TITLE

The genomic imprint of chromosomal inversions and demographic history in island populations of deer mice

AUTHORS AND AFFILIATIONS

Emma K. Howell[1*], Felix Baier[2,3], Hopi E. Hoekstra[2], Bret A. Payseur[1*]

1. Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI, USA
2. Department of Molecular and Cellular Biology; Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA
3. Present address: Max Planck Institute for Brain Research, Frankfurt, Germany

*correspondence: ekhowell@wisc.edu; bret.payseur@wisc.edu

KEYWORDS

Peromyscus maniculatus, inversions, demographic inference, island evolution, balancing selection, whole genome sequencing

ABSTRACT

Populations that colonize islands experience novel selective pressures, fluctuations in size, and changes to their connectivity. Owing to their unique geographic setting, islands can function as natural laboratories in which to examine the interactions between demographic history and natural selection replicated across isolated populations. We used whole genome sequences of wild-caught deer mice (*Peromyscus maniculatus*) from two islands (Saturna and Pender) and one mainland location (Maple Ridge) in the Gulf Islands region of coastal British Columbia to investigate two primary determinants of genome-wide diversity: chromosomal inversions and non-equilibrium demographic history. We found that segregating inversions produce characteristic, large-scale distortions in allele frequencies and linkage disequilibrium that make it possible to identify and characterize them from short-read sequence data. Patterns of variation within and between arrangements indicate that six inversion polymorphisms have been maintained by a shared history of balancing selection in both island and mainland populations. Whereas the estimated timing of contemporary population splits is consistent with the isolation of island populations following the Last Glacial Maximum, ancestral island and mainland lineages are inferred to have diverged much earlier. These aspects of demographic history suggest that shared inversions existed long ago in a common ancestor or spread via limited gene flow between ancestral island and mainland lineages. Our results raise the possibility that inversions segregating among Gulf Islands populations are on similar evolutionary trajectories, providing a contrast to previous findings in mainland *P. maniculatus* and contributing to the emerging portrait of inversion evolution in this species.

INTRODUCTION

Island colonization alters the evolutionary trajectory of a population in multiple ways. Ecological differences between island and mainland environments in variables such as species composition, habitat diversity, resource availability, and climate impart novel selective challenges on colonizing populations (Losos and Ricklefs 2009). Coincident with these ecological shifts, founder effect bottlenecks amplify the effects of genetic drift and reductions in gene flow driven by island isolation increase differentiation between populations (Losos and Ricklefs 2009). Together, these unique selective and demographic changes that accompany evolution on islands shape the fates of standing variants and new mutations alike.

One source of genetic variation that may be particularly sensitive to these features of island evolution is structural changes to the genome. Compared to single nucleotide mutations, structural variants can be orders of magnitude larger in their genomic scope and change the collinearity of chromosome copies. Chromosomal inversions, which occur when a DNA segment is reinserted in the reverse orientation, can disrupt genes that stretch across breakpoints and alter the proximity of genetic elements within and outside of breakpoints, achieving substantial shifts in genome function and architecture in a single mutational step (reviewed in Berdan et al. 2023). Since Alfred Sturtevant's pioneering discovery of inversions in *Drosophila* (Sturtevant 1921), this class of structural variation has been the focus of intensive evolutionary study aimed at understanding the origin, spread, and maintenance of inversions in populations (Hoffmann and Rieseberg 2008).

Central to the evolutionary significance of inversions is their effect on recombination. Crossing over that occurs within inversion breakpoints can produce unbalanced gametes in individuals that carry both the inverted and standard orientation (McClintock 1931). This underdominant behavior in heterokaryotypes, in addition to potentially deleterious positional and breakpoint effects, suggests that inversions may be more likely to spread in populations experiencing increased genetic drift (Lande 1979; Hedrick 1981; Bush et al. 1977), such as those inhabiting islands. Indeed, accelerated rates of inversion-driven karyotypic evolution in gibbons (Carbone et al. 2014) and annual sunflowers (Burke et al. 2004) have been attributed to dramatic fluctuations in population size. At the same time, inversions can promote adaptive evolution through their suppression of recombination. By preserving associations between alleles with epistatic fitness effects, inversions can maintain co-adapted variation (Dobzhansky 1948; Dobzhansky 1949; Dobzhansky 1950). Even in the absence of such epistatic interactions, inversions that capture locally adaptive haplotypes can spread when there is gene flow between populations experiencing divergent selective regimes (Kirkpatrick and Barton 2006). Empirical evidence from diverse taxa suggests that inversions are an important source of adaptive divergence between populations (reviewed in Wellenreuther and Bernatchez 2018).

Peromyscus provides a powerful system for examining chromosomal evolution on islands. As one of the most abundant, widespread rodents in North America (Osgood 1909; Bedford and Hoekstra 2015), the range of *Peromyscus* extends to a wide array of coastal islands that line the continent (Redfield 1976; Gill 1980; Adler and Tamarin 1983; Adler et al. 1986; Lucid and Cook 2004; Álvarez-Castañeda et al. 2010; Nolfo-Clements et al. 2017). Although all species share a 2n = 48 karyotype, there is substantial interspecific and intraspecific variation among chromosomes driven by pericentric inversions and heterochromatic additions (Ohno et al 1966; Robbins and Baker 1981; Stangl and Baker 1984; Greenbaum et al. 1994). In the Pacific Northwest, this karyotypic variation has helped to clarify taxonomic relationships between the deer mouse, *Peromyscus maniculatus*, and Keen's mouse, *P. keeni*, where cytogenetic analyses have revealed differences both within and between species in the fundamental number of chromosome arms (Gunn and Greenbaum 1986; Allard and Greenbaum 1988; Gunn 1987; Hogan et al. 1993). Our understanding of such variability is being refined and expanded in the genomics era. Using a combination of long-read sequencing and population genomic data, Harringmeyer and Hoekstra (2022) provide a detailed characterization of 21 large-scale inversions in *P. maniculatus*, including evidence for substantial geographic variability among mainland populations in both the presence and frequencies of inversion polymorphisms. This genomic profile of inversions provides a useful framework for investigating how this important source of variation is partitioned among island and mainland populations.

Deer mice inhabiting the Gulf Islands of British Columbia offer compelling subjects for studying the interaction between demographic history and inversion evolution in island populations. Pleistocene-era glacial activity played an important role in shaping the contemporary distributions of species occupying higher latitudes (Hewitt 2000; Lessa et al. 2003). Glacial cover of the Gulf Islands during this period (Clague 1980; Clague and James 2002; Hutchinson et al. 2004; Fedje et al. 2009; James et al. 2009) suggests that island colonization could have been accompanied by population size fluctuations as mice dispersed from unglaciated regions or glacial refugia. For low vagility species like *P. maniculatus*, subsequent isolation of the islands from each other and from nearby mainland areas could impose formidable barriers to gene flow (Redfield 1976) that may enable inversions to evolve independently among populations.

Using whole genome sequences from two island populations and one mainland population, we show that segregating inversions have dramatic effects on patterns of variation that span a large portion of the genome. We leverage these genomic signatures to detect polymorphic inversions in short-read population genomic data and to draw inferences about their evolutionary history. By reconstructing the joint demographic history of island and mainland populations, we provide context for our evolutionary interpretations. Together, our genomic characterization of the Gulf Islands populations enriches and expands our current understanding of inversion evolution and population history in *P. maniculatus*.

RESULTS

Population sampling and structure

The Gulf Islands are situated at the southern end of the Strait of Georgia between Vancouver Island (to the west) and mainland British Columbia (to the east). Our sampling of *Peromyscus maniculatus* spans two of the southern Gulf Islands, Saturna (31 km^2) and Pender (34 km^2), and one nearby mainland site at Maple Ridge (Figure 1). Subsequent genomic analyses are based on the subset of unrelated individuals identified among mice sampled from each location (n=20 for Saturna Island, n=10 for Pender Island, and n=17 for mainland Maple Ridge; see Materials and Methods).

To characterize genetic relationships within and among island and mainland sites, we conducted principal component analysis (PCA) and global ancestry inference using genome-wide, unlinked single nucleotide polymorphisms (SNPs). PCA revealed three distinct clusters along the first and second principal component axes that separated mice according to sampling location (Figure 2A). Within this combined sample, cross-validation analyses performed with ADMIXTURE (Alexander et al. 2009) recovered three ancestral components (k=3) that almost entirely delineate individuals by sampling location (Figure 2E-F). In contrast, cross-validation analyses performed separately for each site supported a single ancestral component (k=1) (Figure 2B-D). Together, these findings provide evidence for structure between, but not within, each of the sampled locations.

Inversions drive large-scale distortions in allele frequencies and linkage disequilibrium

The non-collinearity induced by inversions can cause patterns of variation spanning polymorphic loci to depart substantially from genome-wide expectations. Due to the suppression of recombination in heterokaryotypes, variation accumulated on inverted or standard haplotypes may remain private to either class, driving divergence between arrangements in a manner analogous to local population subdivision. In Gulf Islands *P. maniculatus*, we found that this effect is sufficient to distort the site frequency spectrum (SFS) at the genome-wide level. Within each population, the folded SFS constructed from genome-wide variant calls exhibits subtle elevations in SNP counts at intermediate minor allele frequencies (Figure 3A-C; Supplementary Figure S1). Partitioning the SFS by chromosome, we found that these genome-wide irregularities are driven by extreme distortions in the chromosome-wide SFS for five autosomes: 7, 14, 15, 21, and 22 (Figure 3A-C; Supplementary Figure S1), all of which are known to harbor polymorphic inversions within mainland populations of *P. maniculatus* (Hager et al. 2022; Harringmeyer and Hoekstra 2022). In each case, the chromosome-wide SFS exhibits an unusually high density of

144 SNPs at a specific minor allele frequency (MAF) bin, consistent with an excess of fixed
145 differences between inverted and standard arrangements.

146 These same chromosomes also exhibit unusual patterns of linkage disequilibrium (LD). We
147 found that pairwise r^2 across these five chromosomes is elevated over intervals as large as 40
148 Mbp (Figure 3D; Supplementary Figure S1). Such distortions in LD decay could result from the
149 mis-orientation of sampled sequences during their alignment to the reference assembly: when
150 short reads derived from within an inversion are mapped to an inversion-free reference, the re-
151 assignment of genomic coordinates causes nearby SNPs to appear farther apart (and vice versa).
152 In addition to this artificial elevation in LD, the observed patterns of LD decay may also reflect
153 reductions in recombination across regions harboring polymorphic inversions.

154 SNPs that contribute to the aberrant MAF bins in each chromosome-wide SFS localize to
155 discrete genomic intervals that span the breakpoints identified by Harringmeyer and Hoekstra
156 (2022) for inversions inv7.2, inv14.0, inv15.0, inv21.0, and inv22.0 (Figure 4A-E; Supplementary
157 Figures S2-S3). These MAF-partitioned SNPs exhibit extreme genotype continuity across the
158 entirety of each interval, such that each mouse almost exclusively carries heterozygous or
159 homozygous genotypes at this subset of variants (Figure 4A-E; Supplementary Figures S2-S3).
160 Together, these findings suggest that the observed elevations in LD and distortions in the SFS
161 are driven by segregating inversion polymorphisms. We used this diagnostic pattern to screen
162 the Gulf Islands samples for additional segregating inversions by partitioning autosomal
163 genotypes in each population by all possible minor allele frequencies. This approach revealed
164 that a sixth locus on chromosome 3, inv3.0 (Harringmeyer and Hoekstra 2022), is also
165 polymorphic (Figure 4F; Supplementary Figures S2-S3). The shorter span of this inversion is
166 likely insufficient to drive the chromosome-wide distortions in the SFS and LD decay that were
167 observed for the loci on chromosomes 7, 14, 15, 21, and 22. In total, our inversion screen
168 indicates that the same six inversion polymorphisms are segregating in all three Gulf Islands
169 populations.

170 Local PCA conducted using SNPs spanning the breakpoints of inv3.0, inv7.2, inv14.0, inv15.0,
171 inv21.0, and inv22.0 further support the presence of segregating inversions. Within each locus,
172 PCA performed on the combined set of Gulf Islands samples yielded three distinct clusters along
173 the first principal component axis (Figure 5). Whereas genome-wide SNPs group individuals by
174 sampling location (Figure 5A), the clustering pattern observed at these loci instead reflects the
175 three possible inversion genotypes (homozygous inversion, heterozygous, or homozygous
176 standard; Figure 5B-G). Although Harringmeyer and Hoekstra (2022) identified a total of 21
177 large-scale inversion polymorphisms in their broad geographic sampling of mainland *P.*
178 *maniculatus*, we find little evidence that additional polymorphisms beyond the six described
179 above are segregating in the Gulf Islands populations. With the possible exception of inv9.0 in
180 the mainland Maple Ridge population, PCA conducted using SNPs within the 21 intervals

defined by Harringmeyer and Hoekstra (2022) failed to produce clustering patterns consistent with distinct inversion genotypes (Supplementary Figure S4).

Polarizing inversion polymorphisms

To determine the orientation of the homozygous genotypic classes relative to the *P. maniculatus* reference assembly (HU_Pman_2.1.3), we inspected each mouse's genotype proportions for SNPs segregating at the MAF of the rearrangement. Because these SNPs represent fixed differences between arrangements, we reasoned that mice with an enrichment of homozygous reference (REF/REF) genotypes carry the same orientation as the reference assembly while those that are enriched for homozygous alternate (ALT/ALT) genotypes carry the opposite orientation. For inv3.0, inv7.2, inv14.0, inv15.0, and inv22.0, the homozygous classes are clearly distinguished by their proportion of ALT/ALT genotypes (Supplementary Figure S5) and these assignments are visually apparent among MAF-partitioned genotypes (Figure 4A-C and 4E-F). In contrast, neither homozygous class shows a consistent enrichment of ALT/ALT genotypes across all three populations at the inv21.0 locus (Supplementary Figure S5). Instead, one of the homozygous classes (hereby assumed to carry the alternate arrangement) shows an enrichment of ALT alleles at MAFs approaching that of the rearrangement (Supplementary Figures S6-S8) in each population. This pattern might indicate that although private variants have achieved high frequencies within either arrangement subpopulation, the inversion is sufficiently young that few of these private differences have fixed.

To determine whether the arrangements in the *P. maniculatus* reference genome reflect the standard or inverted haplotype, we inspected the orientation of the six segregating inversions in genomes from four additional *Peromyscus* species. We performed pairwise whole genome alignments between the *P. maniculatus* reference genome (HU_Pman_2.1.3) and reference genomes of *P. polionotus* (HU_Ppol_1.3.3), *P. leucopus* (UCI_PerLeu_2.1), *P. eremicus* (PerEre_H2_v1), and *P. californicus* (ASM782708v3) to identify structural rearrangements between species (see Materials and Methods). Comparisons between *P. maniculatus* and these other species suggest an abundance of large-scale inversions, duplications, and translocations (Supplementary Figures S9-S12). In some cases, these inter-species rearrangements span over 50 Mbp (Supplementary Figure S10). Using established phylogenetic relationships among these species (Bradley et al. 2007; Platt et al. 2015), we assessed collinearity across inversion intervals to determine whether the *P. maniculatus* reference genome carries the ancestral (i.e., standard) or derived (i.e., inverted) orientation at each of the six inversions segregating among the Gulf Islands populations (see Materials and Methods). Our findings suggest that the *P. maniculatus* reference genome carries the derived (i.e., inverted) orientation at inv14.0 and the ancestral (i.e., standard) orientation at inv3.0, inv7.2, inv15.0, inv21.0, and inv22.0 (Supplementary Figures S9-S12). We also uncovered evidence that certain segregating inversions are shared across *Peromyscus* species. Collinearity between the *P. maniculatus* and *P. polionotus* genomes

218 along the majority of the inv14.0 interval suggests that this rearrangement is shared between
219 these sister species (Supplementary Figure S9). In addition, although the *P. maniculatus*
220 assembly carries the standard orientation at inv22.0, evidence for this inversion within the *P.*
221 *leucopus* genome suggests that this rearrangement is shared across greater phylogenetic
222 distances (Supplementary Figure S10). Based on the polarizations of each segregating inversion
223 polymorphism, Table 1 summarizes the frequencies of the inverted arrangement in each Gulf
224 Islands population.

225 *High divergence between arrangements and variable levels of diversity within inversions*

226 To gain insights about the evolutionary histories of segregating inversions, we compared levels
227 of nucleotide diversity measured between arrangements ($d_{XY}^{subtype}$), within the inverted class
228 (π_{inv}), and within the standard class (π_{std}). There is substantial divergence between
229 arrangements ($d_{XY}^{subtype}$). At all loci, mean $d_{XY}^{subtype}$ for each population exceeds levels of
230 diversity measured within either class of arrangements (Supplementary Table S1). Divergence
231 between arrangements is also large relative to genome-wide levels of diversity. Across all
232 populations and inversion loci, mean $d_{XY}^{subtype}$ exceeds mean nucleotide diversity measured at
233 putatively neutral, non-inverted loci (π_{genome}) (Figure 6). Divergence between arrangements is
234 greatest for inv15.0 and lowest for inv21.0 (Figure 6).

235 For all loci, relative levels of nucleotide diversity among inverted haplotypes (π_{inv}/π_{std}) are
236 similar across all three Gulf Islands populations. For inv3.0, inv7.2, inv15.0, and inv22.0, inverted
237 arrangements show reduced nucleotide diversity compared to their corresponding standard
238 arrangements (Figure 7A, 7B, 7D, and 7F). These reductions are most extreme for inv7.2 and
239 inv15.0, where mean nucleotide diversity among inverted haplotypes (π_{inv}) is less than 14% of
240 mean nucleotide diversity among standard haplotypes (π_{std}) (Figure 7B and 7D). Both inv14.0
241 and inv21.0 exhibit the opposite trend, with mean π_{inv} exceeding mean π_{std} (Figure 7C and 7E).

