0.09 // 0.24	0.19 // 0.18
	PVE	2.6 // 3.8	3 // 3.6	4.5 // 4.1	4 // 3.9	12.1 // 12.8

Models with additional covariates	Trait	**Tail (body, sex, age)**				
	chr.	6	7	12	14	15
	LOD	5.4	5	6.2	4.3	18.7
	pos.	97457639	37681732	16589347	229987	41332882
	CI	(77183524 , 113132464)	(23423646 , 86976095)	(11903424 , 22687967)	(220736 , 41073607)	(473899 , 41465979)
	CI width	35.9	63.6	10.8	40.9	41
Models including all significant QTL // one QTL alone	a	1.24 ± 0.26 // 1.46 ± 0.3	1.31 ± 0.25 // 1.39 ± 0.29	1.74 ± 0.27 // 1.66 ± 0.31	1.37 ± 0.28 // 1.40 ± 0.33	2.76 ± 0.27 // 2.81 ± 0.29
	d	-0.30 ± 0.37 // -0.23 ± 0.43	0.28 ± 0.37 // 0.43 ± 0.43	0.02 ± 0.37 // -0.06 ± 0.43	0.26 ± 0.37 // 0.46 ± 0.44	0.48 ± 0.37 // 0.47 ± 0.4
	abs. d/a	0.24 // 0.16	0.21 // 0.31	0.01 // 0.04	0.19 // 0.33	0.18 // 0.17
	PVE	2.9 // 4.1	3.2 // 3.8	5.1 // 4.7	2.9 // 3.2	12.5 // 13.2

Primary models	Trait	**Dorsal hue**	**Flank hue**	**Flank brightness**
	chr.	15	15	21*
	LOD	60.1	71.5	4
	pos.	11545604	21979520	35087484
	CI	(406222 , 40539040)	(406222 , 39415731)	(25040451 , 69570542)
	CI width	40.1	39	44.5
	a	-1.96 ± 0.11	-1.93 ± 0.09	3.47 ± 1.07
	d	0.52 ± 0.15	0.58 ± 0.13	-4.32 ± 1.51
	abs. d/a	0.26	0.3	1.25
	PVE	40.0	45.6	3.4

Models with additional covariates	Trait	**Dorsal hue (pgm)**	**Flank hue (pgm)**	**Flank brightness (pgm)**
	chr.	15	15	21*
	LOD	67.2	78.1	4.2
	pos.	11545604	10474970	35087484
	CI	(403562 , 40734548)	(475455 , 40590260)	(27589255 , 69570542)
	CI width	40.3	40.1	42
	a	-1.95 ± 0.1	-1.92 ± 0.09	3.17 ± 1.04
	d	0.54 ± 0.14	0.59 ± 0.12	-4.73 ± 1.47
	abs. d/a	0.28	0.31	1.49
	PVE	39.7	45.3	3.3

Table S3. Genes near inversion breakpoint. Annotated genes within 200 kb of the inversion breakpoint at 40.94 Mb.

Gene	**Start position (bp)**	**Gene type**	**Mouse phenotypes (MGI)**
AC163684.1	40928758	lincRNA	unknown
Hspb3-201	40958074	protein coding	decreased circulating free fatty acids level; increased fasting circulating glucose level
Snx18	40990971	protein coding	hyperactivity; preweaning lethality
AC168056.3	41043419	lincRNA	unknown
AC168056.4	41124903	lincRNA	unknown

Table S4. Genes within inversion. All annotated protein-coding genes within the inversion. # non-synon. mutations = number of non-synonymous mutations fixed between the inversion and reference haplotypes within each gene. Conservation scores (PROVEAN) = predicted effect of amino acid change from PROVEAN for each coding change. Pigment phenotypes = lab mouse (*Mus musculus*, MGI) phenotypes related to pigmentation, associated with any mutations within the gene. Tail-length phenotypes = lab mouse (*Mus musculus*, MGI) phenotypes related to tail or long bone length phenotypes, associated with any mutations within the gene.

Gene	Start position (bp)	# non-synon. mutations	Conservation scores (PROVEAN)	Pigment phenotypes	Tail-length phenotypes
Lrrc14b	97735	0			
Ccdc127	99011	0			
Sdha	111478	0			
Pdcd6	148312	0			
Ahrr	175474	4	N394Y: -2.61 (deleterious) S531P: 0.64 (neutral) D552N: -0.65 (neutral) G599E: -1.90 (neutral)		
Exoc3	278606	1	K413R: -0.97 (neutral)		
Slc9a3	312950	0			
Cep72	410719	0			
Tppp	473369	0			
Zdhhc11	520628	1	N203S: 0.64 (neutral)		
Brd9	552899	0			
Trip13	578982	0		abnormal coat/hair pigmentation	kinked tail;short tail;abnormal tail morphology
Nkd2	678968	0			
Slc12a7	707452	1	D377E: 0.63 (neutral)		
Slc6a19	825512	0		abnormal coat/hair pigmentation	
Slc6a18	852997	2	I168T: -2.85 (deleterious) T233A: -2.73 (deleterious)		
Tert	886003	1	K498R: -1.46 (neutral)		
Clptm1l	919398	0			
Slc6a3	971353	0			decreased length of long bones
Lpcat1	1026840	0			
Mrpl36	1236057	0			
Ndufs6	1239439	0			
Irx4	1303075	0			
8030423J24Rik	3635853	0			

