

Background selection under evolving recombination rates

Tom R. Booker*

*Department of Zoology, University of British Columbia

ABSTRACT

Background selection (BGS), the effect that purifying selection exerts on evolution at linked sites, is expected to be ubiquitous across eukaryotic genomes. BGS effects reflect the interplay of fitness effects and rates of deleterious mutations with recombination. Leveraging the theory of BGS to analyse patterns of nucleotide diversity has shed light on central issues in evolutionary biology. Fundamental to theoretical models of BGS are recombination rate estimates and an assumption that recombination rates are invariant over time. However, in some lineages recombination rates evolve very rapidly violating this central assumption. Here, we investigate the effect that recombination rate evolution can have on BGS. We show that recombination rate evolution may have localised effects and cause analyses to underestimate the effects genome-wide effects of BGS. Indeed, we find evidence that rapid recombination rate evolution in the recent history of the house mouse may impact inferences of selection in that species.

KEYWORDS Evolutionary genetics, recombination rate, background selection

Introduction

Different modes of selection (e.g. positive, purifying and balancing) can all affect variation at linked sites (reviewed in Charlesworth 2008). In the case of purifying selection, the removal of deleterious mutations can cause linked neutral variants to be lost along with them through a process referred to as background selection (BGS; [Charlesworth *et al.* 1993](#)). Of the mutations that affect fitness, the vast majority are likely deleterious with a comparatively small proportion of beneficial mutations ([Charlesworth *et al.* 1993](#)). For those reasons, it has been proposed that BGS is likely ubiquitous across eukaryotic genomes and should be incorporated into the null model explaining variation in nucleotide diversity (π) ([Comeron 2017](#); [Johri *et al.* 2020](#)). However, interpreting genome-wide patterns of genetic variability in terms of BGS requires accurate estimates of population genetic parameters, particularly recombination rates.

Recombination rates vary across the genome. Estimates of the local recombination rate (r) can be obtained empirically by examining the inheritance of genetic markers through known pedigrees, as in traditional genetic mapping, or by directly comparing an individual's genome to that of its gametes (e.g. [Sun *et al.* 2019](#)). Both methods directly observe recombination events over one or fairly small number of generations, and thus provide estimates of r for contemporary populations. Alternatively, estimates of r can be obtained indirectly by analysing linkage disequilibrium across the genome (e.g. [Spence and Song 2019](#)) and such estimates reflect both recent and ancestral recombination events. The use of empirical or indirect estimates of recombination rate when analysing genome-wide variation in π in terms of BGS implicitly assumes that the recombination landscape has not changed over the time in which patterns of diversity have been established. However, temporal variation in the recombination rate may influence our ability to explain population genomic data using BGS as our null model as described by [Comeron \(2017\)](#).

Recombination rate changes may influence fine-scale patterns of molecular evolution. For example, chromosomal fusions would decrease recombination rates experienced by individual nucleotides, and thus increase the effects of BGS and other processes mediated by recombination. Consistent with this, [Cicconardi *et al.* \(2021\)](#) found evidence suggesting that chromosomes that underwent fusions in the ancestors of extant *Heliconius* butterfly species now harbour lower π due to reduced recombination rates and presumably greater BGS effects. Following evolution of the recombination rate landscape, there will be a lag period wherein patterns of genetic variability more closely reflect ancestral recombination rates than derived rates. Depending on how recombination rate landscapes evolve, population genomic analysis of lineages that are still within that lag period may be obscured. In this letter, we examine how patterns of neutral genetic variability under BGS respond to evolution of the recombination rate and describe how this may have affected analyses that are used to identify the effects of selection genome wide.

Results and Discussion

Background selection under evolving recombination rates

A change in the recombination rate will alter the effects of BGS. Under strong purifying selection, BGS reduces coalescence times and resembles a localised reduction in the effective population size (N_e). Changes in the recombination rate for a particular genetic region may therefore resemble a localised change in N_e . For a neutral site ν , the combined effects of recombination, mutation and purifying selection cause there to be a reduction of $B_{\nu,0}$ to coalescence times. At time T_0 in the past (in $2N_e$ generations), the population underwent an instantaneous change in the recombination rate so ν now experiences a BGS effect of $B_{\nu,1}$. Equation 1a from [Johri *et al.* \(2020\)](#) describes coalescence times after instantaneous population size change. We modified their expression, incorporating the effects of BGS before and after a change in the recombination rate:

$$B_{\nu,\Delta r} = B_{\nu,1} \left(1 + \left(\frac{B_{\nu,0}}{B_{\nu,1}} - 1 \right) e^{-T_0} \right) \quad (1)$$

Note that [Pool and Nielsen \(2009\)](#) provided similar expressions to those given by [Johri *et al.* \(2020\)](#).

