

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 the recombination rate. 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 (Keightley and Eyre-Walker review). For those reasons, it has been proposed that BGS is likely ubiquitous across eukaryotic genomes and should be incorporated into the null model for molecular evolution (Comeron 2014; Gilbert and Pouyet 2020; Jensen et al 2018; Johri et al 2020).

One effect of BGS is a reduction in genetic diversity for regions of the genome subject to purifying selection. The expected reduction in nucleotide diversity (π) un-

der BGS is proportional to the ratio of the local deleterious mutation rate and the local recombination rate (Charlesworth et al 1993; Nordborg 1996; Hudson and Kaplan 1995). The first empirical evidence that selection at linked sites of any kind influences genetic variation across the genome came from studies in *Drosophila*. Aguadé et al (1989) measured genetic variability in the *yellow-achaete-scute* regions located at the tip of the X-chromosome in *D. melanogaster*. The *yellow-achaete-scute* regions experience restricted crossing-over and Aguadé et al (1989) found that they harbour far less genetic variation than had been reported for more highly recombining regions of the genome. Subsequent studies demonstrated a positive correlation between nucleotide diversity and recombination rate genome-wide in *D. melanogaster* (Begun and Aquadro 1993; Andolfatto 2002) and similar patterns have been reported in numerous other species (Cutter and Payseur 2012).

Interpreting genome-wide patterns of genetic variability in terms of selection at linked sites requires accurate estimates of population genetic parameters. Empirical estimates of the recombination rate can be obtained by exam-

ining 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 a number of generations, and thus provide recombination rate estimates for contemporary populations. Alternatively, recombination rate estimates can be obtained indirectly by analysing linkage disequilibrium across the genome (REFs) and such estimates reflect a 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 recombination rates have not changed over the time in which patterns of diversity have been established under BGS.

However, recombination rate landscapes can evolve rapidly. For example, there has been extensive evolution of recombination rates at broad and fine scales in the house mouse (*Mus musculus*). Due to the requirement of at least one cross-over per chromosome per meiosis in mammals, karyotype evolution likely influences recombination rate landscapes at broad scales. The lineage leading to *M. musculus* ($2n=40$) has experienced large chromosomal rearrangements since it shared a common ancestor with *Mus pahari* ($2n=48$) 5 million years ago (Thybert et al 2018). Moreover, there are differences in recombination rate among populations of *Mus musculus domesticus* that harbour different karyotypes (Vara et al 2021). Recombination events in mice are typically restricted to narrow windows of the genome (on the order of 5-10 Kbp), referred to as hotspots (Paigen et al 2007). The locations of recombination hotspots in mice, as well as humans 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.* harbour various *PRDM9* alleles corresponding to different suites of recombination hotspots (Smagulova et al 2015). Overall, there is clear evidence that recombination rates have evolved at broad and fine scales in mice in the relatively recent past.

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. 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 the derived rates. Depending on how recombination rate landscapes evolve, analysis of BGS in lineages that

are still within that lag period may be obscured. For example, the hallmark signature of selection at linked sites, a positive correlation between nucleotide diversity and recombination rate may not be clearly observed. In this letter, we examine how patterns of neutral genetic variability under BGS respond to evolution of the recombination rate and describe how this would affect 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. When deleterious mutations have large effects on fitness, BGS resembles a localised reduction in the effective population size (N_e). So, changes in the recombination rate for a particular genetic region may resemble a localised change in N_e . Indeed, an expression that describes coalescence times after a change in population size (Equation 2) accurately predicts the effects of BGS under recombination rate evolution (Figure 1). It takes around $4N_e$ generations for nucleotide diversity to closely reflect an increase or a decrease in the recombination rate (Figure 1). Up to $2N_e$ generation after such a change, however, nucleotide diversity under BGS more closely resembles the expectation under the ancestral recombination rate than it does the derived rate (Figure 1). The fit of this breaks down when deleterious mutations have only mild effects on fitness (Figure X) because in such cases BGS does not resemble a simple reduction in the N_e (Good et al 2014; Cvijović et al 2018).