Adamts16	3681221	0			
Ice1	3878541	3	S448L: -1.36 (neutral) R640L: 1.13 (neutral) S1513L: -2.26 (neutral)		
Med10	4659837	0			
Ube2ql1	4742048	0			
Nsun2	4863135	0			decreased length of long bones
Srd5a1	4889482	0			
Papd7	4990329	0			
Adcy2	5537757	0			
1700001L19Rik	5913906	0			
Mtrr	5956981	2	F86L: -0.32 (neutral) E203G: -2.28 (neutral)		
Zfp748	6931249	2	H135Y: -4.68 (deleterious) Q71P: -2.61 (deleterious)		
Rslcan18	7059142	0			
Gm28557	7064524	0			
Zfp85	7080045	0			
Zfp273	7080046	0			
Zfp493	7080047	0			
Zfp708	7080062	0			
Zfp712	7098107	0			
Sema5a	7498427	0			
Tas2r119	8037648	6	I302V: 0.22 (neutral) V273I: 0.66 (neutral) A78T: 0.13 (neutral) S47P: 0.78 (neutral) M13T: -0.31 (neutral) M12T: -0.15 (neutral)		
Fam173b	8583786	0			
Cct5	8607743	0			
Cmbl	8619911	0			increased caudal vertebrae number
Mar6.	8702141	0			
Ropn1l	8750979	0			
Ankrd33b	8837465	0			
Dap	8941921	0			
Ctnnd2	9205936	0			
Dnah5	11653164	2	I2405V: -0.80 (neutral) M1055T: 0.90 (neutral)		
Trio	12088848	0			
Fam105a	12474403	0			
Otulin	12511438	0			

Ank	12550231	0		abnormal skin pigmentation	short femur
Fbxl7	13244480	0			
Mar11.	13742359	0			
Zfp622	14122070	0			
Retreg1	14144136	0			
Myo10	14309808	0		abnormal coat/hair pigmentation;abnormal tail pigmentation;decreased tail pigmentation	kinked tail;decreased caudal vertebrae number
Gm5468	14742936	0			
Basp1	14748147	0			
9230109A22Rik	14997151	0			
Cdh18	16439528	1	F430I: -1.87 (neutral)		
Cdh12	18122603	0			
Cdh6	21784408	2	F7I: 0.35 (neutral) R35K: -0.33 (neutral)		
Drosha	22016681	0			
6030458C11Rik	22136821	0			
Pdzd2	22229287	1	G1645V: -1.62 (neutral)		
Golph3	22620447	0			
1810049J17Rik	22652345	0			
Mtmr12	22698109	0			
Zfr	22801384	0			
Sub1	22981807	0			
Npr3	23065480	0			elongated vertebral body;kinked tail;long tail;abnormal vertebrae morphology;abnormal caudal vertebrae morphology;increased length of long bones
Tars	23556481	0			
Adamts12	23613328	0			
Rxfp3	23903946	0			
Slc45a2	23909990	1	L75F: -0.73 (neutral)	diluted coat color;abnormal coat/hair pigmentation;abnormal pinna hair pigmentation;abnormal eye pigmentation;decreased eye	

				pigmentation;irregular coat pigmentation;abnormal skin pigmentation;decreased skin pigmentation	
Amacr	23943772	0			
C1qtnf3	23957243	0			
Rai14	24191052	0			
Ttc23l	24341347	0			
Rad1	24417203	1	K540*: stop codon		
Brix1	24427085	0			
Dnajc21	24447547	0			short tibia
Agxt2	24497434	0			
Gm21973	24540862	0			
Prlr	24553567	1	S223N: 1.56 (neutral)	abnormal coat/ hair morphology;coarse hair	
Spef2	25081628	3	A361V: -3.20 (deleterious) T886A: 0.05 (neutral) I959M: -1.13 (neutral)		
Il7r	25285657	0			
Capsl	25340036	0			
Cdh9	28951493	0			
Lmbrd2	29570901	0			
Skp2	29623093	0			
Nadk2	29653733	0		abnormal coat/hair pigmentation	
Ranbp3l	29701386	0			
Slc1a3	29973010	0			
Nipbl	30262300	0			decreased length of long bones
2410089E03Rik	30457341	3	A1833T: -0.54 (neutral) A1812V: -0.30 (neutral) V144L: 0.93 (neutral)		abnormal vertebrae morphology;decreased length of long bones
Nup155	30580359	0			
Wdr70	30645638	0			
Gdnf	30921641	0			
Egflam	31348844	0			
Lifr	31535011	1	R764K: 0.12 (neutral)		abnormal long bone morphology
Osmr	31882202	0			
Rictor	31947553	0			