We modelled deleterious mutations occurring in a single functional element and examined π for neutral mutations in and around this region after an instantaneous change in the recombination rate (Figure 1 and S2). Up to $2N_e$ generations after a change in the recombination rate, π more closely resembles the expectation under the ancestral recombination rate than it does the derived rate (Figure 1). After around $4N_e$ generations, coalescence times closely reflect those expected under BGS given the derived recombination rate, as measured by π (Figure 1). The results in Figure 1 were obtained assuming deleterious mutations with constant effect, a distribution of fitness effects for harmful mutations leads to similar results (Figure S2). When deleterious mutations have only mild effects, Equation 1 does not fit the observed data particularly well (Figure X) because in such cases BGS does not resemble a simple reduction in N_e ([Good *et al.* 2014](#); [Cvijović *et al.* 2018](#)).

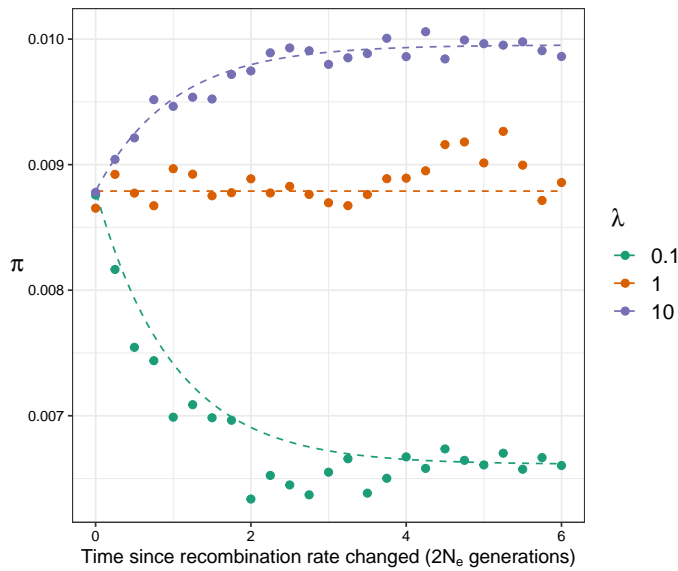


Figure 1 Nucleotide diversity over time after recombination rates change by a factor λ . The dashed lines were calculated using Equation 1, points indicate the mean from 100 replicate simulations.

Patterns of genetic variation after evolution of the recombination landscape

Consider a population that has recently undergone shifts in the recombination rate landscape (i.e. less than $2N_e$ generations ago). Estimates of r from such a population would likely reflect contemporary recombination rates regardless of how they were obtained. Empirical estimates of r always reflect contemporary rates and indirect estimates (i.e. obtained from patterns of LD) can reflect contemporary recombination rates within $0.5N_e$ generations of a change (Figure S3). Depending on the extent and nature of recombination rate evolution, population genomic analyses that compare features of genetic variability to estimates of r may be obscured.

To demonstrate how population genomic analyses may be affected by changes in r , we simulated two scenarios of BGS under evolving recombination rates. In the first, the broadscale landscape of r is rearranged, approximating a burst of karyotype evolution (Figure S4A). In the second, the locations of recombination hotspots are shifted, as if a new PRDM9 allele had fixed in a population (Figure S4B). In both scenarios, deleterious mutations occur across the genome generating widespread BGS such that at equilibrium there is a positive correlation between π and r (Figure 2).

A positive correlation between π and r is the hallmark signature of widespread selection across a genome (Cutter and Payseur 2013), but evolution of the recombination rate may obscure such a pattern. In both of the scenarios we simulated, during the lag period after recombination rate change the correlation between π and r would be misleading (Figure 2). In both cases, a strongly significant positive correlation reestablishes after about $2N_e$ generations (Figure 2). When recombination maps evolved by the movement of hotspots, the scale at which π and r are estimated determined whether a positive correlation between π and r was found (Figure 2). Of course there are reasons why species may not exhibit a positive correlation between π and r (Cutter and Payseur 2013) that have nothing to do with recombination rate evolution. For example wild and domesticated rice (*Oryza spp.*) exhibit negative correlations between π and r , but in those species there is a strong positive correlation between the density of functional sites (i.e. sites subject to purifying selection) and the recombination rate (Flowers et al 2011).