242 *Demographic history*

243 We reconstructed the joint demographic history of Saturna Island, Pender Island, and mainland
244 Maple Ridge populations to estimate population size changes, split times, and migration rates,
245 and to provide context for patterns of variation at inversion polymorphisms. Using *Moments*
246 (Jouganous et al. 2017), we fit a variety of simplified demographic models to joint SFS
247 constructed from putatively neutral SNPs falling outside the boundaries of known inversion
248 polymorphisms. Figure 8 depicts the best-fitting three-population demographic model.
249 Maximum likelihood parameter estimates from this model suggest that ancestral island and
250 mainland populations diverged 2,565,592 generations ago and were connected by a low rate of
251 migration (m = 6.75e-6) until 5,090 generations ago (Figure 8). At this time, the contemporary
252 Saturna Island and Pender Island populations split from their shared ancestor and the mainland
253 Maple Ridge population underwent an over eight-fold size reduction (Figure 8). Although this

best-fit model provides evidence for migration between the contemporary Saturna Island and Pender Island populations (m = 1.63e-7), it does not support migration between the islands and mainland during this recent epoch (Figure 8). Effective population size (N_e) estimates from this model suggest a large N_e for both the ancestral island (N_e = 255,958) and ancestral mainland (N_e = 579,644) populations and smaller N_e for contemporary populations (N_e = 18,607 for Saturna Island, 14,984 for Pender Island, and 69,313 for mainland Maple Ridge) (Figure 8).

Coalescent simulations conducted under this candidate model yielded a close fit to the means of empirical distributions of π, Tajima's *D*, and F_{ST} (Figure 9A-C). Variance among loci in these summary statistic estimates is greater for the empirical data than for the simulated data, which could reflect heterogeneity in the recombination rate and mutation rate not captured by simulations. In contrast to the close agreement between empirical data and simulated data for diversity-related summaries, our best-fit demographic model underestimates levels of LD. In each population, average pairwise r^2 decays slower than predicted by simulated data (Figure 9D). One explanation for this disagreement is that recent population size histories are more dynamic than suggested by our SFS-derived demographic model. To investigate this possibility, we used levels of LD to estimate contemporary N_e, following the approach implemented in *currentNe* (Santiago et al. 2024). Our results indicated that contemporary N_e in each population is more than an order of magnitude smaller than suggested by our SFS-based model (N_e = 679 for Saturna Island, 219 for Pender Island, and 278 for mainland Maple Ridge) (Supplementary Table S2).

Shared history of inversions across populations

Leveraging the shared history of the three Gulf Islands populations, we used patterns of divergence between populations to shed further light on the evolutionary relationships among inverted and standard arrangements. Whereas $d_{XY}^{subtype}$ compares opposite arrangements sampled from the *same* population, divergence measured between inverted (d_{XY}^{inv}) and standard (d_{XY}^{std}) haplotypes compares identical arrangements sampled from *different* populations. At each locus, mean $d_{XY}^{subtype}$ exceeds mean d_{XY}^{inv} observed among population pairs, suggesting that inverted haplotypes share a more recent common ancestor with inversions present in other populations than with the standard haplotypes found in the same population (Figure 10). Although relative levels of divergence among inverted haplotypes ($d_{XY}^{inv}/d_{XY}^{std}$) are similar across population pairs, there is substantial variability in this ratio across loci (Figure 10). For inv3.0, inv7.2, inv15.0, and inv22.0, mean d_{XY}^{inv} is lower than mean d_{XY}^{std}, with inv7.2 and inv15.0 exhibiting the most extreme reductions (Figure 10A, 10B, 10D, and 10F). In contrast, mean d_{XY}^{inv} is greater than mean d_{XY}^{std} at inv14.0 and inv21.0 (Figure 10C and 10E), mirroring the patterns described above for π_{inv}/π_{std} (Figure 7).

DISCUSSION

Our genomic characterization of *P. maniculatus* from the Gulf Islands revealed widespread effects of both polymorphic inversions and demographic history on patterns of variation within and between populations. In the following sections, we discuss how these results enrich our understanding of inversion evolution and the evolution of island populations in this unique system.

A shared history for inversions segregating in island and mainland populations

Our profile of inversions in the Gulf Islands revealed surprising similarities between populations. Despite an estimated divergence time between ancestral island and mainland lineages of 2,560,502 generations ago, all three Gulf Islands populations segregate the same six inversion polymorphisms. These congruencies stand in stark contrast to the geographic variability among polymorphic inversions observed across mainland populations of *P. maniculatus* (Harringmeyer and Hoekstra 2022).

Inversion polymorphisms could be shared among populations because they were inherited from a common ancestor, dispersed by gene flow, or generated by distinct mutations at the same loci. An enrichment of inverted repeats at inversion breakpoints suggests ectopic recombination as a likely mechanism for the origin of inversions in *P. maniculatus* (Harringmeyer and Hoekstra 2022). When combined with evidence for hotspots of transposable element activity within *P. maniculatus* (Gozashti et al. 2023), this observation raises the possibility that some genomic regions in *P. maniculatus* are more susceptible to inversion than others, as has been observed in diverse species (Cáceres et al. 1999; Pevzner and Tesler 2003; Corbett-Detig et al. 2019; Maggiolini et al. 2020). Although the inversions present in the Gulf Islands populations could have originated independently at such “fragile” sites, between-population comparisons revealed that inverted arrangements sampled from different populations have a lower time to most recent common ancestor (TMRCA) than opposite arrangements sampled from the same population. This indicates that inversions segregating in the Gulf Islands populations instead derive from a single origin at each locus. Based on the demographic history we inferred, these inversions either arose in the ultimate island-mainland ancestor or were spread among lineages during the long period of connectivity between ancestral island and ancestral mainland populations. Patterns of variation within each locus provide further support for the shared history of inversions. Despite distinct trends across loci, each locus shows broadly similar levels of within-arrangement diversity and between-arrangement divergence across all three populations.

Evidence for selection on inversions

Multiple findings suggest that the inversions segregating in Gulf Islands populations have been maintained longer than expected under neutrality. First, all six inversion polymorphisms are shared by the three Gulf Islands populations. Second, whole genome alignments of diverse *Peromyscus* reference assemblies suggest at least two inversions, inv14.0 and inv22.0, pre-date species splits. Finally, levels of divergence between inverted and standard arrangements far exceed genome-wide levels of diversity measured at putatively neutral, non-inverted loci, consistent with an elevated TMRCA between arrangements.

Using regions immediately inside inversion breakpoints (where flux between arrangements is minimized), inversion age can be estimated by comparing divergence between arrangements to diversity within the standard arrangement (Hasson and Eanes 1996; Corbett-Detig and Hartl 2012). Unfortunately, this approach assumes that levels of variation within the standard arrangement have remained constant since the emergence of the inversion (Hasson and Eanes 1996; Corbett-Detig and Hartl 2012; Charlesworth 2023), an assumption likely violated by the dynamic history of the Gulf Islands populations. Still, the rank order of mean $d_{XY}^{subtype}$ measured across loci indicates that, of the six polymorphisms segregating in the Gulf Islands populations, inv15.0 and inv7.2 are among the oldest (Figure 6; Supplementary table S1).

Models of long-term balancing selection predict an unusually old TMRCA at the targeted locus and a genealogy characterized by elongated internal branches (Takahata 1990; Hudson 1990; Charlesworth et al. 1997; Bitarello et al. 2023). These features lead to increased neutral variation at the locus, elevated linkage disequilibrium between neutral variants and the selected site, and shifts in the site frequency spectrum toward intermediate-frequency alleles (Bitarello et al 2023; Charlesworth 2023). In most genomic regions, recombination narrows the span of these signals, challenging efforts to detect balancing selection. In regions harboring polymorphic inversions, by contrast, recombination suppression in heterokaryotypes extends the signatures of balancing selection over much greater distances (Charlesworth 2023). At the same time, “flux” between arrangements driven by double crossovers and gene conversion events can cause patterns of neutral diversity to vary with distance from breakpoints and selected sites (Hasson and Eanes 1996; Navarro et al. 1997; Navarro et al. 2000; Corbett-Detig and Hartl 2012; Guerrero et al. 2012; Charlesworth 2023; Kapun et al. 2023).

Population genetic models that account for some of these complexities have provided insights into the effects of balancing selection on inversion loci. Navarro et al. (2000), Guerrero et al. (2012), and Charlesworth (2023) studied the behavior of coalescence times within and between arrangements under different models of balancing selection and population structure. Their findings indicated that expected patterns of diversity and divergence strongly depend on whether the system has reached equilibrium and the equilibrium frequency at which the inversion is maintained (Navarro et al. 2000; Guerrero et al. 2012; Charlesworth 2023). After a selectively favored inversion arises, its approach to equilibrium occurs in two phases (Navarro et

al. 2000; Guerrero et al. 2012; Charlesworth 2023). First, the inversion quickly rises to its balanced frequency, eliminating variation within the inverted class in a manner analogous to a selective sweep (Maynard Smith and Haigh 1974). Once this balanced frequency is reached, variation within the inverted class recovers through the accumulation of new mutations and genetic flux with the standard arrangement. In a symmetric fashion, variation within the standard class is lost because of its reduced frequency in the population. After equilibrium is established, relative levels of diversity within each arrangement class should reflect the frequencies at which they are maintained.

Most of the inversion polymorphisms segregating in the Gulf Islands populations (inv3.0, inv7.2, inv15.0, and inv22.0) exhibit reductions in nucleotide diversity within the inverted class that are consistent with either a recent selective sweep of the inverted arrangement or its long-term maintenance at a low equilibrium frequency. Inv15.0 and inv7.2 show the clearest evidence for recent selective sweeps: in each case, nucleotide diversity within the inverted class is eliminated along the entirety of the locus (Figure 11). As with other locus-wide summaries, the marked reduction in variation at these two loci is seen in all three Gulf Islands populations, raising the possibility that shared selective events occurred within ancestral lineages. Given their old age, the evidence for recent selection acting on these inversions is surprising. If the selective sweeps responsible for the observed reductions in diversity occurred soon after the inversions arose, then there would have been plenty of time for flux between arrangements and the accumulation of new mutations to erode sweep signatures. This logic raises the intriguing possibility that the selective mechanisms responsible for the long-term maintenance of these polymorphisms are different than those that have shaped patterns of variation more recently in the Gulf Islands populations.

We observe the opposite pattern for inv14.0 and inv21.0, where nucleotide diversity is instead reduced within the standard class. This result could suggest that the standard arrangement has been maintained at a low equilibrium frequency. However, given the dynamic population size histories we infer, the equilibrium status of these loci is uncertain. Though reductions in diversity within the standard class seem inconsistent with neutral evolution, they are difficult to reconcile with existing theory. Current models assume that selection on inversions is sufficiently strong that their equilibrium frequency is quickly achieved, and that the selective regime remains constant through time (Navarro et al. 2000; Guerrero et al. 2012; Charlesworth 2023). In reality, the dynamics could depend on both the mechanism of balancing selection (Fijarczyk and Babik 2015; Bitarello et al. 2023) and its targets (inversion breakpoints or alleles captured by the rearrangement) (Said et al. 2018; Kapun et al. 2023), features that may differ from inversion to inversion. Alternatively, the trends we observe at these two loci could reflect errors in our classification of homozygous genotypes. The accuracy of these assignments is influenced both by the age of the inversion and by our capacity to distinguish ancestral and derived states,

398 which could be limited by factors including the quality of reference assemblies and the
399 possibility of incomplete lineage sorting.

400 More broadly, our findings suggest that there could be important regional differences in the role
401 these inversions play in local adaptation. The same inversions we examined show strong
402 evidence for selection in mainland populations of *P. maniculatus* from the Cascades region,
403 where they exhibit clines across a forest-prairie gradient and associations with phenotypic
404 differences between forest and prairie ecotypes (Hager et al. 2022; Harringmeyer and Hoekstra
405 2022). In contrast to their role in ecotype differentiation among Cascades populations, none of
406 the inversion loci exhibit allele frequency differentiation among Gulf Islands populations (Table
407 1). Furthermore, the joint minor allele frequencies of the six inversion polymorphisms across
408 populations (computed by treating the loci as a group) are not unusual with respect to the joint
409 frequencies of neutral, non-inverted SNPs (Supplementary figure S13).

410 Important differences between the Gulf Islands and Cascades could lead to contrasting
411 evolutionary roles for the same inversions in these two systems. Although island and mainland
412 environments may present divergent selective regimes, the selective gradient responsible for
413 establishing inversion clines among Cascades populations might not exist in the Gulf Islands. In
414 addition, whereas forest and prairie populations in the Cascades region exhibit high rates of
415 migration across divergent environments (Hager et al. 2022), we found little support for
416 migration between island and mainland populations during the contemporary epoch of our
417 demographic model. By linking together adaptive allele combinations, inversions may facilitate
418 local adaptation in the face of high gene flow (Kirkpatrick and Barton 2005; Yeaman and Otto
419 2011; Yeaman and Whitlock 2011). Though this phenomenon may position inversions as an
420 important adaptive mechanism in the Cascades region, our demographic inferences indicate
421 that gene flow is not an impediment to adaptive divergence between island and mainland
422 populations in the Gulf Islands. Hager et al. (2022) suggest that inv15.0 pre-dates formation of
423 the current forest and prairie habitat distribution, raising the possibility that its role in local
424 adaptation has resulted from subsequent accumulation of adaptive alleles (Said et al. 2018;
425 Kapun et al. 2023). Future comparisons of the genetic variation carried by these inversions in
426 the Gulf Islands, Cascades, and other populations of *P. maniculatus* could help clarify their roles
427 in local adaptation.

428 *Prospects for inversion characterization*

429 Knowledge of inversion polymorphisms in *P. maniculatus* and the broader karyotypic diversity of
430 the genus pre-dates the genomic era (Ohno et al 1966; Robbins and Baker 1981; Stangl and
431 Baker 1984; Greenbaum et al. 1994), yet important biological properties of these loci remain
432 unknown. Uncertainty about local mutation rates and rates of genetic flux between
433 arrangements preclude more detailed model-based inferences about inversion age and

434 selection in this system. Nevertheless, our work suggests that powerful insights about inversion
435 dynamics can be extracted from short-read sequence data.

436 Our results also emphasize the potential confounding effects of inversions on population
437 genetic inferences about demographic history and selection. Segregating inversions contain an
438 abundance of SNPs that segregate at the same allele frequency, which can shift both the local
439 and genome-wide SFS. At the same time, inversions elevate LD over large genomic distances.
440 These defining characteristics have the potential to mislead inferences of demographic history
441 and selection that focus on the SFS or LD. Proper characterization of structural rearrangements
442 therefore constitutes an important and often overlooked first step in population genomic
443 analysis.

444 *Insights into the evolution of island populations*

445 The demographic history we reconstructed from genome-wide patterns of variation provides
446 new insights into the population dynamics of *P. maniculatus*. Assuming 1-2 generations per year
447 (Wolff 1985), the estimated split time of 5,090 generations between Pender Island and Saturna
448 Island is broadly consistent with the post-glacial establishment of these populations. During the
449 last glacial period, the southern Strait of Georgia and Gulf Islands region was covered by the
450 Cordilleran ice sheet (Clague 1980; Clague and James 2002; Hutchinson et al. 2004; Fedje et al.
451 2009; James et al. 2009). Following its peak extent 14,500 years ago, the retreat of glacial ice
452 drove rapid uplift of underlying land areas, with relative sea level in the southern Strait of
453 Georgia reaching its lowest point around 11,500 years ago (James et al. 2009). Subsequent rises
454 in global sea level driven by glacial melt established the present-day distribution of land area in
455 the region by 6,000 years ago (James et al. 2009). There is conflicting evidence about the extent
456 of land exposure between southeastern Vancouver Island and mainland British Columbia during
457 this period of sea-level adjustment (Gowan 2005; James et al. 2009), leaving unclear whether
458 mainland *P. maniculatus* could have dispersed to the islands via emergent land bridges. In the
459 absence of such overland routes, it is possible that *P. maniculatus* may have dispersed from
460 mainland areas by rafting on floating ice or debris (Steffan 2016). Notably, our island divergence
461 time estimate also post-dates human inhabitation of coastal British Columbia (Mackie et al.
462 2011; Mackie et al. 2018), raising the possibility that island populations could have instead been
463 founded by inadvertent human introductions.

464 Given its closer proximity and greater connectivity to the Gulf Islands during the post-glacial sea
465 level low stand, Vancouver Island may be a more likely source for the Saturna Island and Pender
466 Island populations than mainland British Columbia. Radiocarbon dating of *Peromyscus* fossils
467 from northern Vancouver Island suggest that *P. maniculatus* could have been present as early as
468 14,000 years ago (Steffan 2016). As this estimate pre-dates deglaciation of the region (Clague
469 and James 2002; Hutchinson et al. 2004; Fedje et al. 2009; James et al. 2009), it indicates that

Vancouver Island *Peromyscus* fossils might reflect dispersals from nearby glacial refugia (Steffan 2016). Population genetic analysis of *P. maniculatus* from Vancouver Island and other potential source populations could increase the resolution of demographic inferences for mice from Saturna Island and Pender Island.