Fyb	32080784	2	A556T: -0.39 (neutral) T555N: -0.19 (neutral)		
C9	32239104	0			
Dab2	32285904	0			
Ptger4	33256282	0			
Ttc33	33279747	0			
Prkaa1	33321611	0			
Rpl37	33385957	0			
Card6	33394923	1	H767P: -3.24 (deleterious)		
C7	33445299	3	Y392F: -0.76 (neutral) G571E: 2.80 (neutral) E804A: 0.72 (neutral)		
Mroh2b	33522989	0			
C6	33622704	2	P844S: 0.13 (neutral) L832S: -0.16 (neutral)		
Plcxd3	33784059	0			
Oxct1	34122325	0			
AW549877	34283658	0			
Fbxo4	34301897	0			
Ghr	34596191	0			abnormal long bone hypertrophic chondrocyte zone;abnormal long bone epiphyseal plate morphology;decreased length of long bones
Zfp131	35074673	0			
Nim1k	35119732	0			
Hmgcs1	35179155	0			
Ccl28	35223996	2	E21G: -1.93 (neutral) A16V: -0.62 (neutral)		
Tmem267	35278620	0			
4833420G17Rik	35290937	0			
Paip1	35316504	0			
Nnt	35364192	0			
AC154550.1	35364240	0			
Fgf10	35877825	0		abnormal hair shaft morphology;abnormal hair follicle morphology;decreased hair follicle number;increased hair follicle apoptosis;small hair follicle bulb	

B430218F22Rik	36286956	0			
Mrps30	36287656	1	G485V: -0.47 (neutral)		
Hcn1	36629720	0			
Gm17509	37410673	0			
Parp8	37591979	0			
Isl1	38300246	0			
Itga1	39484743	0			
AC175538.2	39484786	0			
Pelo	39496241	0			
Itga2	39655386	1	A99T: 1.88 (neutral)		
Mocs2	39757993	1	K34N: -0.37 (neutral)	abnormal hair growth	
Fst	40123123	0			
Ndufs4	40183595	1	T39I: -1.51 (neutral)	sparse hair;premature hair loss	straub tail;tail dragging
Arl15	40432621	0			

Table S5. Groups used to classify habitat type. Habitat Group = grouping used for analysis, after binning categories across age groups. Habitat = habitat names used in the source model. Value = numeric code used in the source model.

Habitat Group	Habitat	Value
alpine	Alpine	20
	Subalpine Parkland	35
aspen	Quaking Aspen	23
cliffs_canyons	Cliffs and Canyons	62
coastal_spruce_cedar_redwood	Coastal Spruce, Cedar or Redwood mature	63
	Coastal Spruce, Cedar or Redwood medium	66
	Coastal Spruce, Cedar or Redwood old-growth	72
	Coastal Spruce, Cedar or Redwood young	73
crops_and_developed	Urban (High Intensity Developed)	32
	Suburban (Moderate Intensity Developed)	56
	Cultivated Crops	57
	Rural Residential (Low Intensity Developed)	61
douglas_fir_western_hemlock	Douglas Fir - Western Hemlock mature	14
	Douglas Fir - Western Hemlock medium	33
	Douglas Fir - Western Hemlock young	59
	Douglas Fir - Western Hemlock old-growth	75
grassland	Montane Grasslands and Dry Meadows	16
	Alkali and Desert Grasslands	26
	Coastal and Valley Grasslands	36
	Pasture or Hay	58
	Exotic Grasslands and Annuals	65
	Columbia Basin Grasslands and Prairie	74
juniper	Western Juniper	39
lava_dunes_playa_burns_beaches	Coastal Dunes and Beaches	24
	Inland Dunes	25
	Lava	29
	Playa and Barren Ash	30
	Rocky Coast	31
	Burns	69
lodgepole_pine	Lodgepole Pine mature	5
	Lodgepole Pine young	45
mixed_conifer	Mixed Conifer (White or Douglas Fir/Pine) mature	3
	Mixed Conifer (White or Douglas Fir/Pine) medium	4
	Mixed Conifer (White or Douglas Fir/Pine) old-growth	21
	Mixed Conifer (White or Douglas Fir/Pine) young	40
mixed_hardwood_conifer	Mixed Hardwood - Conifer mature	12
	Mixed Hardwood - Conifer old-growth	13
	Mixed Hardwood - Conifer medium	18
	Mixed Hardwood - Conifer young	44
oak	Oak	34
	Mixed Oak - Conifer mature	41
	Mixed Oak - Conifer old-growth	42
	Mixed Oak - Conifer young to medium	43
ponderosa	Ponderosa Pine medium	8
	Ponderosa Pine mature	46
	Ponderosa Pine old-growth	47
	Ponderosa Pine young	48
sagebrush	Big Sagebrush fair - good	2
	Low Sagebrush fair - good	9
	Mountain Big Sagebrush fair - good	10
	Big Sagebrush poor	71
	Low Sagebrush poor	76
	Mountain Big Sagebrush poor	77
shrub	Salt Desert Scrub	11
	Early Shrub-Tree	17
	Chaparral	19
	Canyon & Montane Shrubland	22
silver_fir_mtn_hemlock	Silver Fir - Mountain Hemlock medium	6
	Silver Fir - Mountain Hemlock mature	49
	Silver Fir - Mountain Hemlock old-growth	50
	Silver Fir - Mountain Hemlock young	51
siskiyou	Siskiyou Mixed Conifer medium	7
	Siskiyou Mixed Conifer old-growth	52
	Siskiyou Mixed Conifer mature	53
	Siskiyou Mixed Conifer young	54
spruce	Spruce - Subalpine Fir old-growth	1
	Spruce - Subalpine Fir young	55
	Spruce - Subalpine Fir medium to mature	64
water_riparian	Marshes, Bogs and Emergent Wetlands	15
	Open Water (Big Rivers and Reservoirs)	27
	Saltmarsh	28
	Coastal and Valley Riparian	37
	Montane Wetlands	38
	Interior Lowland and Foothill Riparian	67
	Lowland Woody Wetlands and Swamps	68
	Bays and Estuaries	70