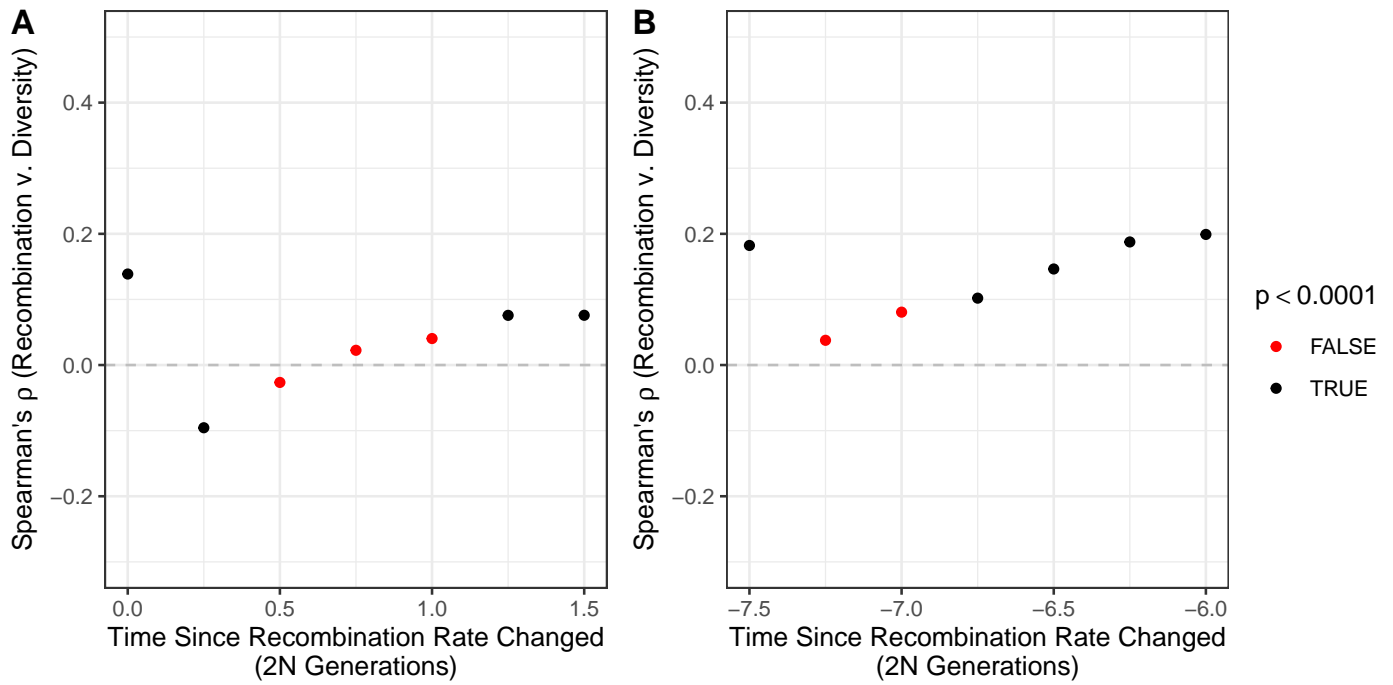


Figure 2 Spearman's correlation between nucleotide diversity (π) and recombination rate (r) over time after recombination rates evolve. Results are shown for 100 Kbp analysis windows.

Rapid recombination rate evolution in house mice

Recombination rate landscapes can evolve very rapidly. For example, due to the requirement of at least one cross-over per chromosome per meiosis in mammals, changes in chromosome length induce changes in r and thus karyotype evolution likely influences the recombination rate. The lineage leading to *Mus musculus* ($2n=40$) has experienced large chromosomal rearrangements since it shared a common ancestor with *Mus pahari* ($2n=48$) 3-5 million years ago (Thybert *et al.* 2018). Moreover, there are differences in r among populations of *Mus musculus domesticus* that harbour different karyotypes (Vara *et al.* 2021). Chromosomal fusions can exhibit meiotic drive (Chmátal *et al.* 2014) so new karyotypes may spread to fixation very rapidly. The locations of actual recombination events in mice are typically restricted to narrow windows of the genome (on the order of 1-5 Kbp), referred to as hotspots (Paigen *et al.* 2008). The locations of recombination hotspots in mice, and many other vertebrates, are determined by the binding of a protein encoded by the *PRDM9* gene to specific DNA motifs (Baudat *et al.* 2010; Baker *et al.* 2017). There is evidence that *PRDM9* has undergone recurrent bouts of positive selection in mice (Oliver *et al.* 2009) and natural populations of *M. musculus spp.* possess various *PRDM9* alleles corresponding to different suites of recombination hotspots (Smagulova *et al.* 2016). Overall, there is clear evidence from mice that recombination rates can evolve rapidly at broad and fine scales.

It is plausible that rapid evolution of recombination rates in *Mus musculus* has influenced our ability to identify the effects of selection across the species' genome. Kartje *et al.* (2020) recently demonstrated that natural populations of *M. m. domesticus* exhibit a very weak correlation between π and r and concluded that there was fairly limited effects of BGS genome-wide. This is somewhat unexpected, however, because wild mice are thought to have large effective population sizes (Leffler *et al.* 2012) and genome-wide evidence for selection is thought to be apparent in species with large N_e (Cutter and Payseur 2013). If the burst of karyotype evolution that occurred over the last few million years in the ancestors of house mice affected recombination rates, then mouse populations may be within the lag period described by Equation 1. If patterns of genetic diversity in mice are still adjusting to changes in the recombination rate, we might see a stronger correlation between π and r in genomic regions that have not undergone dramatic changes in the recombination rate.