The ancient divergence time between ancestral island and ancestral mainland populations we recovered pre-dates the last inter-glacial period substantially (Clague 1980), leaving its connection to the geological history of the region unclear. Though there is phylogeographic evidence for deep divergence among certain geographic clades of *P. maniculatus* (Sawyer et al. 2017), our results are surprising given the proximity of the island and mainland sites we sampled. Demographic analysis of *P. maniculatus* from across the Cascade mountain range produced similar divergence time estimates (2.2 million generations ago) between western populations of the forest ecotype and eastern prairie-ecotype populations (Hager et al. 2022). The similarity in divergence times and the broad proximity of the Gulf Islands and Cascades regions suggest that island-mainland divergence and forest-prairie divergence in these two systems were driven by shared geographic features. Indeed, the Fraser River, which bisects south central British Columbia and northern Washington along its eastern margin, could have shaped population structure among *P. maniculatus* in the Pacific Northwest (Zheng et al. 2003). Demographic modeling that incorporates both Gulf Islands and Cascades mice could be used to further investigate this hypothesis.

Our results also speak to demographic processes operating more recently. Analyses of LD in the Gulf Islands populations indicated that values of contemporary N_e are 1-2 orders of magnitude lower than historical estimates based on the SFS. The causes of these recent, severe declines in population size are unknown, but they suggest a reduced capacity for selection to remove deleterious mutations and to spread beneficial mutations in these populations. In addition, our findings serve as a reminder that different aspects of genomic variation respond to evolutionary processes happening on disparate timescales (Beichman et al. 2018; Nadachowska-Brzyska 2022). We found a similar discordance between N_e estimates based on allele frequencies and patterns of linkage disequilibrium in island populations of the white-footed mouse, *P. leucopus* (Howell et al. 2025). Commonalities between these systems, such as the fragmentation and urbanization of habitat areas (Eis and Craigdallie 1980; Richburg and Patterson 2005), might suggest that fluctuations in effective population size on contemporary timescales are driven by shared processes. Similar comparisons in natural populations of other species could reveal the extent to which these disparities are a universal reflection of human activity.

Finally, our inferred demographic history provides context for interpreting patterns of phenotypic evolution on islands. Mice on Saturna Island exhibit multiple hallmarks of the "island syndrome" (Adler and Levins 1994), including larger body sizes, higher population densities, lower dispersal, and reduced intraspecific aggression (Redfield 1976; Sullivan 1977;

507 Halpin and Sullivan 1978). Saturna Island mice are 35% heavier than mainland mice when raised
508 in a common laboratory environment, indicating a heritable difference in body size (Baier and
509 Hoekstra 2019). The ancient split time between island and mainland lineages substantially
510 extends the temporal window during which the mice that now inhabit Saturna Island and
511 Pender Island could have evolved island syndrome phenotypes. Comparisons of molar sizes
512 between ancient and modern specimens suggest that Pleistocene-era *P. maniculatus* from
513 Vancouver Island were significantly larger than their modern counterparts (Steffan 2016). This
514 inference raises the intriguing possibility that the large body sizes of Saturna Island and Pender
515 Island mice (Redfield 1976; Sullivan 1977; Baier and Hoekstra 2019) are relics of this ancient
516 characteristic.

517

518 **MATERIALS AND METHODS**

519 *Fieldwork and sample collection*

520 We sampled deer mice (n=120) in the Gulf Island National Park Reserve on Saturna Island and
521 Pender Island (*P. m. saturatus*) and in the Malcolm Knapp Research Forest (MKRF) on the
522 mainland (*P. m. austerus and P. keeni*) in October 2014. Of these, we collected 101 individuals,
523 and either sacrificed and froze them in the field or transported them alive to Harvard University.
524 Specimens were accessioned in the Harvard Museum of Comparative Zoology (MCZ 69713–
525 69813) and liver samples were stored in ethanol. We verified species identity of specimens
526 using a fragment of the cytochrome b gene following Zheng et al. (2003) (see electronic
527 supplementary material, Figure S1, in Baier and Hoekstra (2019) for details). The samples in the
528 present study were previously examined by Baier and Hoekstra (2019). We sequenced 79 of the
529 101 mice for which liver tissue was collected (n=31 for Saturna Island, n=16 for Pender Island,
530 and n=32 for mainland Maple Ridge).

531 *Population sampling and sequencing*

532 DNA was extracted from liver samples using Gentra Puregene Tissue Kits (Qiagen, Germantown,
533 MD, USA). Sample purity was assessed by measuring 260/280 absorbance ratios using a
534 NanoDrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). DNA
535 concentrations were measured using the Qubit dsDNA HS Assay (Life Technologies, Carlsbad,
536 CA, USA). We used agarose gel electrophoresis to assess DNA sample quality and confirm the
537 presence of high-molecular weight DNA prior to sequencing. Library preparation and
538 sequencing were conducted by the University of Wisconsin-Madison Biotechnology Center. The
539 Illumina DNA Library Prep Kit was used to construct sequencing libraries. Purified and pooled
540 libraries were sequenced on the Illumina NovaSeq X Plus platform to generate paired-end 150
541 bp reads. Samples were sequenced to a targeted coverage of 30X per individual. As a quality

542 control measure, we re-sequenced four mice: one male each from Saturna (FSB85), Pender
543 (FSB114), and Maple Ridge (FSB42) as well as a single female from Maple Ridge (FSB6). These
544 technical replicates were sequenced using unique libraries, allowing us to capture the effects of
545 sequencing/genotyping error introduced at all laboratory and bioinformatic steps downstream
546 of PCR-based library preparation.

547 *Sequence data processing, variant calling, and variant filtering*

548 Prior to alignment and variant calling, raw sequencing reads were pre-processed to remove
549 adapter sequences and low-quality base calls. BBDuk v38.96
550 (https://sourceforge.net/projects/bbmap/) was used to remove Illumina universal adapter
551 sequences and Trimmomatic v0.39 (Bolger et al. 2014) was used to trim low-quality base calls.
552 Processed sequencing reads were then aligned to the *P. maniculatus* NCBI RefSeq assembly
553 (HU_Pman_2.1.3; GCF_003704035.1) using BWA-MEM v0.7.17 (Li and Durbin 2009).

554 We based our variant calling approach on the Genome Analysis Toolkit (GATK) Best Practices
555 recommendations for germline short variant discovery (McKenna et al. 2010; Van der Auwera
556 and O'Connor 2020). Prior to variant identification, we used GATK v4.2.0.0 MarkDuplicates to
557 identify and exclude duplicate reads from downstream variant calling. We then computed
558 genotype likelihoods on a per-individual basis using GATK v4.2.0.0 HaplotypeCaller with the
559 Reference Confidence Model set to “GVCF” mode (Poplin et al. 2017). The resulting GVCF call
560 sets were combined using GATK v4.2.0.0 CombineGVCFs. GATK v4.2.0.0 GenotypeGVCFs was
561 then used to perform joint variant calling separately for the Saturna (n=32), Pender (n=17), and
562 Maple Ridge (n=34) cohorts (Poplin et al. 2017). This approach identified a total of 110,458,672
563 SNPs in the Saturna cohort, 100,102,299 SNPs in the Pender cohort, and 151,071,903 SNPs in
564 the Maple Ridge cohort.

565 We filtered our variant calls with GATK v4.2.0.0 VariantFiltration using site-level annotation
566 thresholds based on GATK Best Practices recommendations for hard filtering (“FisherStrand” >
567 60.0, “StrandOddsRatio” > 3.0, “RMSMappingQuality” < 40.0, “MappingQualityRankSumTest” <
568 -12.5, and “ReadPosRankSumTest” < -8.0; Van der Auwera and O'Connor 2020) with
569 modifications (“QualByDepth” < 5.0). We then used GATK v4.2.0.0 SelectVariants to remove
570 filtered sites and restrict each call set to autosomal, biallelic, single-nucleotide variants. We
571 retained 78,214,290 SNPs in the Saturna cohort, 72,836,785 SNPs in Pender, and 106,361,479
572 SNPs in Maple Ridge following these basic variant filtering and selection procedures.

573 The population-specific joint calling approach we employed can yield incongruent call sets when
574 there is sufficient allele frequency differentiation between populations because variants private
575 to one cohort will not be represented among the variant calls of other cohorts. Variants that are
576 “missing” between call sets could be monomorphic for the reference allele within their cohort,
577 or they could be sites for which high-quality genotype calls could not be obtained. To avoid this

578 ambiguity and ensure that joint allele frequencies were properly calibrated for downstream
579 analyses, we "rescued" genotypes at these missing sites. Briefly, we used the bcftools v1.8
580 (Danecek et al. 2021) "isec" function to extract the locations of private variants observed in
581 each call set. We then used the "-all-sites" mode of GATK v4.2.0.0 GenotypeGVCFs with the "-
582 stand-call-conf" threshold set to 0 to emit monomorphic reference calls at these private variant
583 sites in each of the other cohorts. We used GATK v4.2.0.0 VariantFiltration and SelectVariants to
584 apply the same site-level filtering thresholds to these "rescued" calls (excluding
585 "QualByDepth"). The resulting high-quality invariant calls were then combined with the
586 polymorphic calls of their respective cohort and resulting monomorphic + polymorphic call sets
587 were merged across cohorts, yielding a single multi-population call set that represented the
588 union of variant sites observed across the three populations.

589 *Identification and exclusion of close relatives*

590 Using KING v2.2.8 (Manichaikul et al. 2010), we estimated pairwise kinship coefficients among
591 individuals within each population cohort. Prior to kinship analysis, each of the filtered single-
592 population autosomal call sets described above was pruned for LD by removing SNPs with
593 pairwise $r^2 > 0.1$ in 50 bp sliding windows using the "--indep-pairwise 50 10 0.1" option in plink
594 v1.90 (Chang et al. 2015). This LD-pruning left a total of 2,853,858 SNPs in the Saturna cohort,
595 2,079,267 in Pender, and 5,503,047 in Maple Ridge. We applied the KING-Robust algorithm to
596 each LD-pruned call set and determined relationships from the resulting kinship coefficient
597 estimates using the bounds suggested by the KING documentation (Manichaikul et al. 2010).
598 We identified a large number of closely related individuals within all three cohorts, with 4.44%
599 of Saturna Island pairs, 10.29% of Pender Island pairs, and 6.95% of mainland Maple Ridge pairs
600 exceeding the lower bound for 2nd degree relatives. We did not find close relatives among
601 individuals sampled from different locations. To mitigate the potential biases that close relatives
602 can introduced to population genetic analyses (Wang 2017), we excluded related individuals
603 from downstream analyses. To do so, we used the "relatednessFilter" function of the R package
604 plinkQC v0.3.4 to identify the maximum set of unrelated individuals (defined here as pairs $\geq 3^{rd}$
605 degree relatives) within each sample. This left us with unrelated sample sizes of n=20 for
606 Saturna, n=10 for Pender, and n=17 for Maple Ridge.

607 *Population structure analysis*

608 We evaluated population structure within and between sampling locations by assessing PCA
609 clustering patterns and conducting model-based inference of global ancestry proportions with
610 ADMIXTURE (Alexander et al. 2009). Since both approaches require unlinked SNPs, we
611 conducted the same LD-pruning described above on the filtered multi-population autosomal
612 call set, restricting our analyses to the unrelated subset of individuals identified in each cohort.
613 PCA was performed on this combined call set using the "--pca" option in plink v1.90 (Chang et

614 al. 2015). Ancestry proportions were inferred either for each population sample separately or
615 for the combined cohort with ADMIXTURE v1.3.0 (Alexander et al. 2009). Five-fold cross-
616 validation was performed for a range of ancestral components (k=1 to k=5) using the "--cv"
617 option. Individual ancestry proportions were then inferred for the k value that produced the
618 lowest cross-validation error.

619 *Inversion identification*

620 We used GATK v4.2.0.0 VariantsToTable to extract information about position, allele count,
621 allele number, allelic states, and individual genotypes at each SNP in the filtered autosomal call
622 sets described above. The resulting genotype tables were used to inspect properties of the
623 genome-wide folded SFS within each population. To investigate the source of "bumps" observed
624 in each population's SFS, we partitioned genotype tables by chromosome in order inspect
625 chromosome-wide allele frequencies. To assess LD decay along each chromosome, we used the
626 "--r2" option in plink v1.90 (Chang et al. 2015) to measure pairwise r^2 among SNPs sampled
627 from each chromosome (set by the "--chr" argument) in the filtered VCF call sets. Singleton
628 variants and sites with missing data were excluded from these LD calculations by setting the "--
629 mac" argument to 1 and the "--geno" argument to 0.

630 In order to "screen" each population sample for evidence of segregating inversions, we looked
631 for intervals that exhibited a high density of either homozygous or heterozygous SNPs in each
632 individual at a specific MAF, consistent with an excess of fixed differences between inverted and
633 standard arrangements. To perform this screen, we partitioned each population's genotype
634 table by chromosome and by every possible MAF. Coloring SNPs based on their genotype in
635 each individual, we then constructed maps of these MAF-partitioned SNPs and assessed
636 whether regions exhibiting high SNP density and genotype continuity overlapped the
637 boundaries of previously characterized inversion polymorphisms (Harringmeyer and Hoekstra
638 2022). MAF-partitioned genotype maps were constructed and visualized using ggplot v3.5.1
639 (Wickham 2016) in R v4.2.1 (R Core Team 2022).

640 To confirm the presence of the inversion polymorphisms identified by our screen and to
641 determine whether additional loci may be segregating in any of the Gulf Islands populations, we
642 conducted PCA within each of the 21 inversion loci characterized by Harringmeyer and Hoekstra
643 (2022). Starting with the filtered multi-population call set, we used the "--extract" function in
644 plink v1.90 (Chang et al. 2015) to extract SNPs falling within each inversion's breakpoints, then
645 performed PCA on each subset of SNPs using the "--pca" function. Because segregating loci
646 carry an abundance of fixed differences that distinguish standard and inverted arrangements,
647 we did not prune SNPs within inversion intervals for LD prior to analysis. For comparison, we
648 also conducted PCA on genome-wide SNPs falling outside of known inversion intervals. This
649 subset of SNPs was pruned for LD using the same parameters described previously. For each

650 locus, we evaluated whether PCA clustering patterns were consistent with three genotypic
651 classes: homozygous ALT, heterozygous, and homozygous REF. Though inv7.3 exhibits patterns
652 consistent with a segregating inversion (Supplementary figure S4), we note that this is likely
653 driven by polymorphism at the larger inv7.2 locus in which it is nested, and as such do not treat
654 it as a separate polymorphism. For each polymorphic locus, we assigned individuals to one of
655 three genotypic classes (homozygous ALT, heterozygous, and homozygous REF) based on their
656 proportion of ALT alleles at (Supplementary figure S5) or approaching (Supplementary figures
657 S6-S8) the MAF of the inversion.

658 *Inversion polarization*

659 In order to connect the genotype assignments described above (homozygous REF, heterozygous,
660 and homozygous ALT) to arrangement status in each individual (inv/inv, inv/std, and std/std), we
661 polarized segregating inversion loci to determine whether the *P. maniculatus* assembly carries
662 the inverted (i.e., derived) or standard (i.e., ancestral) orientation at each locus. To assess
663 evolutionary relationships among existing *Peromyscus* reference genomes, we ran Mashtree v
664 1.4.6 (Katz et al. 2019) on all of the existing *Peromyscus* NCBI RefSeq assemblies: *P. maniculatus*
665 (HU_Pman_2.1.3; GCF_003704035.1), *P. polionotus* (HU_Ppol_1.3.3; GCA_003704135.2), *P.*
666 *leucopus* (UCI_PerLeu_2.1; GCF_004664715.2), *P. californicus* (ASM782708v3;
667 GCF_007827085.1), *P. eremicus* (PerEre_H2_v1; GCF_949786415.1), *P. nudipes* (Pnud_10x_v1;
668 GCA_902168325.1), *P. melanophrys* (Pmel_10x_v1; GCA_902168415.1), *P. attwateri*
669 (Patt_10x_v1; GCA_902168425.1), and *P. aztecus* (Pazt_10x_v1; GCA_902168405.1). The
670 resulting distance-based tree (Supplementary figure S14) recovers previously characterized
671 phylogenetic relationships by Bradley et al. (2007) and Platt et al. (2015). We focused our
672 polarization analysis on four of these species: *P. polionotus*, *P. leucopus*, *P. californicus*, and *P.*
673 *eremicus*, which collectively span 1.5 to 5 million years of divergence from *P. maniculatus*
674 (Bradley et al. 2007; Platt et al. 2015).

675 We used the nucmer module of the MUMmer v4.0.0 alignment package (Marçais et al. 2018) to
676 conduct pairwise, whole genome alignments between *P. maniculatus* and each of the four focal
677 *Peromyscus* species. Each pair of reference assemblies was aligned using the following
678 parameters: "--maxmatch", "-c 500", "-b 500", and "-l 100". The resulting alignments were
679 filtered using the delta-filter module with parameters "-m", "-i 90", and "-l 100". We then used
680 the show-coords module with parameters "-THrd" to summarize the filtered alignments.