Table S6. Taqman assay probes. Custom Taqman SNP genotyping assays used to genotype museum specimens at four SNPs differentiating the inversion.

Taqman assay name	SNP location	Forward primer	Reverse primer	Probe	Reference allele	Inversion allele
March6	chr15_8729600	AGCCCACCAAGCCATACTC	AAGGCCGGCGGTAAGG	TGAAAGCC[A/G]ACAGGTC	A - VIC	G - FAM
Zfp622	chr15_14131681	TCCATGGTTTCATCATCACATTCCA	GCAGGACCCTAGTGAGATTGAG	TTTTCGGA[G/A]AATAAC	G - VIC	A - FAM
Slc45a2	chr15_23939629	TGCAGCAGGAATCCCAAGATG	GCCTGCCCAAGAGTTTGTACA	CATGGTGTGG[C/T]TCCTAA	C - VIC	T - FAM
Skp2	chr15_29639255	CTGTCTGAGTGCTCCAAGCT	ACAATGGGATCCGAAAGTTGCA	CTTCCAG[G/A]CTTAGATTC	G - VIC	A - FAM

References and Notes

1. G. Turesson, The species and the variety as ecological units. *Hereditas* **3**, 100–113 (1922). doi:10.1111/j.1601-5223.1922.tb02727.x

2. D. B. Lowry, Ecotypes and the controversy over stages in the formation of new species. *Biol. J. Linn. Soc.* **106**, 241–257 (2012). doi:10.1111/j.1095-8312.2012.01867.x

3. A. Tigano, V. L. Friesen, Genomics of local adaptation with gene flow. *Mol. Ecol.* **25**, 2144–2164 (2016). doi:10.1111/mec.13606 Medline

4. J. B. S. Haldane, A mathematical theory of natural and artificial selection. (Part VI, Isolation.). *Math. Proc. Camb. Philos. Soc.* **26**, 220–230 (1930). doi:10.1017/S0305004100015450

5. R. Bürger, A. Akerman, The effects of linkage and gene flow on local adaptation: A two-locus continent-island model. *Theor. Popul. Biol.* **80**, 272–288 (2011). doi:10.1016/j.tpb.2011.07.002 Medline

6. S. Yeaman, M. C. Whitlock, The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**, 1897–1911 (2011). doi:10.1111/j.1558-5646.2011.01269.x Medline

7. W. H. Osgood, A revision of the mice of the American genus *Peromyscus*. *N. Am. Fauna* **28**, 1–285 (1909). doi:10.3996/nafa.28.0001

8. W. F. Blair, Ecological factors in speciation of *Peromyscus*. *Evolution* **4**, 253–275 (1950). doi:10.1111/j.1558-5646.1950.tb01396.x

9. L. R. Dice, Ecologic and genetic variability within species of *Peromyscus*. *Am. Nat.* **74**, 212–221 (1940). doi:10.1086/280889

10. S. W. Calhoun, I. F. Greenbaum, K. P. Fuxa, Biochemical and karyotypic variation in *Peromyscus maniculatus* from western North America. *J. Mammal.* **69**, 34–45 (1988). doi:10.2307/1381745

11. D. S. Yang, G. Kenagy, Population delimitation across contrasting evolutionary clines in deer mice (*Peromyscus maniculatus*). *Ecol. Evol.* **1**, 26–36 (2011). doi:10.1002/ece3.3 Medline

12. Materials and methods are available as supplementary materials.

13. R. Villoutreix, D. Ayala, M. Joron, Z. Gompert, J. L. Feder, P. Nosil, Inversion breakpoints and the evolution of supergenes. *Mol. Ecol.* **30**, 2738–2755 (2021). doi:10.1111/mec.15907 Medline