Using an alignment of genomes from closely related species, Thybert *et al.* (2018) identified chromosomes

Window	Population	Whole Genome		Conserved Chromosomes		Non-Conserved Chromosomes	
		Spearman's ρ	p -value	Spearman's ρ	p -value	Spearman's ρ	p -value
5Kbp	Gough Island	0.007 67	4.28×10^{-5}	0.008 80	0.0102	0.004 86	0.0302
5Kbp	France	0.004 08	0.0295	0.0403	6.10×10^{-32}	-0.0107	1.76×10^{-6}
5Kbp	Germany	0.007 52	6.05×10^{-5}	0.0152	9.63×10^{-6}	0.003 86	0.0849
1Mbp	Gough Island	0.0536	0.009 46	0.0588	0.124	0.0437	0.0748
1Mbp	France	0.0450	0.0294	0.135	0.000 400	0.009 99	0.684
1Mbp	Germany	0.0535	0.009 53	0.0775	0.0428	0.0426	0.0828

Table 1 The correlation between nucleotide diversity (π) and recombination rate for three populations of house mice (*Mus musculus domesticus*) calculated from all autosomes, conserved chromosomes that exhibit no syntenic breaks between *M. musculus* and *M. pahari* and the non-conserved chromosomes as identified by [Thybert et al. \(2018\)](#).

in the *M. musculus* genome that have or have not undergone dramatic rearrangements in the last 5 million years. We re-analysed data from [Kartje et al. \(2020\)](#) and found that the correlation between π and r is stonger and more significant on chromosomes that have not undergone rearrangements in the last 3-5 million years (Table 1) for *M. m. domesticus* individuals from France and Germany. This pattern holds when looking at analysis windows of 5Kbp and 1Mbp (Table 1). *M. m. domesticus* colonised Gough Island in the 19th century and experienced a severe population bottleneck (Grey et al 2014), which may have further obscured patterns of variability.

Our re-analysis suggests that mice are still within the lag period after evolution of the recombination rate, such that π in *M. m. domesticus* does not fully reflect contemporary recombination rates in *Mus musculus*. In contrast, the ancestors of *Heliconius* butterflies also underwent largescale karyotype evolution, but gross patterns of π versus chromosome length in those species suggest that patterns of variation have largely re-equilibrated after changes in r ([Cicconardi et al. 2021](#)).

As pointed out by ([Comeron 2017](#)) and Smukowski-Heil et al (201

In comparison to mammals, birds have fairly conserved karyotypes and in some cases highly conserved recombination landscapes (REF; Singhal et al 2017).

Methods

Model

Background selection has been modelled as the reduction in effective population size (N_e) at a neutral site due to the removal of deleterious variants. The effects of background selection are often expressed as $B = \frac{N_e}{N_0}$, where N_e is the effective population size and N_0 is the expected population size under strict neutrality. In a non-recombining genome, B is proportional to the ratio of the deleterious mutation rate to the strength of selection acting on harmful mutations (Charlesworth et al 1993). For a neutral site present on a recombining chromosome, the effects of background selection depend on the density of functional sites (i.e. those that can mutate to generate deleterious alleles), the mutation rate, the strength of selection and the recombination rate (Hudson and Kaplan 1995; Nordborg et al 1996; Nordborg 1997). For a neutral locus v linked to x functional sites, the reduction in N_e has been described with the following equation:

$$B_v = \frac{N_e}{N_0} = \exp\left[-\sum_x \frac{u_x}{t(1 + (1 - t)r_{x,v}/t)^2}\right] \quad (2)$$

Where u_x is the deleterious mutation rate at functional site x , t is the heterozygous fitness effect of a deleterious mutation (i.e. 0.5s in the case of semi-dominance) and $r_{x,v}$ is the recombination map distance between the

neutral locus and functional site x . In the above equation, deleterious mutations have fixed effects, but it is straightforward to incorporate a distribution of fitness effects (Nordborg *et al.* 1996). The above equation holds when selection is sufficiently strong such that random drift does not overwhelm selection ($N_e s > 1$) (Good *et al.* 2014).

Simulations

We simulated BGS under recombination rate evolution using two types of simulations in *SLiM* v3.2 (Haller and Messer 2019). Diploid populations of $N_e = 5,000$ individuals were simulated and in all cases we scaled the parameters by population size to species with large larger N_e .

The first set of simulations was designed to examine how long it takes for patterns of neutral diversity under BGS to equilibrate after the recombination rate evolves. In these simulations, the genome was 25Kbp long with a 5Kbp functional element in the centre. Mutations occurred in the functional element at rate $\mu = 2.5 \times 10^{-6}$ and had semi-dominant fitness effects with a fixed selection coefficient of $s = -0.01$. We also simulated cases with varying fitness effects using a gamma distribution with mean (\bar{s}) of -0.1 and a shape parameter of 0.1. Recombination occurred at a uniform rate of $r = 2.5 \times 10^{-6}$ across the chromosome. After 15,000 generations, we simulated an instantaneous change in the recombination rate, multiplying r by λ , giving $r = \lambda 2.5 \times 10^{-6}$. We simulated cases with $\lambda = 0.1, 1.0$ and 10.0 . Simulated populations were sampled every 500 generations after the recombination rate changed and we performed 200 replicates for each set of parameters tested. Note that these simulations were not designed to be particularly realistic, but to provide clear cut patterns to test the theoretical predictions.