681 Supplementary figures S15-S18 depict the resulting dotplots from our initial whole genome
682 alignments. Though the closely related *P. polionotus* assembly exhibits one-to-one homology at
683 the chromosome level (Supplementary figure S15), we identified major inconsistencies between
684 chromosome name and order between *P. maniculatus* and the more diverged *P. leucopus*
685 (Supplementary figure S16) and *P. eremicus* assemblies (Supplementary figure S17). Both the *P.*

686 *leucopus* and *P. eremicus* assemblies support the fusion of chromosomes 16 and 21 into a single
687 scaffold (chr16_21) and the splitting of chromosome 8 into two separate scaffolds (chr8a and
688 chr8b). According to Long et al. (2019), chromosome assignments for the *P. leucopus* reference
689 were based on the mapping of markers from the *P. maniculatus* linkage map (Kenney-Hunt et al.
690 2014; Brown et al. 2018) and based on the results of their Hi-C contact analysis, which
691 supported two separate chromosome 8 scaffolds and a fused chromosome 16 and chromosome
692 21 scaffold. Neither of these fusions or splits are reflected in the *P. maniculatus* assembly.
693 Instead, we found that chr16_21 in both *P. eremicus* and *P. leucopus* is homologous with chr21
694 in *P. maniculatus* (Supplementary figures S16 and S17). For the split chromosome 8, we find that
695 chr8a and chr8b in *P. eremicus* align with chr8 and chr16 in *P. maniculatus*, respectively
696 (Supplementary figure S17). In contrast, chr8a and chr8b in *P. leucopus* align with chr16 and
697 chr8 in *P. maniculatus* (Supplementary figure S16). We also find that the assignments of chr11
698 and chr15 in both *P. eremicus* and *P. leucopus* are switched in *P. maniculatus* (Supplementary
699 figures S16 and S17). We used the "replace" function in seqkit v2.9.0 (Shen et al. 2016) to re-
700 assign chromosome names in both *P. eremicus* and *P. leucopus* reference FASTAs to reflect their
701 homology with *P. maniculatus*. For the scaffold-level *P. californicus* assembly FASTA, we assigned
702 names to the 24 largest scaffolds to reflect their alignment to *P. maniculatus* chromosomes
703 (Supplementary figure S18).

704 In addition to these chromosome naming inconsistencies, we found that pairwise alignments
705 with *P. leucopus*, *P. eremicus*, and *P. californicus* all exhibit reverse-strand errors in which
706 homologous chromosomes in each pair are represented by the forward strand in one species
707 and the reverse strand in the other (Supplementary figures S15-S18). These errors manifest as
708 whole-chromosome inversions and can confound the identification of intra-chromosomal
709 rearrangements. To correct these errors, we used the "seq -r -p" option in seqkit v2.9.0 (Shen et
710 al. 2016) to reverse-complement chr1, chr2, chr5, chr8, chr9, chr13, chr16, chr17, chr22, and
711 chrX in the alignment with *P. californicus* and chr1, chr5, chr9, chr11, chr12, chr17, chr18, chr20,
712 and chrX in the alignments with both *P. eremicus* and *P. leucopus*.

713 Following these operations on the reference FASTAs, we re-ran each alignment with the same
714 MUMmer parameters described above. We used SyRI v1.7.1 (Goel et al. 2019) to identify large-
715 scale structural rearrangements between species from these revised alignments and visualized
716 the resulting synteny plots using plotsr v1.1.1 (Goel and Schneeberger 2022).

717 *Population genomic analyses of inversions*

718 Taken together, the genotype assignments described above, which reflect the status of each
719 arrangement relative to the *P. maniculatus* reference assembly (i.e., homozygous ALT,
720 heterozygous, or homozygous REF), and our polarization results, which establish the orientation
721 of each locus in the reference, enabled us to determine the inversion genotype carried by each

722 mouse (i.e., inv/inv, inv/std, or std/std). These designations are summarized in Supplementary
723 table S3. To avoid switch errors that could be introduced by the phasing of inverted and
724 standard haplotypes in heterozygous individuals, we restricted our analyses of diversity and
725 divergence to the two homozygous classes of individuals (inv/inv and std/std).

726 Using these homozygous subsets, we computed nucleotide diversity π (Tajima 1983) within each
727 arrangement (π_{inv} and π_{std}) and divergence d_{XY} (Nei and Li 1979) measured either between
728 arrangements ($d_{XY}^{subtype}$) or between populations within each arrangement class (d_{XY}^{inv} and
729 d_{XY}^{std}). We used scikit-allel v1.3.6 (https://github.com/cggh/scikit-allel) to compute each of these
730 summaries in non-overlapping 5 kbp physical windows along each of the six segregating
731 inversion loci: inv3.0, inv7.2, inv14.0, inv15.0, inv21.0, inv22.0. Because the Maple Ridge sample
732 did not include any inversion homozygotes for inv14.0, we excluded within- and between-
733 arrangement summary statistic comparisons involving this population at this locus.

734 *Obtaining a high-confidence, putatively neutral SNP call set*

735 To obtain a high-confidence SNP call set for demographic analyses, we leveraged the technical
736 replicates of re-sequenced mice to adjust SNP filtering thresholds in a manner that achieved low
737 genotyping error rates while maintaining high SNP retention. In addition to the *site*-level
738 filtering described previously, we focused our more rigorous quality filtering on two *individual*-
739 level annotations: coverage depth (DP), which measures individual read count at a given SNP,
740 and genotype quality score (GQ), which measures the difference between the second lowest
741 and the lowest phred-scaled genotype likelihood. Whereas a low coverage depth can indicate
742 poor support for an individual's genotype call, artificially inflated coverage could reflect copy
743 number variation that is poorly resolved in the reference assembly (Li 2014; Pfeifer 2017). By
744 measuring pairwise genotype concordance between technical replicates, we used the rate of
745 genotype mismatch within each pair of duplicate samples as a proxy for genotyping/sequencing
746 error. To achieve an optimum balance of sensitivity and specificity, we tested a range of
747 maximum DP (41X, 46.5X, 52X, and 57.5X), minimum DP (2.5X, 8X, 13.5X, and 19X), and
748 minimum GQ (0, 25, and 50) thresholds. The impact of these filtering thresholds on SNP
749 retention and error rates is given in Supplementary table S4. Based on these findings, we
750 imposed the following thresholds using GATK v4.2.0.0 VariantFiltration: "DP" > 8, "DP" < 52, and
751 "GQ" > 25. These individual-level filters left a total of 46,287,217 SNPs in Saturna, 50,060,722
752 SNPs in Pender, and 57,424,563 SNPs in Maple Ridge while reducing the error rates in each
753 sample (measured as one-half the rate of genotype mismatch within each pair of duplicate
754 samples) to 0.06%, 0.05%, and 0.04%, respectively.

755 We also restricted downstream analyses to variants falling outside of annotated repetitive
756 elements, which can cause errors in read mapping and variant calling (Pfeifer 2017). We used
757 the union of the RepeatMasker (hub_2100979_repeatMasker; https://repeatmasker.org),

758 WindowMasker (hub_2100979_windowMasker; Morgulis et al. 2006), TandemRepeatFinder
759 (hub_2100979_simpleRepeat; Benson 1999), and Tandem Duplicates
760 (hub_2100979_tandemDups) annotation tracks of the UCSC Genome Browser
761 (http://genome.ucsc.edu; Karolchik et al 2004) to exclude SNPs within these regions. Repetitive
762 DNA that is poorly resolved in the reference genome can also manifest as regions of high
763 heterozygosity when reads derived from diverged repeats map to the same position. To mitigate
764 this issue, we used GATK v4.2.0.0 VariantFiltration to remove sites with "ExcessHet" > 15.0,
765 corresponding to a p-value less than 0.05 for the Hardy-Weinberg exact test of excess
766 heterozygosity (Wigginton et al. 2005).

767 To reduce the impact of direct and linked selection, we restricted our analysis to putatively
768 neutral SNPs by masking annotated genes in the *P. maniculatus* reference assembly. Using the
769 NCBI RefSeq track (hub_2100979_ncbiRefSeq) of the UCSC Genome Browser
770 (http://genome.ucsc.edu; Karolchik et al 2004), we excluded SNPs falling within 1 kbp of
771 predicted and curated genes. Because inversions may experience distinct recombination and
772 selective landscapes, SNPs falling within these intervals also violate the assumptions of SFS-
773 based demographic inference frameworks. Using the breakpoints defined by Harringmeyer and
774 Hoekstra (2022), we excluded sites within each of the 21 inversion polymorphisms they
775 characterized from downstream analyses. Following these additional masking and filtering
776 procedures, we were left with a total of 2,470,844 SNPs in Saturna, 2,315,952 SNPs in Pender,
777 and 3,455,232 SNPs in Maple Ridge to use for demographic inference.

778 *Demographic inference*

779 We employed the SFS-based inference approach of *Moments* (Jouganous et al. 2017) to
780 reconstruct the joint demographic history of the Gulf Islands populations. Using the high-
781 confidence, putatively neutral variant call set described above, we constructed 2D joint SFS
782 (jSFS) for each pair of populations and a 3D jSFS for the combined set of populations. jSFS were
783 folded to avoid errors in ancestral state identification and all sites containing missing genotype
784 information for one or more individuals were excluded from analysis. To estimate demographic
785 parameters from these jSFS, we used the maximum likelihood framework implemented in
786 *Moments* v1.1.15.

787 We began by fitting simplified two-population demographic histories to the 2D jSFS for each pair
788 of populations (Supplementary figures S19-S21). Each model featured the divergence of
789 contemporary populations from a shared ancestor, followed by either constant population sizes,
790 discrete changes in size, continuous changes in size, or a combination of discrete and
791 continuous changes in size (Supplementary figure S19). On top of these basic population size
792 histories, we fit additional parameters that allowed for either symmetric (Supplementary figure

S20) or asymmetric (Supplementary figure S21) migration between populations following their split.

To fit each of the tested two-population models, we used the BFGS optimization routine in *Moments* v1.1.15 to find the parameter values that maximized the likelihood of each observed 2D jSFS. To ensure thorough exploration of the parameter space, starting parameter values were permuted across 10,000 independent searches. For each tested model, we retained the top three combinations of parameter values that yielded the highest likelihood. After fitting each model, we inspected parameter convergence, model likelihoods, and residual fit to select between contrasting models. First, we eliminated models where the coefficient of variation for one or more parameters exceeded 0.2 across the top three highest likelihood estimates. Next, we compared model likelihoods either directly (for models with equivalent numbers of parameters), or through adjusted likelihood ratio tests (for models with nested parameters; Coffman et al. 2016). For remaining models, we inspected their fit to both the joint and marginal SFS using Anscombe residuals and selected the simplest model that yielded the best fit to all partitions of the jSFS.

Our two-population model fitting and selection procedure indicated that the island-island history was best captured by a model involving recent population divergence (Supplementary figure S22). In contrast, both island-mainland comparisons yielded best-fit models that involved ancient population divergence followed by recent discrete population size reductions (Supplementary figures S23 and S24). All three comparisons supported a history of symmetric gene flow between populations (Supplementary figures S22-S24).

To reconcile the extreme differences in divergence times suggested by the island-island and island-mainland comparisons, we used these candidate two-population histories to construct a model that reflects the shared history of all three Gulf Islands populations. This three-population model featured an older epoch in which the ancestral island and mainland lineages diverge from each other, followed by a more recent epoch reflecting the establishment of the contemporary island and mainland populations (Supplementary figure S25). Given the evidence for gene flow, we fit all possible parameterizations of symmetric migration among ancestral and contemporary populations (Supplementary figure S25) to the 3D jSFS. We used the same model fitting and model selection approach described above to obtain demographic parameter estimates and select among contrasting models. Residual differences between the joint and marginal SFS observed for each population/population-pair and the expected SFS under the best-fitting three-population demographic model are illustrated in Supplementary figure S26.

For both two- and three-population candidate models, the ancestral effective population sizes were calculated as $N_{Anc} = \theta/(4*\mu*L)$ using an assumed mutation rate, μ, of 5e-9 per-site, per-generation and an effective sequence length, L, that represents the total length of "callable" sequence. To approximate L, we multiplied the total length of the *P. maniculatus* autosomal

genome (2,302,074,966 bp) by the average percentage of SNPs (2.34%) that were retained following both the site- and individual-level filtering as well as the masks we applied to repetitive, genic, and inverted DNA. This yielded an approximate L of 53,870,000 bp. Using the formula above, estimates of population size and time parameters were scaled by N_{Anc} and $2*N_{Anc}$ to convert them into units of individuals and generations, respectively. Parameter uncertainties were estimated using the Godambe Information Matrix methods described in Coffman et al. (2016). We used uncertainty propagation techniques (Ku 1966) to obtain the standard deviations for parameters that are functions of multiple estimated quantities.

We conducted predictive simulations with *msprime* v1.2.0 (Baumdicker et al. 2022) to evaluate the fit of the inferred three-population demographic history. Using the maximum likelihood parameter estimates obtained for the best-fit three-population model, we conducted 2,500 independent simulations of a 1 Mbp genomic element for each component population assuming uniform recombination and mutation rates of 5e-9 per-site, per-generation. Simulated sample sizes were set to match those of the empirical data for each population. We computed nucleotide diversity π (Tajima 1983), Tajima's *D* (Tajima 1989), and F_{ST} (Hudson et al. 1992) in non-overlapping 5 kbp physical windows using scikit-allel v1.3.6 (https://github.com/cggh/scikit-allel) for both the simulated and empirical data. Together, the quality-based filtering we conducted on our empirical SNP call set and the masks we applied to repetitive, genic, and inverted DNA introduced variability in the number of sites "observed" among the 5 kbp physical windows that we used to compute each summary statistic. To account for this variability, we used the "is_accessible" argument in scikit-allel to compute per-site statistics based on the total number of observable sites. To measure the decay of LD in both simulated and empirical data, we used the "--r2" option in plink v1.90 (Chang et al. 2015) to compute pairwise r^2. We restricted this analysis to SNPs separated by at most 1 Mbp by setting the "--ld-window" and "--ld-window-kb" arguments to 1,000,000 and 1,000, respectively. Singleton variants were excluded from these calculations by setting the "--mac" argument to 1.

Using the same jSFS employed for demographic inference, we assessed whether the joint minor allele frequencies of the six segregating inversion loci were unusual compared to neutral, non-inverted loci. We summarized joint frequencies by computing the Euclidean distance between each population's minor allele frequencies at the six inversion loci. To obtain the null distributions of this distance measure, we constructed 1,000 random samples of six SNPs from the jSFS and computed the Euclidean distance between minor allele frequencies for each population pair using the dist() function in R v4.2.1 (R Core Team 2022).

Contemporary effective population size estimation

We used the LD-based approach of *currentNe* (Santiago et al. 2024) to estimate contemporary effective population sizes among Gulf Islands populations. Excluding close relatives in a random

sample of individuals can artificially inflate estimates of contemporary effective population size obtained through LD-based methods (Waples and Anderson 2017; Santiago et al. 2024). Since the excess of close relatives we found in each sample could be driven by small contemporary effective population sizes or nonrandom trapping, we chose to conduct these analyses on both the full sample (excluding technical replicates) and on the subset of unrelated individuals from each population (obtained by excluding relationships $\leq 2^{nd}$ degree). The same SNP call set employed for demographic inference was used as input to these analyses. We used the "-s" argument to restrict each analysis to 50,000 randomly sampled SNPs and restricted LD calculations to pairs of SNPs located on different chromosomes. When conducting analyses on samples that included close relatives, we used the "-k" option to jointly estimate contemporary effective population size and k, which represents the average number of full siblings each individual has in the sample (Santiago et al. 2024). We use the artificial neutral network approach implemented in *currentNe* to obtain 90% confidence intervals for each point estimate of contemporary effective population size (Santiago et al. 2024).

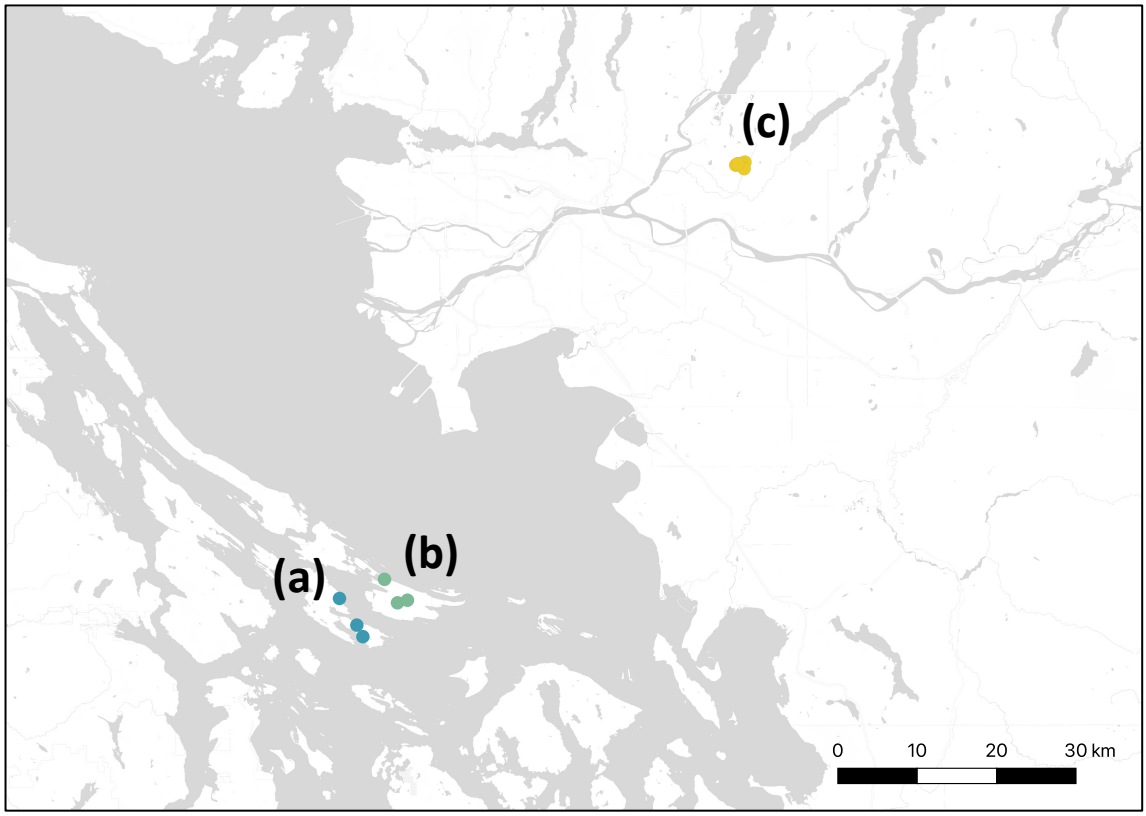

Figure 1. Map of the Gulf Islands in southwestern British Columbia. Land areas are depicted in white and water areas are depicted in gray. Points indicate sampled areas on Pender Island (a; blue), Saturna Island (b; green), and mainland Maple Ridge (c; yellow).