14. P. Johri, K. Riall, H. Becher, L. Excoffier, B. Charlesworth, J. D. Jensen, The impact of purifying and background selection on the inference of population history: Problems and prospects. *Mol. Biol. Evol.* **38**, 2986–3003 (2021). doi:10.1093/molbev/msab050 Medline

15. S. Wright, Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931). doi:10.1093/genetics/16.2.97 Medline

16. G. Hewitt, The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000). doi:10.1038/35016000 Medline

17. M. Kirkpatrick, N. Barton, Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006). doi:10.1534/genetics.105.047985 Medline

18. B. Charlesworth, N. H. Barton, The spread of an inversion with migration and selection. *Genetics* **208**, 377–382 (2018). doi:10.1534/genetics.117.300426 Medline

19. S. N. Vignieri, J. G. Larson, H. E. Hoekstra, The selective advantage of crypsis in mice. *Evolution* **64**, 2153–2158 (2010). doi:10.1111/j.1558-5646.2010.00976.x Medline

20. E. P. Kingsley, K. M. Kozak, S. P. Pfeifer, D. S. Yang, H. E. Hoekstra, The ultimate and proximate mechanisms driving the evolution of long tails in forest deer mice. *Evolution* **71**, 261–273 (2017). doi:10.1111/evo.13150 Medline

21. E. L. Berdan, A. Blanckaert, R. K. Butlin, C. Bank, Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *PLOS Genet.* **17**, e1009411 (2021). doi:10.1371/journal.pgen.1009411 Medline

22. B. Charlesworth, J. D. Jensen, Effects of selection at linked sites on patterns of genetic variability. *Annu. Rev. Ecol. Evol. Syst.* **52**, 177–197 (2021). doi:10.1146/annurev-ecolsys-010621-044528

23. B. E. Horner, Arboreal adaptations of *Peromyscus*, with special reference to use of the tail. *Contrib. Lab. Vertebr. Biol.* **61**, 1–85 (1954).

24. S. T. Mincer, G. A. Russo, Substrate use drives the macroevolution of mammalian tail length diversity. *Proc. R. Soc. B.* **287**, 20192885 (2020). doi:10.1098/rspb.2019.2885 Medline

25. E. R. Hager, H. E. Hoekstra, Tail length evolution in deer mice: Linking morphology, behavior, and function. *Integr. Comp. Biol.* **61**, 385–397 (2021). doi:10.1093/icb/icab030 Medline

26. R. D. H. Barrett, S. Laurent, R. Mallarino, S. P. Pfeifer, C. C. Y. Xu, M. Foll, K. Wakamatsu, J. S. Duke-Cohan, J. D. Jensen, H. E. Hoekstra, Linking a mutation to survival in wild mice. *Science* **363**, 499–504 (2019). doi:10.1126/science.aav3824 Medline

27. R. Faria, K. Johannesson, R. K. Butlin, A. M. Westram, Evolving Inversions. *Trends Ecol. Evol.* **34**, 239–248 (2019). doi:10.1016/j.tree.2018.12.005 Medline

28. M. J. Thompson, C. D. Jiggins, Supergenes and their role in evolution. *Heredity* **113**, 1–8 (2014). doi:10.1038/hdy.2014.20 Medline

29. A. H. Sturtevant, A case of rearrangement of genes in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **7**, 235–237 (1921). doi:10.1073/pnas.7.8.235 Medline

30. Z. Fang, T. Pyhäjärvi, A. L. Weber, R. K. Dawe, J. C. Glaubitz, Jde. J. González, C. Ross-Ibarra, J. Doebley, P. L. Morrell, J. Ross-Ibarra, Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics* **191**, 883–894 (2012). doi:10.1534/genetics.112.138578 Medline

31. K. Huang, R. L. Andrew, G. L. Owens, K. L. Ostevik, L. H. Rieseberg, Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Mol. Ecol.* **29**, 2535–2549 (2020). doi:10.1111/mec.15428 Medline

32. D. B. Lowry, J. H. Willis, A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLOS Biol.* **8**, e1000500 (2010). doi:10.1371/journal.pbio.1000500 Medline

33. C.-R. Lee, B. Wang, J. P. Mojica, T. Mandáková, K. V. S. K. Prasad, J. L. Goicoechea, N. Perera, U. Hellsten, H. N. Hundley, J. Johnson, J. Grimwood, K. Barry, S. Fairclough, J. W. Jenkins, Y. Yu, D. Kudrna, J. Zhang, J. Talag, W. Golser, K. Ghattas, M. E. Schranz, R. Wing, M. A. Lysak, J. Schmutz, D. S. Rokhsar, T. Mitchell-Olds, Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nat. Ecol. Evol.* **1**, 0119 (2017). doi:10.1038/s41559-017-0119