The second set of simulations was designed to examine how patterns of π versus r varied over time when recombination rates evolved at fine and/or broad scales. For these simulations, we modelled chromosomes that were 10Mbp long. Neutral mutations occurred at random across the length of the sequence at a rate of 5×10^{-6} . Deleterious mutations occurred at random across the length of the sequence at a rate of 2×10^{-8} with semi-dominant fitness effects drawn from a gamma distribution with a mean (\bar{s}) of -0.1 and a shape parameter of 0.1. The deleterious mutation rate was chosen so that 4% of the genome was subject to purifying selection. Populations evolved under background selection for 100,000 generations (i.e. $20N_e$ generation). In generation 100,000 there was instantaneous evolution the recombination landscape after which we recorded the tree-sequence of the population every 5,000 generations for . We incorporated two models of recombination rate variation and evolution of the recombination map:

- We modelled recombination rate evolution at broad scales by rearrangement of the recombination landscape. Recombination rates vary across the genome (Shapley et al 2018). For example, recombination rates vary by a factor of 3 across chromosome 1 in mice. In these simulations, recombination varied from $r = 2.08 \times 10^{-7}$ to $r = 6.24 \times 10^{-7}$. When the recombination landscape evolved, we reversed the order of recombination rates across the genome (Figure S4A).
- We modelled the evolution the recombination map by the movement of hotspots. Recombination occurred at a uniform rate of $r = 6 \times 10^{-8}$ except in 5Kbp hotspots where it occurred at a rate of $r = 6 \times 10^{-6}$. At the beginning of a simulation, a Poisson number of hotspots was sampled with an expectation of 120. Hotspots were placed at random across the simulated chromosome. When the recombination landscape evolved, we resampled the locations of hotspots (Figure S4B).

In both cases, rates were chosen such that the total map length was similar to one that recombined at a constant rate of $4N_e r = 0.008$, the value reported for wild mice (Booker et al 2017).

For all simulations, we used the tree sequence recording option in *SLiM* and neutral mutations were added to the resulting tree-sequences at a rate of 5×10^{-7} using *PySLiM* and *msprime* (Haller *et al.* 2019; Kelleher *et al.* 2016). Nucleotide diversity (π) was calculated in windows of varying size using *sci-kit-allel*. We used the

program *PyRho* (Spence and Song 2019) to estimate recombination rates from samples of 10 diploid individuals. Spearman's ρ between π and r was calculated using R. All figures were made using ggplot2.

Acknowledgements

I would like to extend thanks to Nadia Singh and Judith Mank for inviting me to present this work at vSMBE 2021. Add thanks to everyone who has read the manuscript.

References

Literature Cited

- Baker, Z., M. Schumer, Y. Haba, L. Bashkirova, C. Holland, *et al.*, 2017 Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *eLife* **6**.
- Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober, *et al.*, 2010 PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice. *Science* **327**: 836–840.
- Campos, J. L., L. Zhao, and B. Charlesworth, 2017 Estimating the parameters of background selection and selective sweeps in *Drosophila* in the presence of gene conversion. *Proceedings of the National Academy of Sciences* **114**: E4762–E4771.
- Charlesworth, B., M. T. Morgan, and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- Chmátal, L., S. I. Gabriel, G. P. Mitsainas, J. Martínez-Vargas, J. Ventura, *et al.*, 2014 Centromere Strength Provides the Cell Biological Basis for Meiotic Drive and Karyotype Evolution in Mice. *Current Biology* **24**: 2295–2300.
- Cicconardi, F., J. J. Lewis, S. H. Martin, R. D. Reed, C. G. Danko, *et al.*, 2021 Chromosome Fusion Affects Genetic Diversity and Evolutionary Turnover of Functional Loci but Consistently Depends on Chromosome Size. *Molecular Biology and Evolution* **38**: 4449–4462.
- Comeron, J. M., 2017 Background selection as null hypothesis in population genomics: insights and challenges from *Drosophila* studies. *Philos Trans R Soc Lond B Biol Sci* **372**.
- Cutter, A. D. and B. A. Payseur, 2013 Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat Rev Genet* **14**: 262–274.
- Cvijović, I., B. H. Good, and M. M. Desai, 2018 The Effect of Strong Purifying Selection on Genetic Diversity. *Genetics* **209**: 1235–1278.
- Elyashiv, E., S. Sattath, T. T. Hu, A. Strutsovsky, G. McVicker, *et al.*, 2016 A genomic map of the effects of linked selection in *Drosophila*. *PLoS Genetics* **12**: e1006130.
- Good, B. H., A. M. Walczak, R. A. Neher, and M. M. Desai, 2014 Genetic Diversity in the Interference Selection Limit. *PLoS Genetics* **10**.
- Haller, B. C., J. Galloway, J. Kelleher, P. W. Messer, and P. L. Ralph, 2019 Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Molecular Ecology Resources* **19**: 552–566.
- Haller, B. C. and P. W. Messer, 2019 SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. *Molecular Biology and Evolution* **36**: 632–637.
- Johri, P., B. Charlesworth, and J. D. Jensen, 2020 Toward an evolutionarily appropriate null model: Jointly inferring demography and purifying selection. *Genetics* **215**: 173–192.
- Kartje, M. E., P. Jing, and B. A. Payseur, 2020 Weak Correlation between Nucleotide Variation and Recombination Rate across the House Mouse Genome. *Genome Biology and Evolution* **12**: 293–299.
- Kelleher, J., A. M. Etheridge, and G. McVean, 2016 Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLOS Computational Biology* **12**: e1004842.
- Nam, K., K. Munch, T. Mailund, A. Nater, M. P. Greminger, *et al.*, 2017 Evidence that the rate of strong selective sweeps increases with population size in the great apes. *Proceedings of the National Academy of Sciences of the United States of America* **114**: 1613–1618.
- Nordborg, M., B. Charlesworth, and D. Charlesworth, 1996 The effect of recombination on background selection. *Genetical Research* **67**: 159–174.