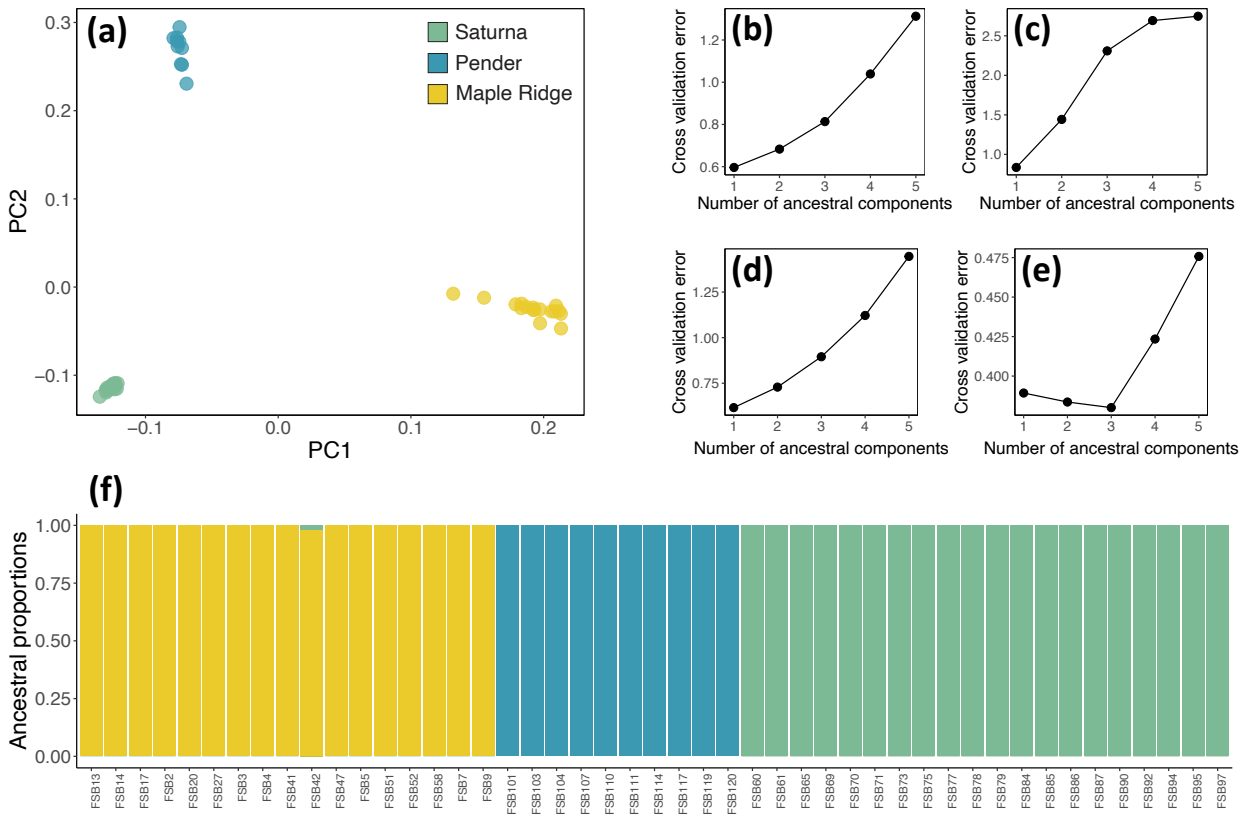

Figure 2. Evidence for structure among island and mainland sampling locations. Panel a depicts the results of PCA conducted using genome-wide unlinked SNPs. Each point represents an individual's position along the first (x-axis) and second (y-axis) principal component. Individuals are colored according to sampling location (yellow for mainland Maple Ridge, blue for Pender Island, and green for Saturna Island). Percentage of variance explained by PC1=7.73% and by PC2=4.65%. Panels b-e present the results of five-fold cross validation analyses performed with ADMIXTURE (Alexander et al. 2009) to identify the best-fit number of ancestral components within the Saturna (b), Pender (c), Maple Ridge (d), and combined (e) samples. Panel f depicts global ancestry proportions estimated for the combined sample, assuming k=3 ancestral components. Each individual is represented by a colored bar where the height of each colored segment within the bar gives the proportion of ancestry an individual derives from each component. Ancestral component colors reflect sampling location (as in Panel a).

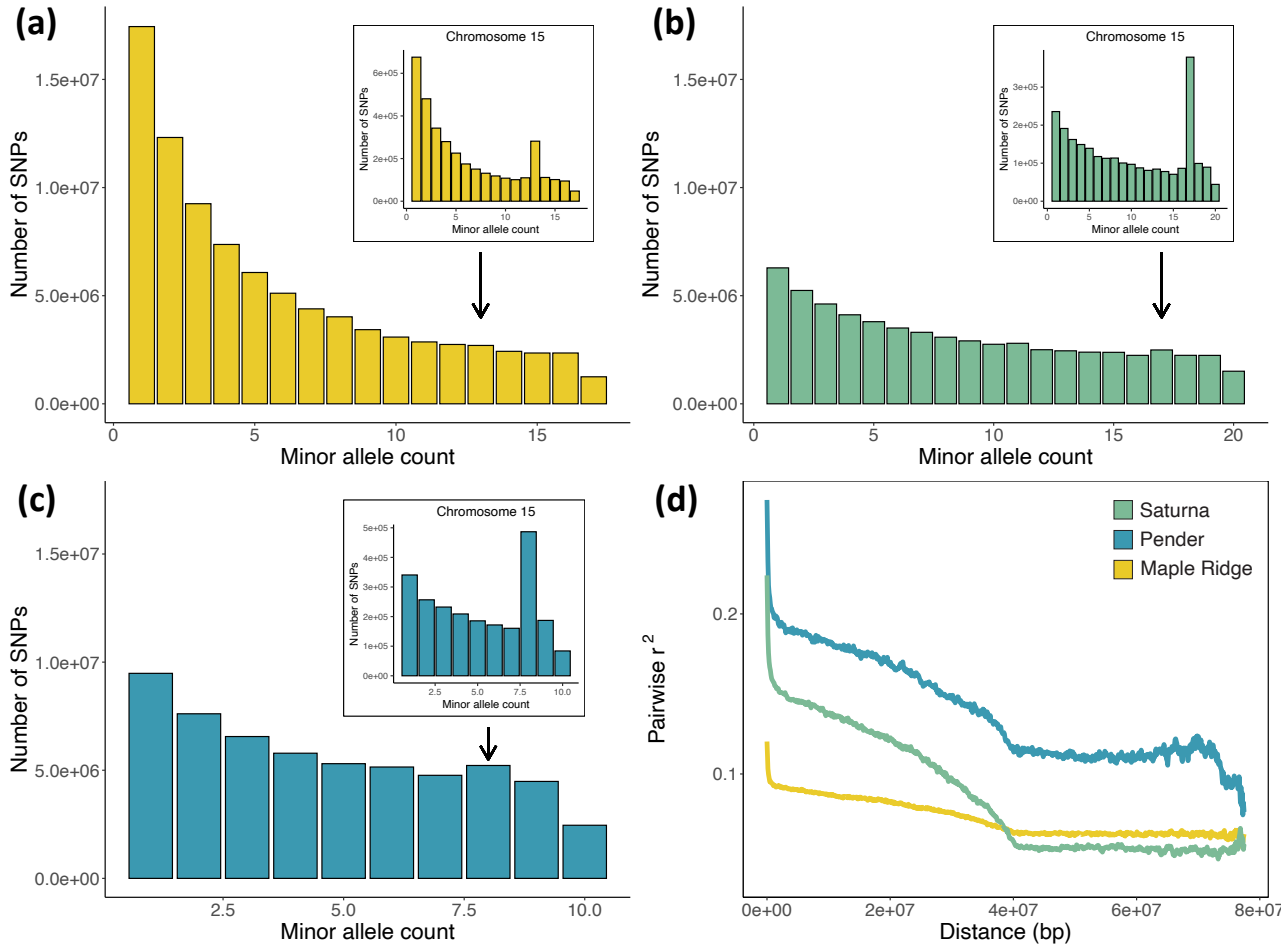

Figure 3. Inversion polymorphisms drive irregularities in the SFS and cause distortions in LD decay. Panels a-c provide an example of how the inversion polymorphism on chromosome 15 (inv15.0) affects both local and genome-wide allele frequencies in the Maple Ridge (a), Saturna Island (b), and Pender Island (c) samples. Main histograms plot the folded SFS constructed from genome-wide SNPs. Arrows indicate minor allele frequency bins in the genome-wide SFS that exhibit elevated SNP counts. Insets illustrate how these minor distortions in the genome-wide SFS correspond to major distortions in the local SFS for chromosome 15 SNPs. Panel d plots average pairwise r^2 (y-axis) between chromosome 15 SNPs separated by increasing distance (x-axis). Each line represents LD decay within a different population sample (yellow for Maple Ridge, green for Saturna, and blue for Pender).

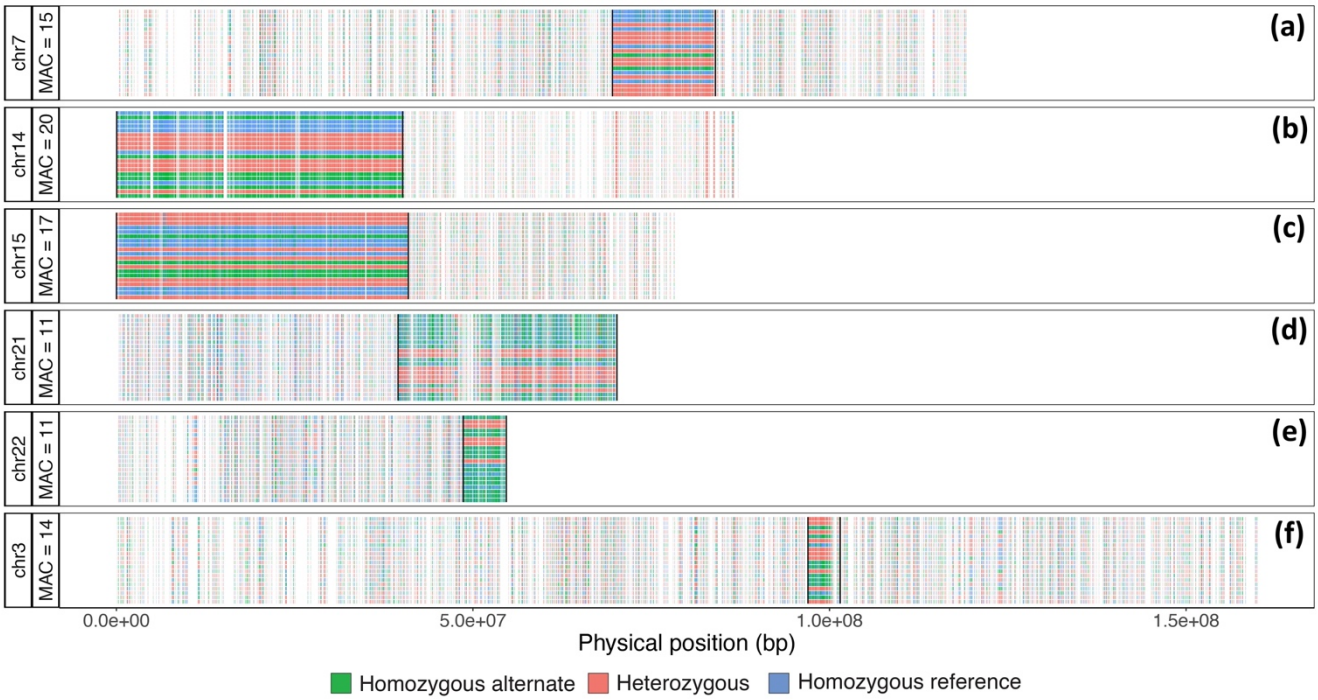

Figure 4. Partitioning SNP genotypes by minor allele frequency enables the detection and genotyping of polymorphic inversions. Plots illustrate the positive results for the inversion screen conducted in the Saturna Island sample. Each panel represents a different autosome: chr7 (a), chr14 (b), chr15 (c), chr21 (d), chr22 (e), and chr3 (f). Labels beneath each autosome name indicate the corresponding minor allele count that revealed the presence of the inversion polymorphism. X-axes measure the physical position along each autosome. Within a panel, each vertical tick mark indicates the location of a SNP with the specified minor allele count and each row denotes the genotype of that SNP in each individual. Blue tick marks represent homozygous reference genotypes, pink tick marks represent heterozygous genotypes, and green tick marks represent homozygous alternate genotypes. Thick black lines denote the breakpoints defined by Harringmeyer and Hoekstra (2022) for inv7.2 (a), inv14.0 (b), inv15.0 (c), inv21.0 (d), inv22.0 (e), and inv3.0 (f).

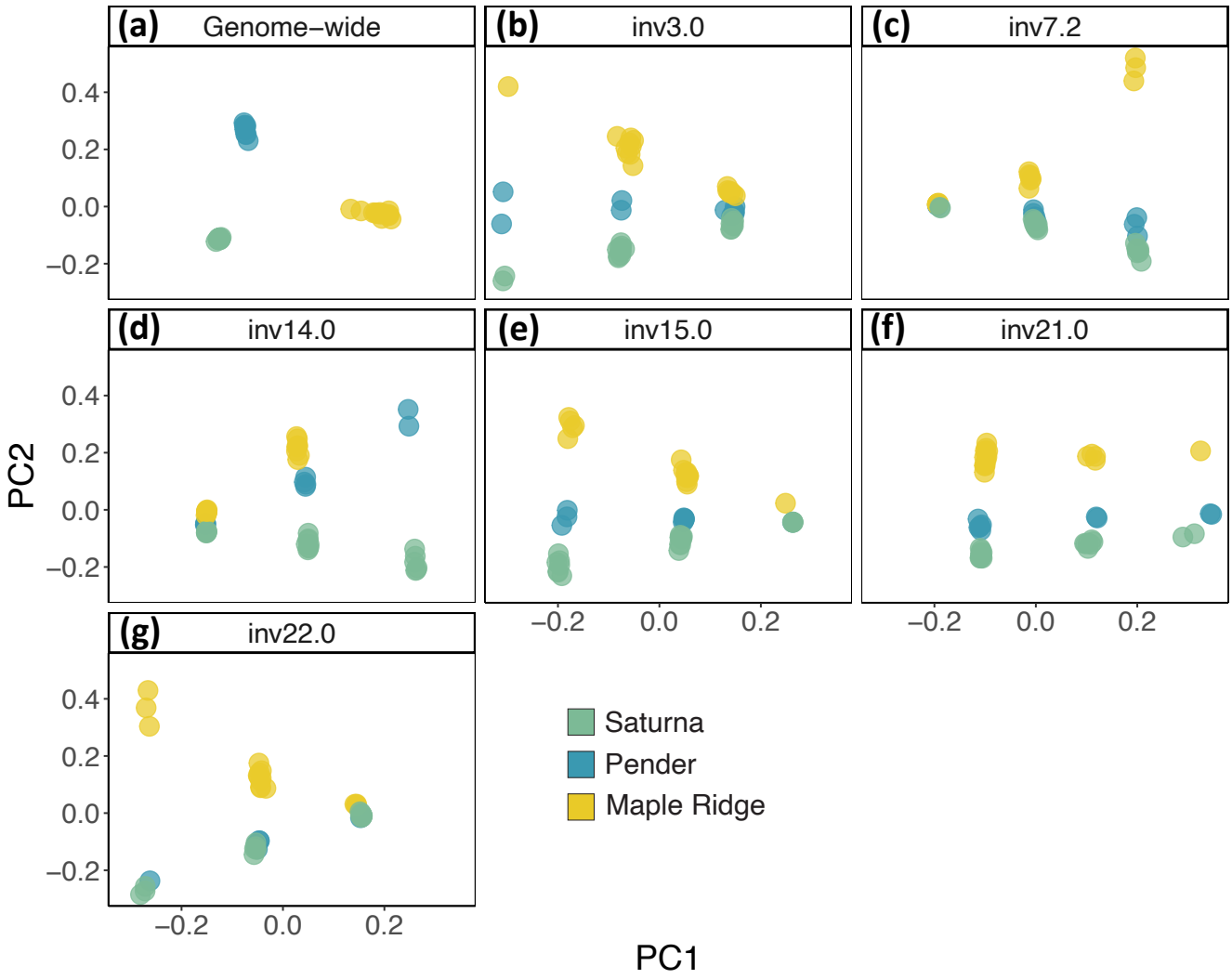

Figure 5. Individuals from distinct populations cluster according to inversion genotype at each polymorphic locus. Panels compare the results of PCA conducted on genome-wide SNPs falling outside of known inversions (a) to PCA performed on SNPs within each polymorphic inversion locus: inv3.0 (b), inv7.2 (c), inv14.0 (d), inv15.0 (e), inv21.0 (f), and inv22.0 (g). Each point denotes an individual's position along the first (x-axis) and second (y-axis) principal component. Points are colored according to population sample (yellow for Maple Ridge, blue for Pender Island, and green for Saturna Island).