34. M. Joron, L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee, C. R. Haag, A. Whibley, M. Becuwe, S. W. Baxter, L. Ferguson, P. A. Wilkinson, C. Salazar, C. Davidson, R. Clark, M. A. Quail, H. Beasley, R. Glithero, C. Lloyd, S. Sims, M. C. Jones, J. Rogers, C. D. Jiggins, R. H. ffrench-Constant, Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203–206 (2011). doi:10.1038/nature10341 Medline

35. K. Kunte, W. Zhang, A. Tenger-Trolander, D. H. Palmer, A. Martin, R. D. Reed, S. P. Mullen, M. R. Kronforst, *doublesex* is a mimicry supergene. *Nature* **507**, 229–232 (2014). doi:10.1038/nature13112 Medline

36. Z. Yan, S. H. Martin, D. Gotzek, S. V. Arsenault, P. Duchen, Q. Helleu, O. Riba-Grognuz, B. G. Hunt, N. Salamin, D. Shoemaker, K. G. Ross, L. Keller, Evolution of a supergene that regulates a trans-species social polymorphism. *Nat. Ecol. Evol.* **4**, 240–249 (2020). doi:10.1038/s41559-019-1081-1 Medline

37. C. Mérot, E. L. Berdan, H. Cayuela, H. Djambazian, A. L. Ferchaud, M. Laporte, E. Normandeau, J. Ragoussis, M. Wellenreuther, L. Bernatchez, Locally adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly. *Mol. Biol. Evol.* **38**, 3953–3971 (2021). doi:10.1093/molbev/msab143 Medline

38. C. Cheng, B. J. White, C. Kamdem, K. Mockaitis, C. Costantini, M. W. Hahn, N. J. Besansky, Ecological genomics of *Anopheles gambiae* along a latitudinal cline: A population-resequencing approach. *Genetics* **190**, 1417–1432 (2012). doi:10.1534/genetics.111.137794 Medline

39. J. L. Feder, J. B. Roethele, K. Filchak, J. Niedbalski, J. Romero-Severson, Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* **163**, 939–953 (2003). doi:10.1093/genetics/163.3.939 Medline

40. M. Kapun, T. Flatt, The adaptive significance of chromosomal inversion polymorphisms in *Drosophila melanogaster*. *Mol. Ecol.* **28**, 1263–1282 (2019). doi:10.1111/mec.14871 Medline

41. E. L. Koch, H. E. Morales, J. Larsson, A. M. Westram, R. Faria, A. R. Lemmon, E. M. Lemmon, K. Johannesson, R. K. Butlin, Genetic variation for adaptive traits is associated with polymorphic inversions in *Littorina saxatilis*. *Evol. Lett.* **5**, 196–213 (2021). doi:10.1002/evl3.227 Medline

42. D. Lindtke, K. Lucek, V. Soria-Carrasco, R. Villoutreix, T. E. Farkas, R. Riesch, S. R. Dennis, Z. Gompert, P. Nosil, Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect. *Mol. Ecol.* **26**, 6189–6205 (2017). doi:10.1111/mec.14280 Medline

43. T. Dobzhansky, Adaptive changes induced by natural selection in wild populations of *Drosophila*. *Evolution* **1**, 1–16 (1947). doi:10.1111/j.1558-5646.1947.tb02709.x

44. A. Brelsford, J. Purcell, A. Avril, P. Tran Van, J. Zhang, T. Brütsch, L. Sundström, H. Helanterä, M. Chapuisat, An ancient and eroded social supergene is widespread across *Formica* ants. *Curr. Biol.* **30**, 304–311.e4 (2020). doi:10.1016/j.cub.2019.11.032 Medline

45. D. Ayala, R. F. Guerrero, M. Kirkpatrick, Reproductive isolation and local adaptation quantified for a chromosome inversion in a malaria mosquito. *Evolution* **67**, 946–958 (2013). doi:10.1111/j.1558-5646.2012.01836.x Medline

46. F. C. Jones, M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, E. S. Lander, F. Di Palma, K. Lindblad-Toh, D. M. Kingsley, The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012). doi:10.1038/nature10944 Medline

47. M. Matschiner, J. M. I. Barth, O. K. Tørresen, B. Star, H. T. Baalsrud, M. S. O. Brieuc, C. Pampoulie, I. Bradbury, K. S. Jakobsen, S. Jentoft, Supergene origin and maintenance in Atlantic cod. *Nat. Ecol. Evol.* **6**, 469–481 (2022). doi:10.1038/s41559-022-01661-x Medline

48. C. Küpper, M. Stocks, J. E. Risse, N. Dos Remedios, L. L. Farrell, S. B. McRae, T. C. Morgan, N. Karlionova, P. Pinchuk, Y. I. Verkuil, A. S. Kitaysky, J. C. Wingfield, T. Piersma, K. Zeng, J. Slate, M. Blaxter, D. B. Lank, T. Burke, A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **48**, 79–83 (2016). doi:10.1038/ng.3443 Medline