- Oliver, P. L., L. Goodstadt, J. J. Bayes, Z. Birtle, K. C. Roach, *et al.*, 2009 Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. *PLoS Genetics* **5**: e1000753.
- Paigen, K., J. P. Szatkiewicz, K. Sawyer, N. Leahy, E. D. Parvanov, *et al.*, 2008 The recombinational anatomy of a mouse chromosome. *PLoS Genetics* **4**: e1000119.
- Pool, J. E. and R. Nielsen, 2009 Correction for pool and nielsen (2007). *Evolution* **63**: 1671–1671.
- Smagulova, F., K. Brick, P. Yongmei, R. D. Camerini-Otero, and G. V. Petukhova, 2016 The evolutionary turnover of recombination hotspots contributes to speciation in mice. *Genes & Development* **30**: 277–280.
- Spence, J. P. and Y. S. Song, 2019 Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Science Advances* **5**.
- Sun, H., B. A. Rowan, P. J. Flood, R. Brandt, J. Fuss, *et al.*, 2019 Linked-read sequencing of gametes allows efficient genome-wide analysis of meiotic recombination. *Nature Communications* **10**: 1–9.
- Thybert, D., M. Roller, F. C. Navarro, I. Fiddes, I. Streeter, *et al.*, 2018 Repeat associated mechanisms of genome evolution and function revealed by the *Mus caroli* and *Mus pahari* genomes. *Genome Research* **28**: 448–459.
- Tigano, A., R. Khan, A. D. Omer, D. Weisz, O. Dudchenko, *et al.*, 2021 Chromosome size affects sequence divergence between species through the interplay of recombination and selection. *bioRxiv* p. 2021.01.15.426870.
- Vara, C., A. Paytuví-Gallart, Y. Cuartero, L. Álvarez-González, L. Marín-Gual, *et al.*, 2021 The impact of chromosomal fusions on 3D genome folding and recombination in the germ line. *Nature Communications* **2021 12:1 12**: 1–17.