Table 1. Inversion genotypes at polymorphic loci					
Locus	Population	Inv/Inv	Inv/Std	Std/Std	Inversion frequency
inv3.0	Saturna	8	10	2	0.650
	Pender	6	2	2	0.700
	Maple Ridge	7	9	1	0.676
inv7.2	Saturna	2	11	7	0.375
	Pender	3	4	3	0.500
	Maple Ridge	8	6	3	0.647
inv14.0	Saturna	6	8	6	0.500
	Pender	2	4	4	0.400
	Maple Ridge	0	8	9	0.235
inv15.0	Saturna	4	9	7	0.425
	Pender	1	6	3	0.400
	Maple Ridge	1	11	5	0.382
inv21.0	Saturna	11	7	2	0.725
	Pender	6	2	2	0.700
	Maple Ridge	12	4	1	0.824
inv22.0	Saturna	12	5	3	0.725
	Pender	3	6	1	0.600
	Maple Ridge	4	10	3	0.529

Table 1. All three Gulf Islands populations are polymorphic at the same inversion loci. Observed genotype frequencies at each polymorphic inversion locus. Columns three through five summarize the number of inversion homozygotes (Inv/Inv), heterozygotes (Inv/Std), and standard homozygotes (Std/Std) observed in each population sample at each locus. Individual genotypes at each inversion locus were obtained by partitioning SNP genotypes by minor allele frequency (see Figure 4). The orientation of each arrangement was determined based on alignments between the *P. maniculatus* reference assembly and those of *P. polionotus*, *P. leucopus*, *P. eremicus*, and *P. californicus*.

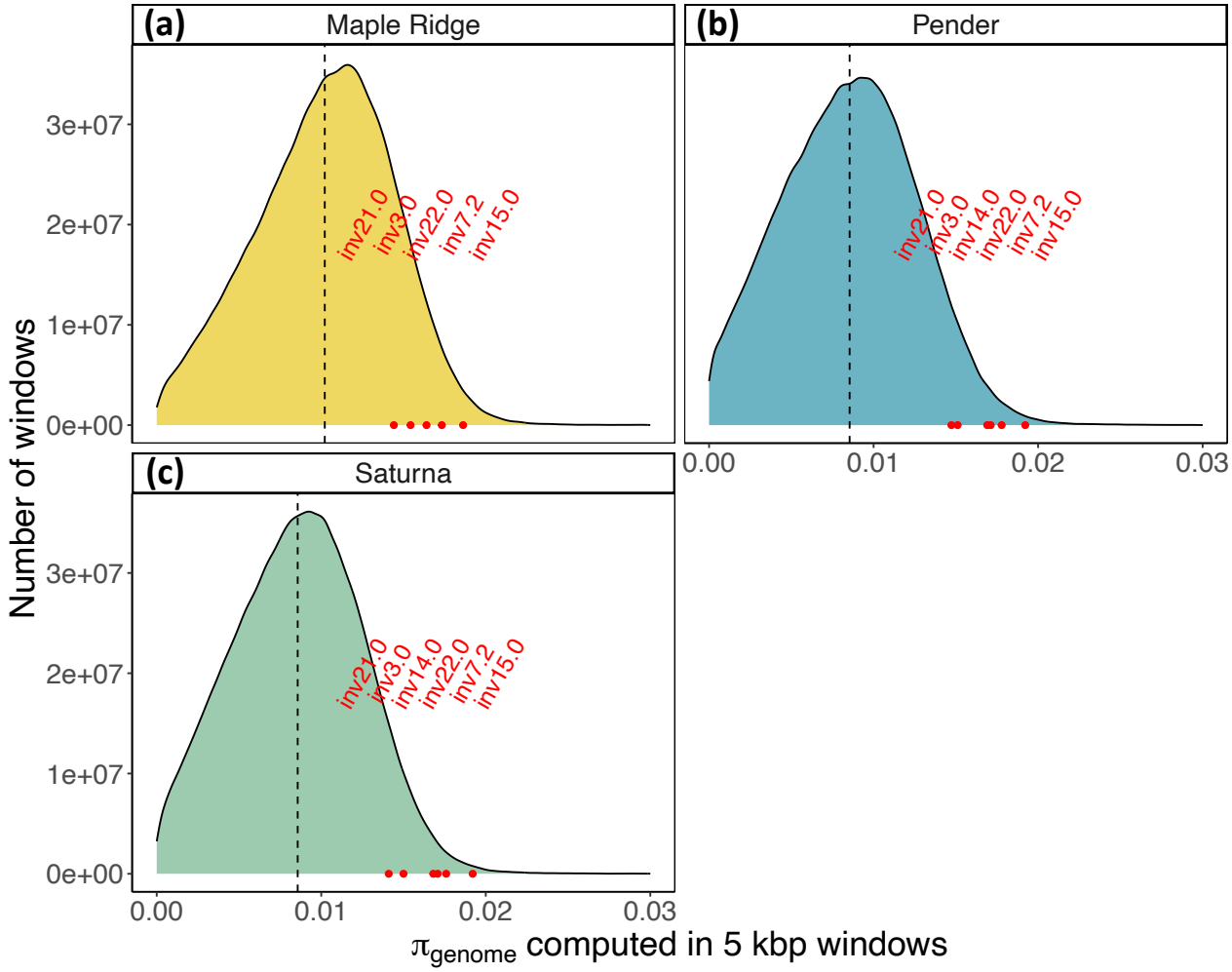

Figure 6. Divergence between arrangements at polymorphic inversions exceeds mean genome-wide diversity. Density plots illustrate the distribution of nucleotide diversity computed in genome-wide non-overlapping 5 kbp windows outside of known inversion loci (π_{genome}) for the Maple Ridge (a), Pender Island (b), and Saturna Island (c) samples. Dashed black lines denote the mean π_{genome} observed in each population sample. Red points demonstrate where mean divergence between arrangements ($d_{XY}^{subtype}$) falls along the distribution of genome-wide nucleotide diversity for each polymorphic inversion.

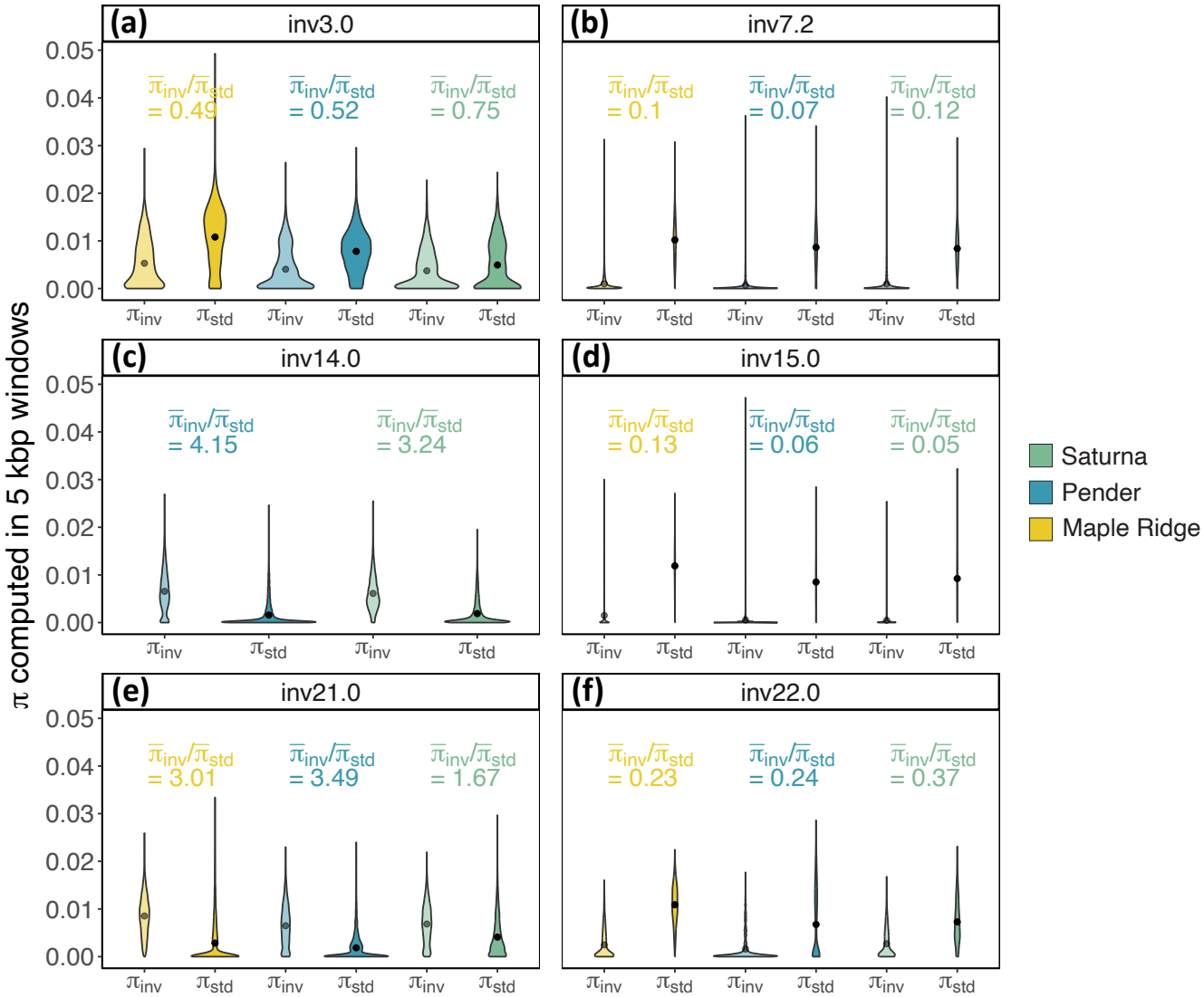

Figure 7. Comparisons of nucleotide diversity between inverted and standard arrangements. Y-axes plot nucleotide diversity measured in non-overlapping 5 kbp windows within either the inverted (π_{inv}) or standard arrangement (π_{std}) at six inversion loci: inv3.0 (a), inv7.2 (b), inv14.0 (c) inv15.0 (d), inv21.0 (e), and inv22.0 (f). Yellow, blue, and green violins represent the distributions of each diversity measure for the Maple Ridge, Pender Island, and Saturna Island samples, respectively. Opaque distributions represent nucleotide diversity within the standard arrangement (π_{std}) and transparent distributions represent nucleotide diversity within the inverted arrangement (π_{inv}). Black points denote the mean value observed across windows. Annotations provide the ratio of mean nucleotide diversity within the inverted arrangement to mean nucleotide diversity within the standard arrangement. Nucleotide diversity distributions for Maple Ridge are excluded for inv14.0 due to a lack of inversion homozygotes (see Table 1).

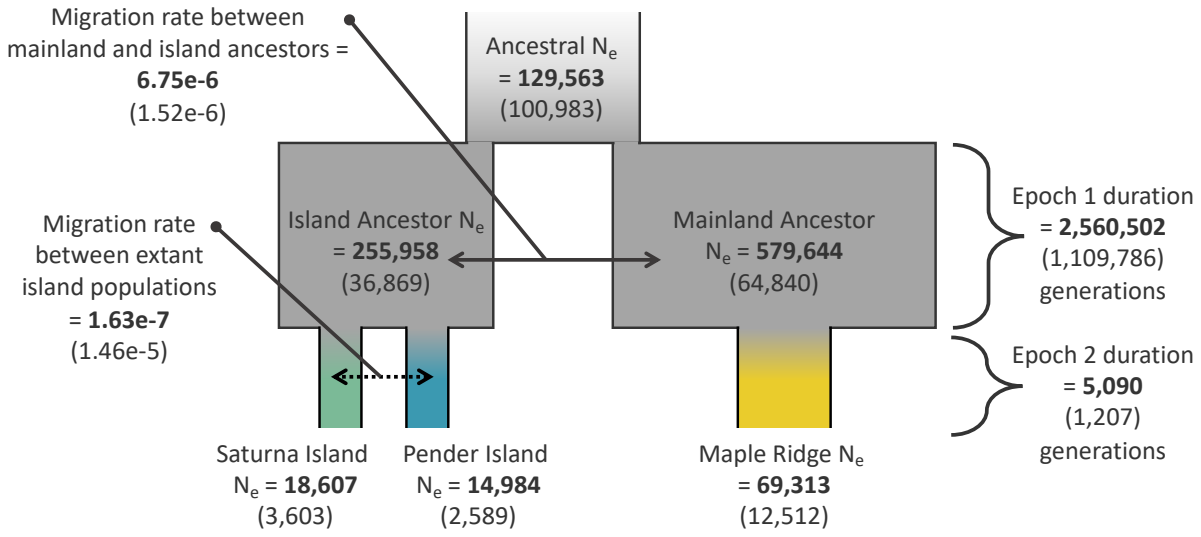

Figure 8. Parameter estimates for the candidate demographic model. Diagram reflects the structure of the best-fit three population demographic model. Boxes reflect population size changes through time (not drawn to scale). Gray shading reflects ancestral populations included in the model. Green, blue, and yellow shading reflects the contemporary Saturna Island, Pender Island, and mainland Maple Ridge populations. Arrows denote symmetric migration parameters included in the model. Maximum likelihood estimates for each parameter are bolded and standard deviations are given in parentheses. Standard deviations were obtained using the Godambe Information Matrix (Coffman et al. 2016) as an alternative to conventional bootstrapping.

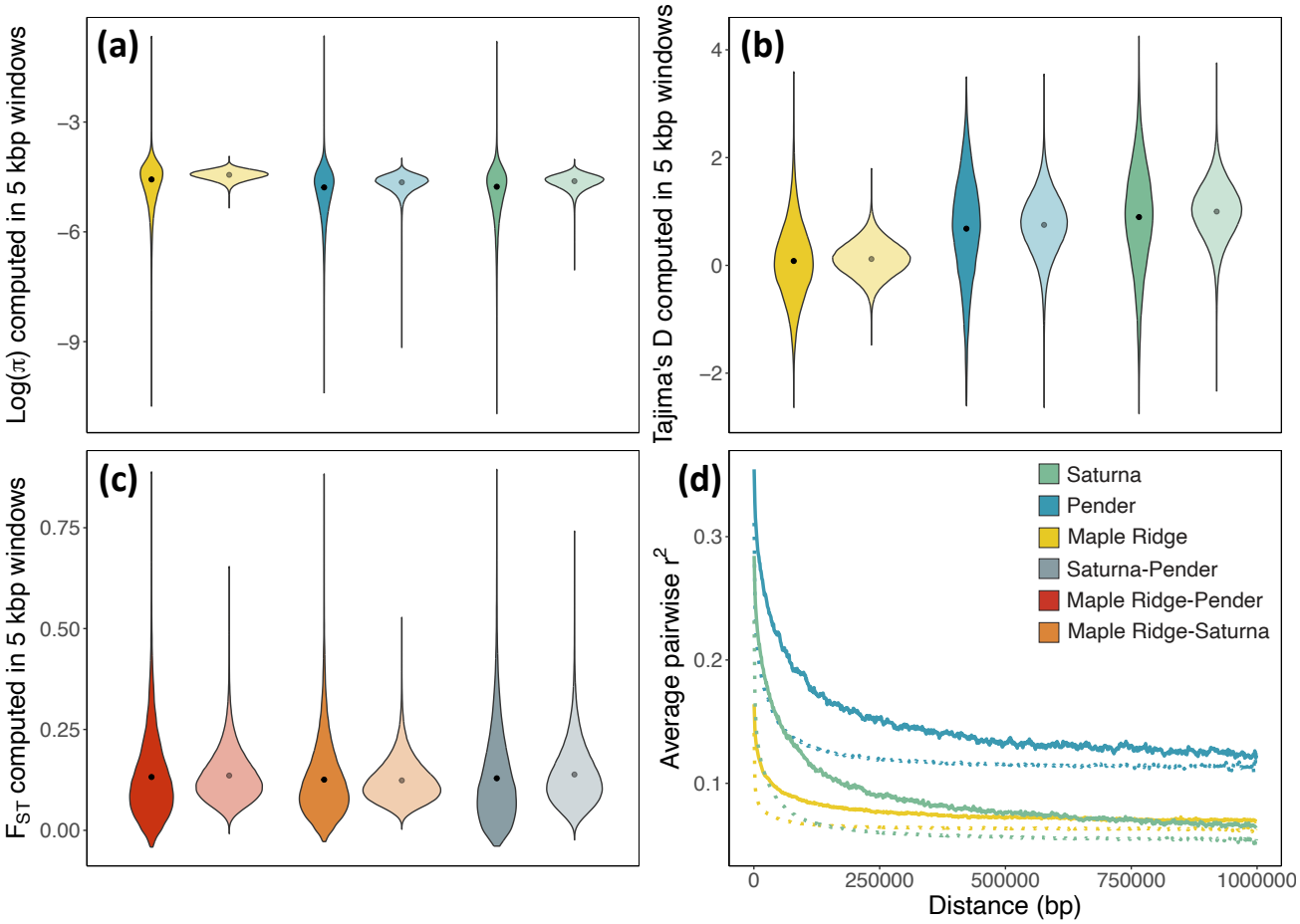

Figure 9. Inferred demographic model provides a close fit to diversity-based summaries but underpredicts levels of LD. Coalescent simulations conducted with the inferred demographic parameters were used to evaluate the fit of the model to empirical estimates of π, Tajima's D, F_{ST}, and LD. Violin plots compare simulated and empirical distributions of π (a), Tajima's D (b), and F_{ST} (c) computed in non-overlapping 5 kbp windows. The color of each violin represents the population (yellow for Maple Ridge, blue for Pender Island, and green for Saturna Island) or population pair (red for Maple Ridge-Pender, orange for Maple Ridge-Saturna, and gray for Saturna-Pender) for which the given summary statistic was measured. Opaque distributions represent empirical results, and transparent distributions represent simulated results. Black points denote the mean value of each distribution. Panel d compares simulated and observed LD decay. The y-axis plots average pairwise r^2 measured between SNPs separated by increasing distances (up to 1 Mbp; x-axis). Solid lines represent empirical results, and dashed lines represent simulated results. Curves are colored according to population (as in panels a-b).