49. H. B. Thorneycroft, Chromosomal polymorphism in the white-throated sparrow, *Zonotrichia albicollis* (Gmelin). *Science* **154**, 1571–1572 (1966). doi:10.1126/science.154.3756.1571 Medline

50. S. Lamichhaney, G. Fan, F. Widemo, U. Gunnarsson, D. S. Thalmann, M. P. Hoeppner, S. Kerje, U. Gustafson, C. Shi, H. Zhang, W. Chen, X. Liang, L. Huang, J. Wang, E. Liang, Q. Wu, S. M. Y. Lee, X. Xu, J. Höglund, X. Liu, L. Andersson, Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* **48**, 84–88 (2016). doi:10.1038/ng.3430 Medline

51. E. R. Funk, N. A. Mason, S. Pálsson, T. Albrecht, J. A. Johnson, S. A. Taylor, A supergene underlies linked variation in color and morphology in a Holarctic songbird. *Nat. Commun.* **12**, 6833 (2021). doi:10.1038/s41467-021-27173-z Medline

52. I. Sanchez-Donoso, S. Ravagni, J. D. Rodríguez-Teijeiro, M. J. Christmas, Y. Huang, A. Maldonado-Linares, M. Puigcerver, I. Jiménez-Blasco, P. Andrade, D. Gonçalves, G. Friis, I. Roig, M. T. Webster, J. A. Leonard, C. Vilà, Massive genome inversion drives

coexistence of divergent morphs in common quails. *Curr. Biol.* **32**, 462–469.e6 (2022). doi:10.1016/j.cub.2021.11.019 Medline

53. G. Dobigny, J. Britton-Davidian, T. J. Robinson, Chromosomal polymorphism in mammals: An evolutionary perspective. *Biol. Rev.* **92**, 1–21 (2017). doi:10.1111/brv.12213 Medline
54. H. Stefansson, A. Helgason, G. Thorleifsson, V. Steinthorsdottir, G. Masson, J. Barnard, A. Baker, A. Jonasdottir, A. Ingason, V. G. Gudnadottir, N. Desnica, A. Hicks, A. Gylfason, D. F. Gudbjartsson, G. M. Jonsdottir, J. Sainz, K. Agnarsson, B. Birgisdottir, S. Ghosh, A. Olafsdottir, J. B. Cazier, K. Kristjansson, M. L. Frigge, T. E. Thorgeirsson, J. R. Gulcher, A. Kong, K. Stefansson, A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129–137 (2005). doi:10.1038/ng1508 Medline
55. E. R. Hager, O. S. Harringmeyer, oharring/chr15_inversion: v1.0.0, version 1.0.0, Zenodo (2022); https://doi.org/10.5281/zenodo.6354918.
56. K. G. Rapp, HAN-rotation, a new system for rigorous outbreeding. *Z. Versuchstierkd.* **14**, 133–142 (1972). Medline
57. J. A. Endler, On the measurement and classification of colour in studies of animal colour patterns. *Biol. J. Linn. Soc.* **41**, 315–352 (1990). doi:10.1111/j.1095-8312.1990.tb00839.x
58. R. Montgomerie, CLR, version 1.05 (2008).
59. R Core Team, R: A Language and Environment for Statistical Computing (2019).
60. K. W. Broman, H. Wu, S. Sen, G. A. Churchill, R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003). doi:10.1093/bioinformatics/btg112 Medline
61. B. K. Peterson, J. N. Weber, E. H. Kay, H. S. Fisher, H. E. Hoekstra, Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS ONE* **7**, e37135 (2012). doi:10.1371/journal.pone.0037135 Medline
62. P. Andolfatto, D. Davison, D. Erezyilmaz, T. T. Hu, J. Mast, T. Sunayama-Morita, D. L. Stern, Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res.* **21**, 610–617 (2011). doi:10.1101/gr.115402.110 Medline
63. K. W. Broman, S. Sen, *A Guide to QTL Mapping with R/qtl* (Springer, 2009).
64. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014). doi:10.1186/s12859-014-0356-4 Medline
65. J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009). doi:10.1126/science.1162986 Medline

66. S. Koren, B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, A. M. Phillippy, Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017). doi:10.1101/gr.215087.116 Medline

67. G. Marçais, A. L. Delcher, A. M. Phillippy, R. Coston, S. L. Salzberg, A. Zimin, MUMmer4: A fast and versatile genome alignment system. *PLOS Comput. Biol.* **14**, e1005944 (2018). doi:10.1371/journal.pcbi.1005944 Medline

68. F. J. Sedlazeck, P. Rescheneder, M. Smolka, H. Fang, M. Nattestad, A. von Haeseler, M. C. Schatz, Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15**, 461–468 (2018). doi:10.1038/s41592-018-0001-7 Medline

69. B. M. Smalec, T. N. Heider, B. L. Flynn, R. J. O'Neill, A centromere satellite concomitant with extensive karyotypic diversity across the *Peromyscus* genus defies predictions of molecular drive. *Chromosome Res.* **27**, 237–252 (2019). doi:10.1007/s10577-019-09605-1 Medline