Supplementary Material

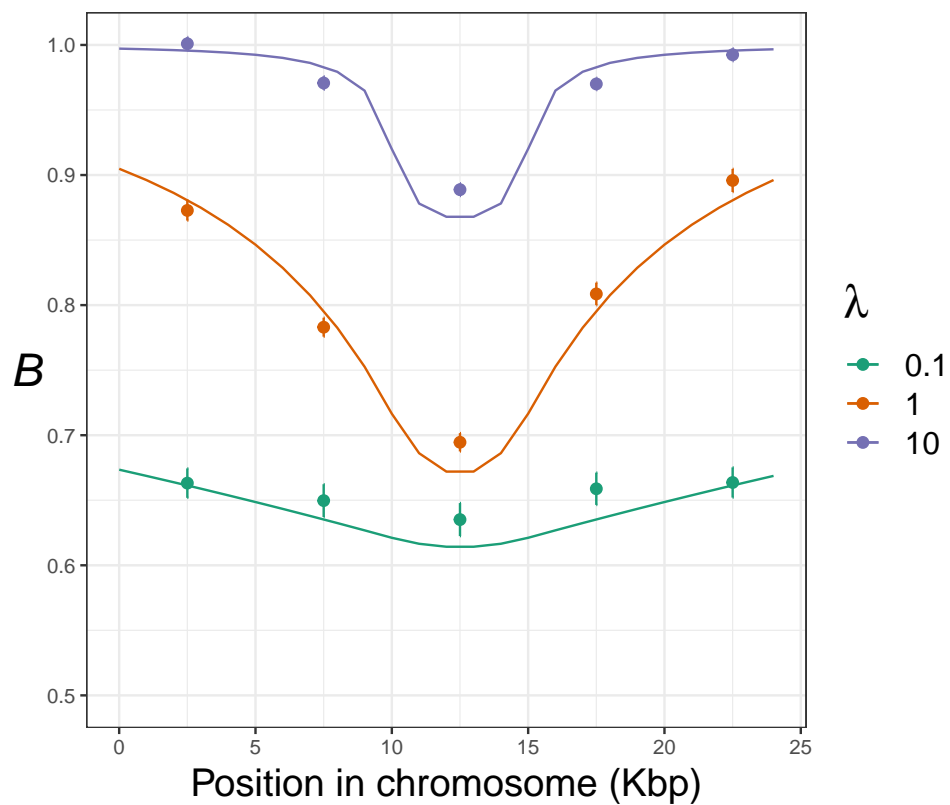


Figure S1 The effects of background selection across simulated chromosomes. B was calculated for simulated data by comparing observed π to the neutral expectation of $4N_e\mu = 0.01$. The lines show the theoretical expectation calculated using formulae from Nordborg et al (1996).



Figure S2 The effects of background selection across simulated chromosomes. B was calculated for simulated data by comparing observed π to the neutral expectation of $4N_e\mu = 0.01$. The lines show the theoretical expectation calculated using formulae from Nordborg et al (1996). The labels on the top of each panel indicate the location in the simulated data being analysed.