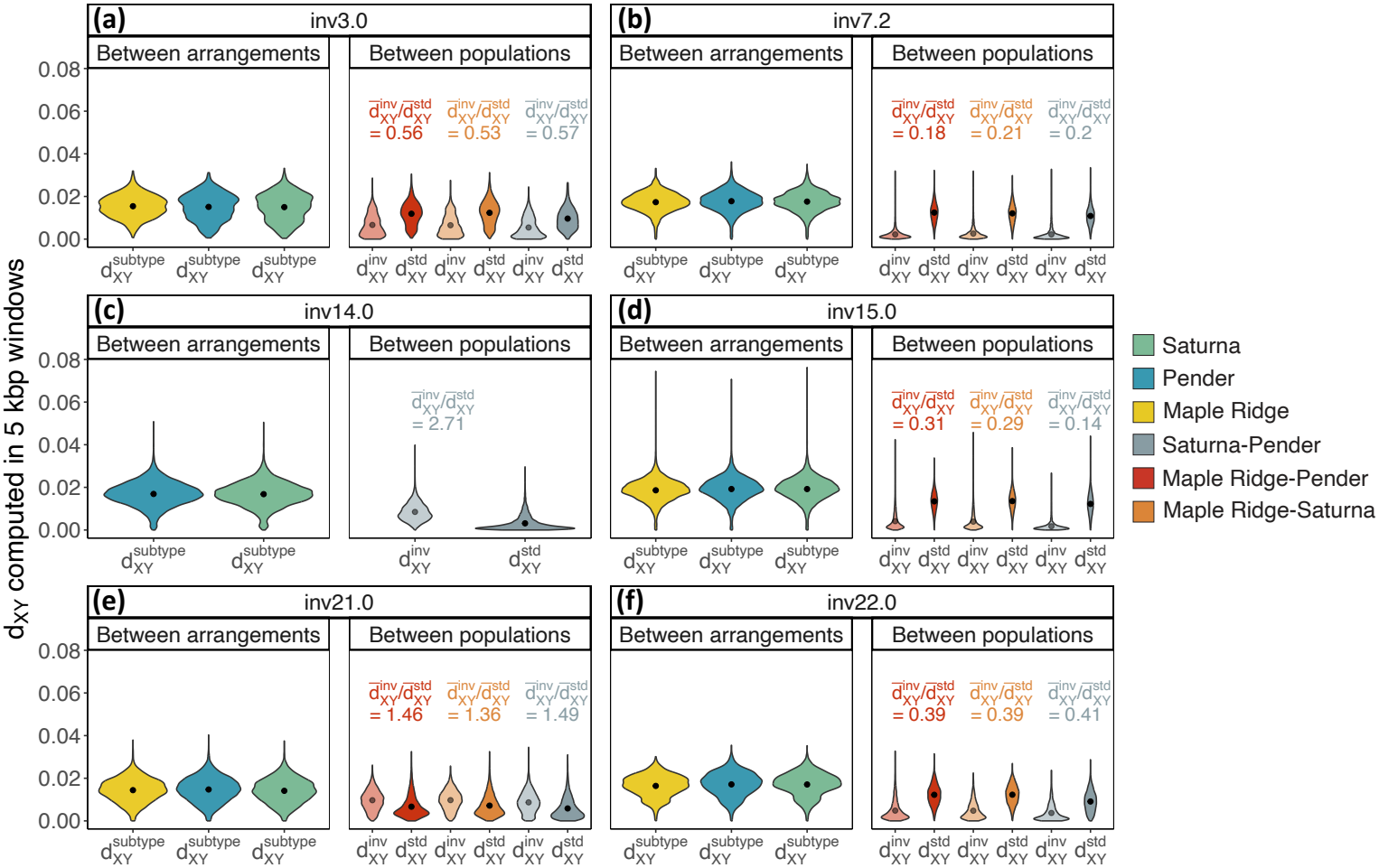

Figure 10. Divergence within and between arrangements provide insights about the emergence of inversion polymorphisms. Y-axes plot three measures of divergence computed in non-overlapping 5 kbp windows across six inversion loci: inv3.0 (a), inv7.2 (b), inv14.0 (c), inv15.0 (d), inv21.0 (e), and inv22.0 (f). Yellow, blue, and green violins plot the distribution of divergence between the inverted and standard arrangements ($d_{XY}^{subtype}$) sampled within each population (for Maple Ridge, Pender Island, and Saturna Island, respectively). Red, orange, and gray violins plot the distribution of divergence between populations for either the inverted (d_{XY}^{inv}) or standard (d_{XY}^{std}) arrangement. Maple Ridge-Pender comparisons are plotted in red, Maple Ridge-Saturna comparisons in orange, and Saturna-Pender comparisons in gray. Opaque distributions represent divergence between the standard arrangements (d_{XY}^{std}) and transparent distributions represent divergence between the inverted arrangements (d_{XY}^{inv}). Black points denote the mean value observed across windows. Annotations compare mean divergence between inverted arrangements to mean divergence between standard arrangements. Distributions of divergence within and between arrangements for Maple Ridge are excluded for inv14.0 due to a lack of inversion homozygotes (see Table 1).

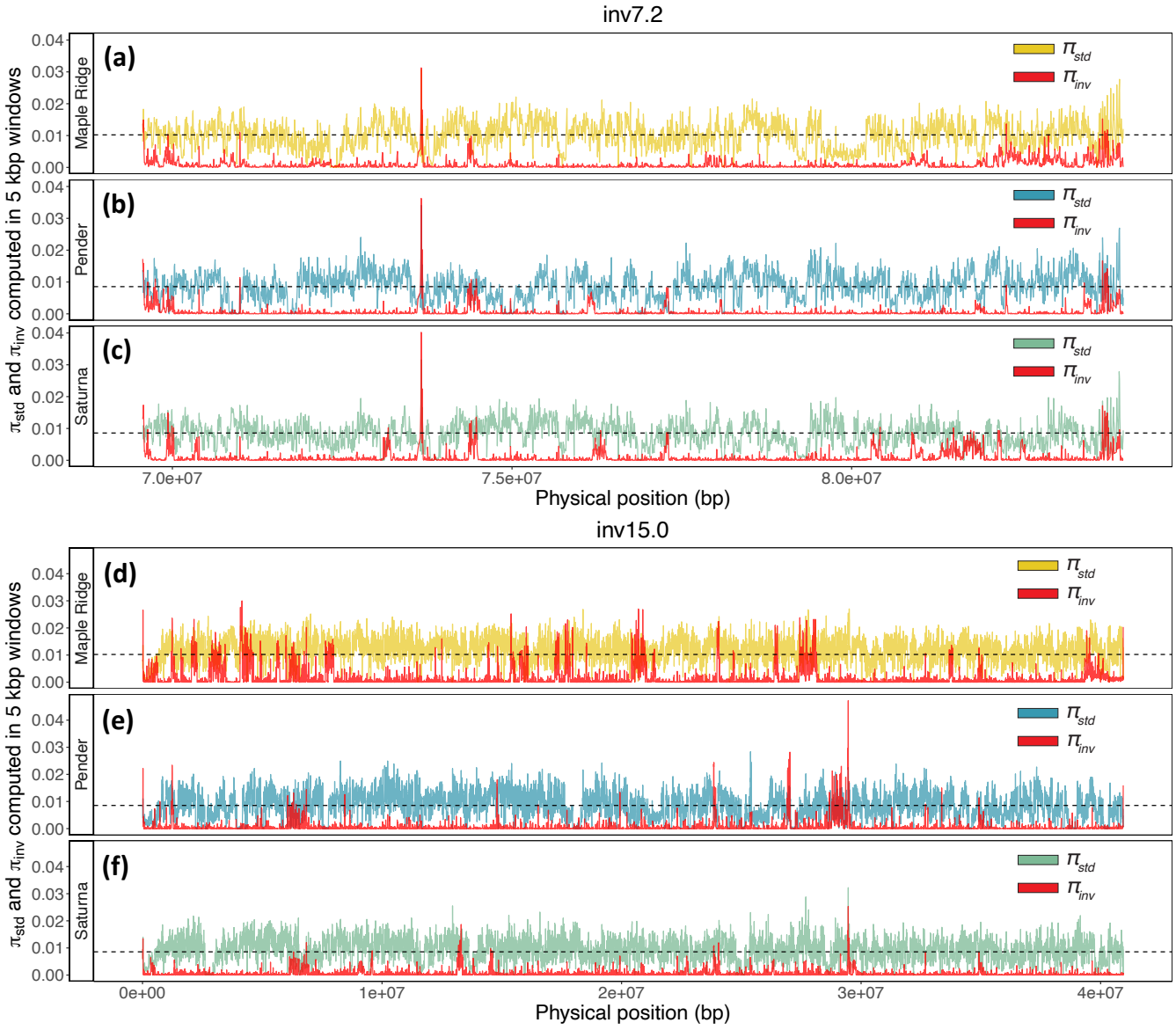

Figure 11. Inversion loci on chromosomes 7 and 15 show strong evidence for recent selection. Nucleotide diversity measured within arrangements for the Maple Ridge (a), Pender Island (b), and Saturna Island (c) samples along the inv7.2 locus. X-axes plot genomic position along the chromosome. Y-axes plot three measures of diversity computed in non-overlapping 5 kbp windows. Yellow, blue, and green lines represent nucleotide diversity within the subset of standard arrangements (π_{std}) sampled from Maple Ridge, Pender, and Saturna. Red lines represent nucleotide diversity within the corresponding sample of inverted arrangements (π_{inv}). Black dotted lines denote the mean genome-wide nucleotide diversity measured outside of known inversion loci (π_{genome}). Panels d, e, and f plot the same information for the inv15.0 locus.

ACKNOWLEDGEMENTS

We thank members of the Payseur lab for their helpful input on this work. This research was funded by National Institutes of Health (NIH) grants R01GM100426 and R35GM139412 (to B.A.P.). E.K.H. was partially supported by the NIH Graduate Training Grant in Genetics at the University of Wisconsin-Madison (T32GM007133). Computational analyses were conducted using resources provided by the University of Wisconsin-Madison's Center for High Throughput Computing. Fieldwork was supported by a Putnam Expedition Grant of the Harvard Museum of Comparative Zoology. F.B. was supported by an HHMI International Student Research Fellowship, a Joan Brockman Williamson Graduate Research Fellowship, and a Herchel Smith Graduate Fellowship. This work was supported by the Howard Hughes Medical Institute, of which H.E.H. was an Investigator.

DATA ACCESSIBILITY

Upon submission, raw sequencing reads will be deposited in the SRA. Code used to conduct the analyses is available on GitHub (https://github.com/PayseurLabUWMadison/gi_demography_inversions).

AUTHOR CONTRIBUTIONS

E.K.H. and B.A.P. designed the study. F.B. and H.E.H acquired the samples used for the study. E.K.H. conducted the research with supervision from B.A.P. E.K.H. and B.A.P. wrote the manuscript with input from F.B. and H.E.H.

REFERENCES

Adler, G. H., & Levins, R. (1994). The Island Syndrome in Rodent Populations. *The Quarterly Review of Biology*, *69*(4), 473–490.

Adler, G. H., & Tamarin, R. H. (1984). Demography and reproduction in island and mainland white-footed mice (Peromyscus leucopus) in southeastern Massachusetts. *Canadian Journal of Zoology*, *62*(1), 58-64.

Adler, G. H., Wilson, M. L., & Derosa, M. J. (1986). Influence of island area and isolation on population characteristics of Peromyscus leucopus. *Journal of Mammalogy*, *67*(2), 406-409.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, *19*(9), 1655-1664.

Allard, M. W., & Greenbaum, I. F. (1988). Morphological variation and taxonomy of chromosomally differentiated Peromyscus from the Pacific Northwest. *Canadian Journal of Zoology*, *66*(12), 2734–2739. https://doi.org/10.1139/z88-400

Álvarez-Castañecla, S. T., Cortés-Calva, P., González-Cózatl, F. X., Rojas, D., & Leyva, I. (2010). Comparison of distribution and habitat characteristics between an endemic and a wide-ranging cryptic species of Peromyscus on the Baja California peninsula. *Western North American Naturalist*, *70*(3), 323-333.

Baier, F., & Hoekstra, H. E. (2019). The genetics of morphological and behavioural island traits in deer mice. *Proceedings of the Royal Society B*, *286*(1914), 20191697.

Baumdicker, F., Bisschop, G., Goldstein, D., Gower, G., Ragsdale, A. P., Tsambos, G., ... & Kelleher, J. (2022). Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, *220*(3), iyab229.

Bedford, N. L., & Hoekstra, H. E. (2015). Peromyscus mice as a model for studying natural variation. *eLife*, *4*, e06813. https://doi.org/10.7554/eLife.06813

Beichman, A. C., Huerta-Sanchez, E., & Lohmueller, K. E. (2018). Using Genomic Data to Infer Historic Population Dynamics of Nonmodel Organisms. *Annual Review of Ecology, Evolution, and Systematics*, *49*(1), 433–456. https://doi.org/10.1146/annurev-ecolsys-110617-062431

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research*, *27*(2), 573-580.

Berdan, E. L., Barton, N. H., Butlin, R., Charlesworth, B., Faria, R., Fragata, I., ... & Flatt, T. (2023). How chromosomal inversions reorient the evolutionary process. *Journal of Evolutionary Biology*, *36*(12), 1761-1782.

Bitarello, B. D., Brandt, D. Y. C., Meyer, D., & Andrés, A. M. (2023). Inferring Balancing Selection From Genome-Scale Data. *Genome Biology and Evolution*, *15*(3), evad032. https://doi.org/10.1093/gbe/evad032

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.

Bradley, R. D., Durish, N. D., Rogers, D. S., Miller, J. R., Engstrom, M. D., & Kilpatrick, C. W. (2007). Toward a Molecular Phylogeny for Peromyscus: Evidence from Mitochondrial Cytochrome-b Sequences. *Journal of Mammalogy*, *88*(5), 1146–1159. https://doi.org/10.1644/06-MAMM-A-342R.1

Brown, J., Crivello, J., & O'Neill, R. J. (2018). An updated genetic map of Peromyscus with chromosomal assignment of linkage groups. *Mammalian Genome*, *29*(5–6), 344–352. https://doi.org/10.1007/s00335-018-9754-7

Burke, J. M., Lai, Z., Salmaso, M., Nakazato, T., Tang, S., Heesacker, A., Knapp, S. J., & Rieseberg, L. H. (2004). Comparative Mapping and Rapid Karyotypic Evolution in the Genus Helianthus. *Genetics*, *167*(1), 449–457. https://doi.org/10.1534/genetics.167.1.449

Bush, G. L., Case, S. M., Wilson, A. C., & Patton, J. L. (1977). Rapid speciation and chromosomal evolution in mammals. *Proceedings of the National Academy of Sciences*, *74*(9), 3942–3946. https://doi.org/10.1073/pnas.74.9.3942

Caceres, M., Ranz, J. M., Barbadilla, A., Long, M., & Ruiz, A. (1999). Generation of a widespread Drosophila inversion by a transposable element. *Science*, *285*(5426), 415-418.

Carbone, L., Alan Harris, R., Gnerre, S., Veeramah, K. R., Lorente-Galdos, B., Huddleston, J., Meyer, T. J., Herrero, J., Roos, C., Aken, B., Anaclerio, F., Archidiacono, N., Baker, C., Barrell, D., Batzer, M. A., Beal, K., Blancher, A., Bohrson, C. L., Brameier, M., … Gibbs, R. A. (2014). Gibbon genome and the fast karyotype evolution of small apes. *Nature*, *513*(7517), 195–201. https://doi.org/10.1038/nature13679

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), s13742-015.

Charlesworth, B. (2023). The effects of inversion polymorphisms on patterns of neutral genetic diversity. *Genetics*, *224*(4), iyad116.

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, *70*(2), 155–174. https://doi.org/10.1017/S0016672397002954

Clague, J. J. (1980). *Late Quaternary geology and geochronology of British Columbia* (No. 13). Energy, Mines and Resources Canada.

Clague, J. J., & James, T. S. (2002). History and isostatic effects of the last ice sheet in southern British Columbia. *Quaternary Science Reviews*, *21*(1–3), 71–87. https://doi.org/10.1016/S0277-3791(01)00070-1

Coffman, A. J., Hsieh, P. H., Gravel, S., & Gutenkunst, R. N. (2016). Computationally Efficient Composite Likelihood Statistics for Demographic Inference. *Molecular Biology and Evolution*, *33*(2), 591–593. https://doi.org/10.1093/molbev/msv255

Corbett-Detig, R. B., & Hartl, D. L. (2012). Population Genomics of Inversion Polymorphisms in Drosophila melanogaster. *PLOS Genetics*, *8*(12), e1003056. https://doi.org/10.1371/journal.pgen.1003056

Corbett-Detig, R. B., Said, I., Calzetta, M., Genetti, M., McBroome, J., Maurer, N. W., ... & Besansky, N. J. (2019). Fine-mapping complex inversion breakpoints and investigating somatic pairing in the Anopheles gambiae species complex using proximity-ligation sequencing. *Genetics*, *213*(4), 1495-1511.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., ... & Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2), giab008.