70. B. Pfeifer, U. Wittelsbürger, S. E. Ramos-Onsins, M. J. Lercher, PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31**, 1929–1936 (2014). doi:10.1093/molbev/msu136 Medline

71. Y. Choi, A. P. Chan, PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747 (2015). doi:10.1093/bioinformatics/btv195 Medline

72. B. S. Arbogast, G. J. Kenagy, Comparative phylogeography as an integrative approach to historical biogeography. *J. Biogeogr.* **28**, 819–825 (2001). doi:10.1046/j.1365-2699.2001.00594.x

73. QGIS Geographic Information System, version 3.4 (Open Source Geospatial Foundation Project, 2019); http://qgis.osgeo.org.

74. Soil Survey Staff, Natural Resources Conservation Service, "U.S. General Soil Map (STATSGO2)" (US Department of Agriculture, 2019); https://sdmdataaccess.sc.egov.usda.gov.

75. Soil Survey Staff, Natural Resources Conservation Service, "Official Soil Series Descriptions" (US Department of Agriculture, 2020); https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/class/data/?cid=nrcs142p2_053585.

76. J. Kagan, K. Zaret, J. Bernert, E. Henderson, "Oregon Habitat Map - 2018" (Oregon Biodiversity Information Center, 2018); https://spatialdata.oregonexplorer.info/geoportal/details;id=4f271c43605a48f3b1edf89f35f0db29.

77. D. Kahle, H. Wickham, ggmap: Spatial Visualization with ggplot2. *R J.* **5**, 144–161 (2013). doi:10.32614/RJ-2013-014

78. L. Skotte, T. S. Korneliussen, A. Albrechtsen, Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013). doi:10.1534/genetics.113.154138 Medline

79. N. M. Kopelman, J. Mayzel, M. Jakobsson, N. A. Rosenberg, I. Mayrose, CLUMPAK: A program for identifying clustering modes and packaging population structure inferences across *K*. *Mol. Ecol. Resour.* **15**, 1179–1191 (2015). doi:10.1111/1755-0998.12387 Medline

80. G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005). doi:10.1111/j.1365-294X.2005.02553.x Medline

81. R. J. Hijmans, geosphere: Spherical trigonometry (2017); https://cran.r-project.org/package=geosphere.

82. E. P. Derryberry, G. E. Derryberry, J. M. Maley, R. T. Brumfield, HZAR: Hybrid zone analysis using an R software package. *Mol. Ecol. Resour.* **14**, 652–663 (2014). doi:10.1111/1755-0998.12209 Medline

83. J. M. Szymura, N. H. Barton, Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution* **40**, 1141–1159 (1986). doi:10.1111/j.1558-5646.1986.tb05740.x Medline

84. M. Plummer, N. Best, K. Cowles, K. Vines, CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* **6**, 7–11 (2006).

85. B. L. Phillips, S. J. E. Baird, C. Moritz, When vicars meet: A narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. *Evolution* **58**, 1536–1548 (2004). doi:10.1111/j.0014-3820.2004.tb01734.x Medline

86. E. M. Ortiz, vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis, version 2.0, Zenodo (2019); https://doi.org/10.5281/zenodo.2540861.

87. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). doi:10.1093/bioinformatics/btu033 Medline

88. A. D. Leaché, B. L. Banbury, J. Felsenstein, A. N. M. de Oca, A. Stamatakis, Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. *Syst. Biol.* **64**, 1032–1047 (2015). doi:10.1093/sysbio/syv053 Medline

89. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019). doi:10.1093/nar/gkz239 Medline

90. A. Uchimura, M. Higuchi, Y. Minakuchi, M. Ohno, A. Toyoda, A. Fujiyama, I. Miura, S. Wakana, J. Nishino, T. Yagi, Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Res.* **25**, 1125–1134 (2015). doi:10.1101/gr.186148.114 Medline

91. L. Excoffier, I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, M. Foll, Robust demographic inference from genomic and SNP data. *PLOS Genet.* **9**, e1003905 (2013). doi:10.1371/journal.pgen.1003905 Medline

92. B. C. Haller, P. W. Messer, SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Mol. Biol. Evol.* **36**, 632–637 (2019). doi:10.1093/molbev/msy228 Medline

93. C. Zhang, S. S. Dong, J. Y. Xu, W. M. He, T. L. Yang, PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788 (2019). doi:10.1093/bioinformatics/bty875 Medline

94. K. Csilléry, O. François, M. G. B. Blum, abc: An R package for approximate Bayesian computation (ABC). *Methods Ecol. Evol.* **3**, 475–479 (2012). doi:10.1111/j.2041-210X.2011.00179.x

95. P. Johri, B. Charlesworth, J. D. Jensen, Toward an evolutionarily appropriate null model: Jointly inferring demography and purifying selection. *Genetics* **215**, 173–192 (2020). doi:10.1534/genetics.119.303002 Medline