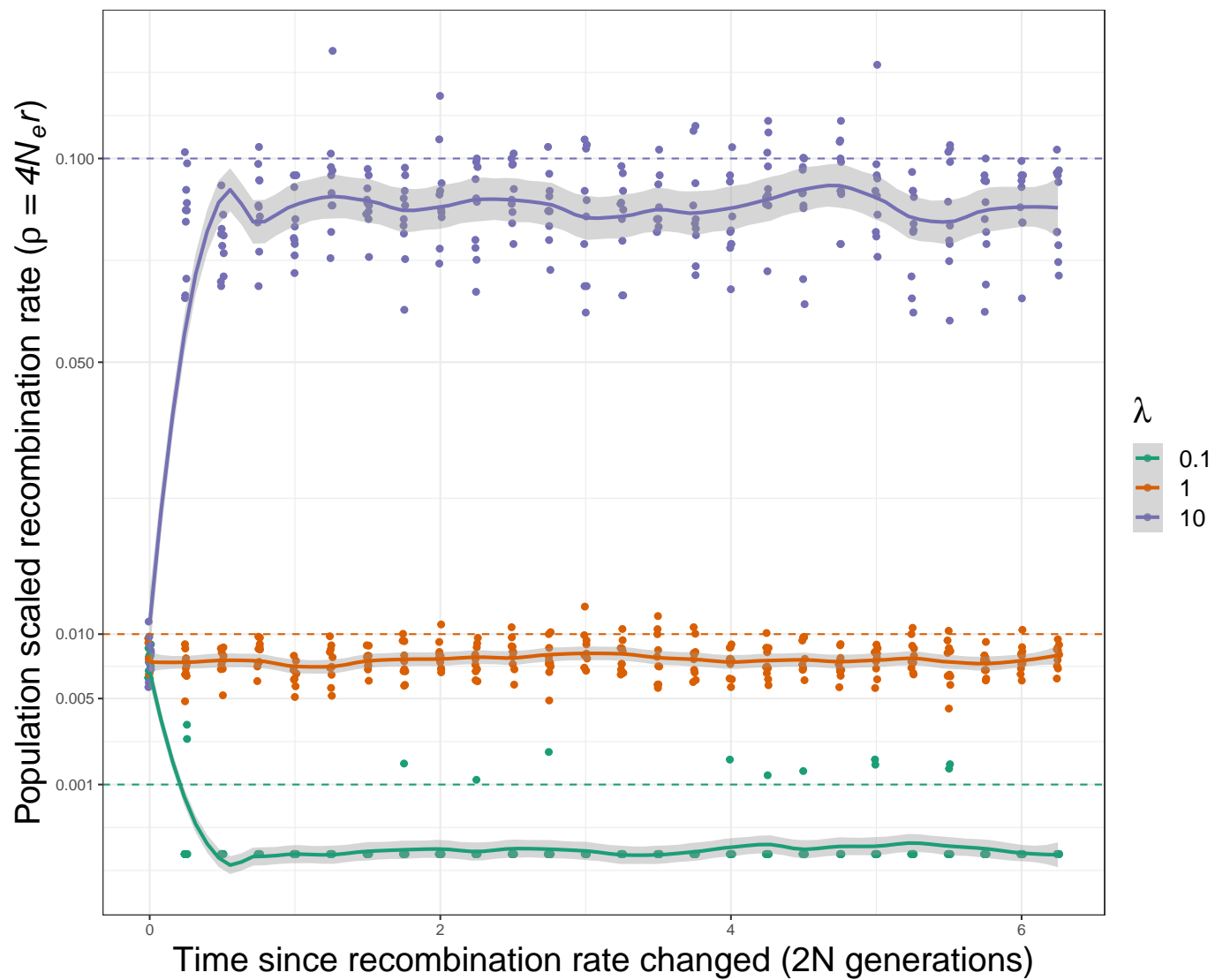


Figure S3 Recombination rates inferred using *PyRho* after an instantaneous change in the recombination rate. Dashed horizontal lines indicate the true recombination rate for the three cases. Smoothed lines with shaded ribbon indicate the fit and error of a LOESS regression.

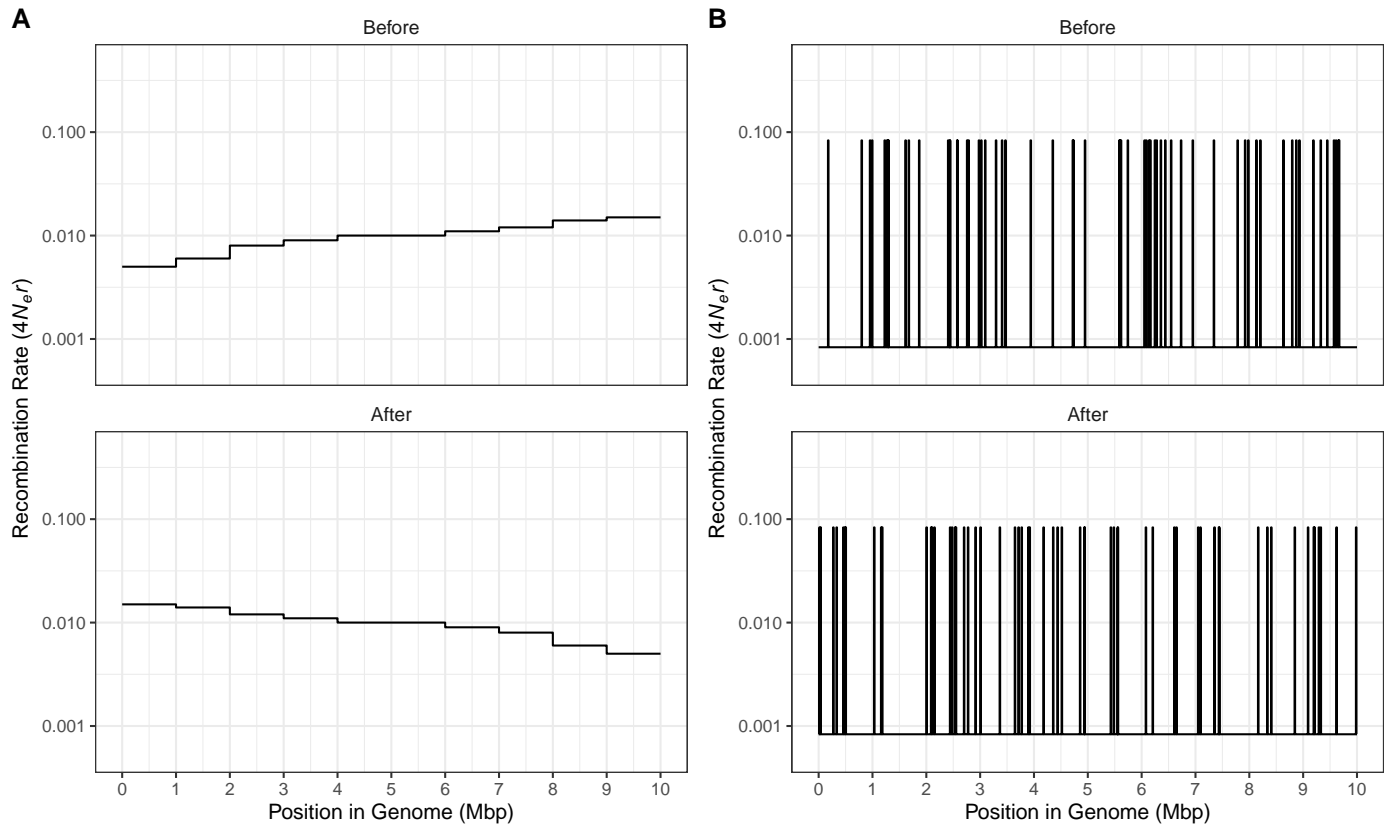


Figure S4 The recombination rate maps used in the simulations modelling BGS across the genome. The upper and lower panels show the recombination rate landscape before and after it evolved in simulations, respectively. A) Evolution of the recombination rate at the Mbp scale. B) Evolution of the recombination rate at the scale of recombination hotspots.