Dobzhansky, T. (1948). Genetics of natural populations. XVIII. Experiments on chromosomes of Drosophila pseudoobscura from different geographic regions. *Genetics*, *33*(6), 588.

Dobzhansky, T. (1949). Observations and experiments on natural selection in Drosophila. *Hereditas*, *35*(S1), 210-224.

Dobzhansky, T. (1950). Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of Drosophila pseudoobscura. *Genetics*, *35*(3), 288.

Eis, S., & Craigdallie, D. (1980). Gulf Islands of British Columbia. A landscape analysis.

Fedje, D. W., Sumpter, I. D., & Southon, J. R. (2009). Sea-levels and Archaeology in the Gulf Islands National Park Reserve. *Canadian Journal of Archaeology / Journal Canadien d'Archéologie*, *33*(2), 234–253.

Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes: limits and prospects. *Molecular ecology*, *24*(14), 3529-3545.

Gill, A. E. (1980). Evolutionary genetics of California islands Peromyscus. In *The California Islands: proceedings of a multidisciplinary symposium* (pp. 719-743). Santa Barbara, CA: Santa Barbara Museum of Natural History.

Goel, M., & Schneeberger, K. (2022). plotsr: Visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics*, *38*(10), 2922–2926. https://doi.org/10.1093/bioinformatics/btac196

Goel, M., Sun, H., Jiao, W.-B., & Schneeberger, K. (2019). SyRI: Finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology*, *20*(1), 277. https://doi.org/10.1186/s13059-019-1911-0

Gowan, E. J. (2007). *Glacio-isostatic adjustment modelling of improved relative sea-level observations in southwestern British Columbia, Canada* (Doctoral dissertation).

Gozashti, L., Feschotte, C., & Hoekstra, H. E. (2023). Transposable Element Interactions Shape the Ecology of the Deer Mouse Genome. *Molecular Biology and Evolution*, *40*(4), msad069. https://doi.org/10.1093/molbev/msad069

Greenbaum, I. F., Gunn, S. J., Smith, S. A., McAllister, B. F., Hale, D. W., Baker, R. J., Engstrom, M. D., Hamilton, M. J., Modi, W. S., Robbins, L. W., Rogers, D. S., Ward, O. G., Dawson, W. D., Elder, F. F. B., Lee, M. R., Pathak, S., & Stangl, Jr., F. B. (1994). Cytogenetic nomenclature of deer mice, *Peromyscus* (Rodentia): Revision and review of the standardized karyotype. *Cytogenetic and Genome Research*, *66*(3), 181–195. https://doi.org/10.1159/000133696

Guerrero, R. F., Rousset, F., & Kirkpatrick, M. (2012). Coalescent patterns for chromosomal inversions in divergent populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 430–438. https://doi.org/10.1098/rstb.2011.0246

Gunn, S. J. (1988). Chromosomal variation and differentiation among insular populations of Peromyscus from the Pacific Northwest. *Canadian Journal of Zoology*, *66*(12), 2726–2733. https://doi.org/10.1139/z88-399

Gunn, S. J., & Greenbaum, I. F. (1986). Systematic Implications of Karyotypic and Morphologic Variation in Mainland Peromyscus from the Pacific Northwest. *Journal of Mammalogy*, *67*(2), 294–304. https://doi.org/10.2307/1380882

Hager, E. R., Harringmeyer, O. S., Wooldridge, T. B., Theingi, S., Gable, J. T., McFadden, S., Neugeboren, B., Turner, K. M., Jensen, J. D., & Hoekstra, H. E. (2022). A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science*, *377*(6604), 399–405. https://doi.org/10.1126/science.abg0718

Halpin, Z. T., & Sullivan, T. P. (1978). Social Interactions in Island and Mainland Populations of the Deer Mouse, Peromyscus maniculatus. *Journal of Mammalogy*, *59*(2), 395–401. https://doi.org/10.2307/1379924

Harringmeyer, O. S., & Hoekstra, H. E. (2022). Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nature Ecology & Evolution*, *6*(12), Article 12. https://doi.org/10.1038/s41559-022-01890-0

Hasson, E., & Eanes, W. F. (1996). Contrasting Histories of Three Gene Regions Associated With *In(3L)Payne* of *Drosophila melanogaster*. *Genetics*, *144*(4), 1565–1575. https://doi.org/10.1093/genetics/144.4.1565

Hedrick, P. W. (1981). The Establishment of Chromosomal Variants. *Evolution*, *35*(2), 322–332. https://doi.org/10.2307/2407841

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*(6789), 907–913. https://doi.org/10.1038/35016000

Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual Review of Ecology, Evolution, and Systematics*, *39*(1), 21–42. https://doi.org/10.1146/annurev.ecolsys.39.110707.173532

Hogan, K. M., Hedin, M. C., Koh, H. S., Davis, S. K., & Greenbaum, I. F. (1993). Systematic and Taxonomic Implications of Karyotypic, Electrophoretic, and Mitochondrial-DNA Variation in Peromyscus from the Pacific Northwest. *Journal of Mammalogy*, *74*(4), 819–831. https://doi.org/10.2307/1382420

Howell, E. K., Nolfo-Clements, L. E., & Payseur, B. A. (2025). Population history across timescales in an urban archipelago. *Genome Biology and Evolution*, evaf048.

Hudson, R. R. (1990). Gene genealogies and the coalescent process. *Oxford surveys in evolutionary biology*, *7*(1), 44.

Hudson, R. R., Slatkin, M., & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, *132*(2), 583-589.

Hutchinson, I., James, T. S., Clague, J. J., Barrie, J. V., & Conway, K. W. (2004). Reconstruction of late Quaternary sea-level change in southwestern British Columbia from sediments in isolation basins. *Boreas*, *33*(3), 183–194. https://doi.org/10.1111/j.1502-3885.2004.tb01140.x

James, T., Gowan, E. J., Hutchinson, I., Clague, J. J., Barrie, J. V., & Conway, K. W. (2009). Sea-level change and paleogeographic reconstructions, southern Vancouver Island, British Columbia, Canada. *Quaternary Science Reviews*, *28*(13), 1200–1216. https://doi.org/10.1016/j.quascirev.2008.12.022

Jouganous, J., Long, W., Ragsdale, A. P., & Gravel, S. (2017a). Inferring the Joint Demographic History of Multiple Populations: Beyond the Diffusion Approximation. *Genetics*, *206*(3), 1549–1567. https://doi.org/10.1534/genetics.117.200493

Kapun, M., Mitchell, E. D., Kawecki, T. J., Schmidt, P., & Flatt, T. (2023). An ancestral balanced inversion polymorphism confers global adaptation. *Molecular biology and evolution*, *40*(6), msad118.

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic acids research*, *32*(suppl_1), D493-D496.

Katz, L. S., Griswold, T., Morrison, S. S., Caravas, J. A., Zhang, S., den Bakker, H. C., Deng, X., & Carleton, H. A. (2019). Mashtree: A rapid comparison of whole genome sequence files. *Journal of Open Source Software*, *4*(44), 10.21105/joss.01762. https://doi.org/10.21105/joss.01762

Kenney-Hunt, J., Lewandowski, A., Glenn, T. C., Glenn, J. L., Tsyusko, O. V., O'Neill, R. J., Brown, J., Ramsdell, C. M., Nguyen, Q., Phan, T., Shorter, K. R., Dewey, M. J., Szalai, G., Vrana, P. B., & Felder, M. R. (2014). A genetic map of Peromyscus with chromosomal assignment of linkage groups (a Peromyscus genetic map). *Mammalian Genome*, *25*(3–4), 160–179. https://doi.org/10.1007/s00335-014-9500-8

Kirkpatrick, M., & Barton, N. (2006). Chromosome Inversions, Local Adaptation and Speciation. *Genetics*, *173*(1), 419–434. https://doi.org/10.1534/genetics.105.047985

Ku, H. H. (1966). Notes on the use of propagation of error formulas. *Journal of Research of the National Bureau of Standards*, *70*(4).

Lande, R. (1979). Effective Deme Sizes During Long-Term Evolution Estimated from Rates of Chromosomal Rearrangement. *Evolution*, *33*(1), 234–251. https://doi.org/10.2307/2407380

Lessa, E. P., Cook, J. A., & Patton, J. L. (2003). Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. *Proceedings of the National Academy of Sciences*, *100*(18), 10331–10334. https://doi.org/10.1073/pnas.1730921100

Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, *30*(20), 2843-2851.

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

Long, A. D., Baldwin-Brown, J., Tao, Y., Cook, V. J., Balderrama-Gutierrez, G., Corbett-Detig, R., ... & Barbour, A. G. (2019). The genome of Peromyscus leucopus, natural host for Lyme disease and other emerging infections. *Science advances*, *5*(7), eaaw6441.

Losos, J. B., & Ricklefs, R. E. (2009). Adaptation and diversification on islands. *Nature*, *457*(7231), 830–836. https://doi.org/10.1038/nature07893

Lucid, M. K., & Cook, J. A. (2004). Phylogeography of Keen's mouse (Peromyscus keeni) in a naturally fragmented landscape. *Journal of Mammalogy*, *85*(6), 1149-1159.

Mackie, Q., Fedje, D., & McLaren, D. (2018). Archaeology and Sea Level Change on the British Columbia Coast. *Canadian Journal of Archaeology / Journal Canadien d'Archéologie*, *42*(1), 74–91.

Mackie, Q., Fedje, D., McLaren, D., Smith, N., & McKechnie, I. (2011). Early Environments and Archaeology of Coastal British Columbia. In N. F. Bicho, J. A. Haws, & L. G. Davis (Eds.), *Trekking the Shore* (pp. 51–103). Springer New York. https://doi.org/10.1007/978-1-4419-8219-3_3

Maggiolini, F. A. M., Sanders, A. D., Shew, C. J., Sulovari, A., Mao, Y., Puig, M., ... & Antonacci, F. (2020). Single-cell strand sequencing of a macaque genome reveals multiple nested inversions and breakpoint reuse during primate evolution. *Genome research*, *30*(11), 1680-1693.

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867-2873.

Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLOS Computational Biology*, *14*(1), e1005944. https://doi.org/10.1371/journal.pcbi.1005944

McClintock, B. (1931). *Cytological observations of deficiencies involving known genes, translocations and an inversion in Zea mays*. University of Missouri, College of Agriculture, Agricultural Experiment Station.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, *20*(9), 1297-1303.

Morgulis, A., Gertz, E. M., Schäffer, A. A., & Agarwala, R. (2006). WindowMasker: window-based masker for sequenced genomes. *Bioinformatics*, *22*(2), 134-141.

Nadachowska-Brzyska, K., Konczal, M., & Babik, W. (2022). Navigating the temporal continuum of effective population size. *Methods in Ecology and Evolution*, *13*(1), 22–41. https://doi.org/10.1111/2041-210X.13740

Navarro, A., Barbadilla, A., & Ruiz, A. (2000). Effect of Inversion Polymorphism on the Neutral Nucleotide Variability of Linked Chromosomal Regions in Drosophila. *Genetics*, *155*(2), 685–698. https://doi.org/10.1093/genetics/155.2.685

Navarro, A., Betrán, E., Barbadilla, A., & Ruiz, A. (1997). Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics*, *146*(2), 695–709. https://doi.org/10.1093/genetics/146.2.695

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, *76*(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Nolfo-Clements, L., Butcher, R., Leite, M., & Clements, M. (2017). Evidence of the island rule and microevolution in white-footed mice (Peromyscus leucopus) in an urban harbor archipelago. *Mammal Research*, *62*, 423-430.

Ohno, S., Weiler, C., Poole, J., Christian, L., & Stenius, C. (1966). Autosomal polymorphism due to pericentric inversions in the deer mouse (peromyscus maniculatus), and some evidence of somatic segregation. *Chromosoma*, *18*(2), 177–187. https://doi.org/10.1007/BF00326866

Osgood, W. H. (1909). *Revision of the mice of the American genus Peromyscus* (No. 28). US Government Printing Office.

Pevzner, P., & Tesler, G. (2003). Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proceedings of the National Academy of Sciences*, *100*(13), 7672-7677.

Pfeifer, S. P. (2017). From next-generation resequencing reads to a high-quality variant data set. *Heredity*, *118*(2), 111-124.

Platt, R. N., Amman, B. R., Keith, M. S., Thompson, C. W., & Bradley, R. D. (2015). What Is *Peromyscus* ? Evidence from nuclear and mitochondrial DNA sequences suggests the need for a new classification. *Journal of Mammalogy*, *96*(4), 708–719. https://doi.org/10.1093/jmammal/gyv067

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., ... & Banks, E. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178.

Redfield, J. A. (1976). Distribution, abundance, size, and genetic variation of *Peromyscus maniculatus* on the Gulf Islands of British Columbia. *Canadian Journal of Zoology*, *54*(4), 463–474. https://doi.org/10.1139/z76-053

Richburg, J. A., & Patterson III, W. A. (2005). Historical Description of the Vegetation of the Boston Harbor Islands: 1600–2000. *Northeastern Naturalist*, *12*(sp3), 13–30. https://doi.org/10.1656/1092-6194(2005)12[13:HDOTVO]2.0.CO;2

Robbins, L. W., & Baker, R. J. (1981). An assessment of the nature of chromosomal rearrangements in 18 species of *Peromyscus* (Rodentia: Cricetidae). *Cytogenetic and Genome Research*, *31*(4), 194–202. https://doi.org/10.1159/000131649

Said, I., Byrne, A., Serrano, V., Cardeno, C., Vollmers, C., & Corbett-Detig, R. (2018). Linked genetic variation and not genome structure causes widespread differential expression associated with chromosomal inversions. *Proceedings of the National Academy of Sciences*, *115*(21), 5492–5497. https://doi.org/10.1073/pnas.1721275115

Santiago, E., Caballero, A., Köpke, C., & Novo, I. (2024). Estimation of the contemporary effective population size from SNP data while accounting for mating structure. *Molecular Ecology Resources*, *24*(1), e13890. https://doi.org/10.1111/1755-0998.13890

Sawyer, Y. E., Flamme, M. J., Jung, T. S., MacDonald, S. O., & Cook, J. A. (2017). Diversification of deermice (Rodentia: Genus Peromyscus) at their north-western range limit: genetic consequences of refugial and island isolation. *Journal of Biogeography*, *44*(7), 1572–1585. https://doi.org/10.1111/jbi.12995

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE*, *11*(10), e0163962. https://doi.org/10.1371/journal.pone.0163962

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, *23*(1), 23-35.

Stangl, F. B., & Baker, R. J. (1984). Evolutionary Relationships in Peromyscus: Congruence in Chromosomal, Genic, and Classical Data Sets. *Journal of Mammalogy*, *65*(4), 643–654. https://doi.org/10.2307/1380846

Steffen, M. L. (2016). Body-size trends in Peromyscus (Rodentia: Cricetidae) on Vancouver Island, Canada, with comments on relictual gigantism. *Paleobiology*, *42*(3), 532–546. https://doi.org/10.1017/pab.2016.14

Sturtevant, A. H. (1921). A case of rearrangement of genes in Drosophila. *Proceedings of the National Academy of Sciences*, *7*(8), 235-237.

Sullivan, T. P. (1977). Demography and Dispersal in Island and Mainland Populations of the Deer Mouse, Peromyscus Maniculatus. *Ecology*, *58*(5), 964–978. https://doi.org/10.2307/1936918

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*(2), 437-460.

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

Takahata, N. (1990). A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proceedings of the National Academy of Sciences*, *87*(7), 2419-2423.

Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the cloud: using Docker, GATK, and WDL in Terra*. O'Reilly Media.

Wang, J. (2017). Effects of sampling close relatives on some elementary population genetics analyses. *Molecular Ecology Resources*, *18*(1), 41-54.

Waples, R. S., & Anderson, E. C. (2017). Purging putative siblings from population genetic data sets: a cautionary view.

Wellenreuther, M., & Bernatchez, L. (2018). Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends in Ecology & Evolution*, *33*(6), 427–440. https://doi.org/10.1016/j.tree.2018.04.002

Wickham, H. (2016). *Ggplot2: Elegrant graphics for data analysis*. Springer.

Wigginton, J. E., Cutler, D. J., & Abecasis, G. R. (2005). A Note on Exact Tests of Hardy-Weinberg Equilibrium. *The American Journal of Human Genetics*, *76*(5), 887–893. https://doi.org/10.1086/429864

Wolff, J. O. (1985). Comparative population ecology of Peromyscus leucopus and Peromyscus maniculatus. *Canadian Journal of Zoology*, *63*(7), 1548-1555.

Yeaman, S., & Otto, S. P. (2011). Establishment and Maintenance of Adaptive Genetic Divergence Under Migration, Selection, and Drift. *Evolution*, *65*(7), 2123–2129. https://doi.org/10.1111/j.1558-5646.2011.01277.x

Yeaman, S., & Whitlock, M. C. (2011). The Genetic Architecture of Adaptation Under Migration–Selection Balance. *Evolution*, *65*(7), 1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x

1229 Zheng, X., Arbogast, B. S., & Kenagy, G. J. (2003). Historical demography and genetic structure of
1230 sister species: Deermice (Peromyscus) in the North American temperate rain forest.
1231 *Molecular Ecology*, *12*(3), 711–724. https://doi.org/10.1046/j.1365-294X.2003.01770.x