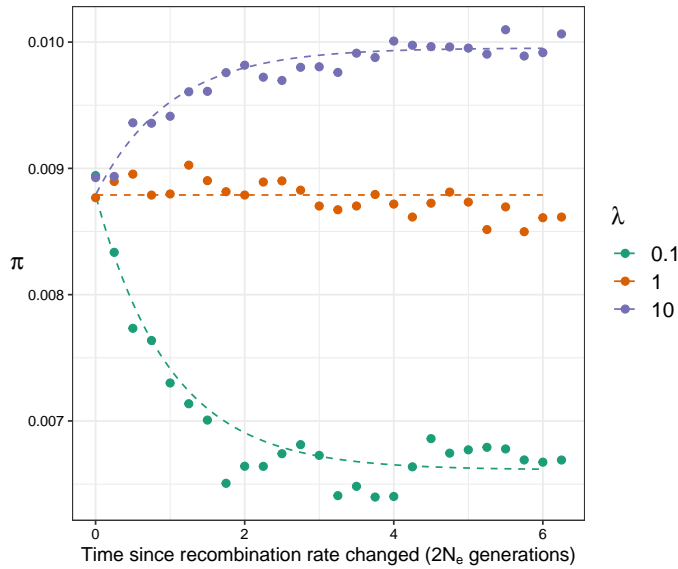


Figure 1 Nucleotide diversity over time after recombination rates change by a factor λ for neutral sites in or around a functional element. The dashed lines were calculated using Equation 2, points indicate the mean from XXX replicate simulations.

For the sake of simplicity, the model we used assumes that recombination rates evolve instantaneously. While that is obviously an oversimplification, there is reason to expect that recombination rate changes may evolve rapidly. Chromosomal fusions can exhibit meiotic drive (Chmátal et al 2014) so may spread to fixation very rapidly and there is evidence that the *PRDM9* gene that controls the location of recombination hotspots in many vertebrate species has undergone recurrent bouts of positive selection (Oliver et al 2009).

Empirical recombination rate estimates (i.e. obtained from direct observation of recombination events) reflect contemporary populations, but indirect estimates may reflect a combination of ancestral and contemporary rates. In Figure XY, we show that indirect estimates recombination rates estimates (Spence et al 2018) very quickly

Evolving recombination rates and evidence for natural selection

When a population is sampled during the lag period after evolution of the recombination rate (e.g. before $2N_e$ generations have elapsed), inferences made from genomic data about molecular evolution may be obscured. The effect that recombination rate evolution will have on inferences made about molecular evolution will, of course, depend on how recombination rates change. We performed two sets of simulations to demonstrate effects of recombination rate evolution. In the first, we simulated a dramatic rearrangement of the recombination rate map at broad scales (Figure S2). In the second, we modelled fine-scale recombination rate evolution by shifting the landscape of

recombination hotspots (Figure S2).

The lineage leading to *Mus musculus* underwent a bout of karyotype evolution around 3 MYA (Thybert et al 2018). Thybert et al (2018) identified ancestral chromosomal rearrangements in the lineage leading to *M. musculus* using an alignment of rodent genomes. We classified *M. musculus* chromosomes as either "non-conserved" or "conserved" based on whether Thybert et al (2018) identified ancestral rearrangements or not, respectively.

We compared the correlation between π and r on the conserved and non-conserved chromosomes of *Mus musculus domesticus* using previously analysed data from Kartje et al (2020)(Table 1). We found a pattern suggesting that π does not fully reflect the current recombination rate regime in *Mus musculus*. The correlation between π and r was weaker and less significant for the non-conserved chromosomes compared to the conserved chromosomes in all cases tested (Table 1).

For Gough island mice, the correlation was suggests that *M. musculus* has

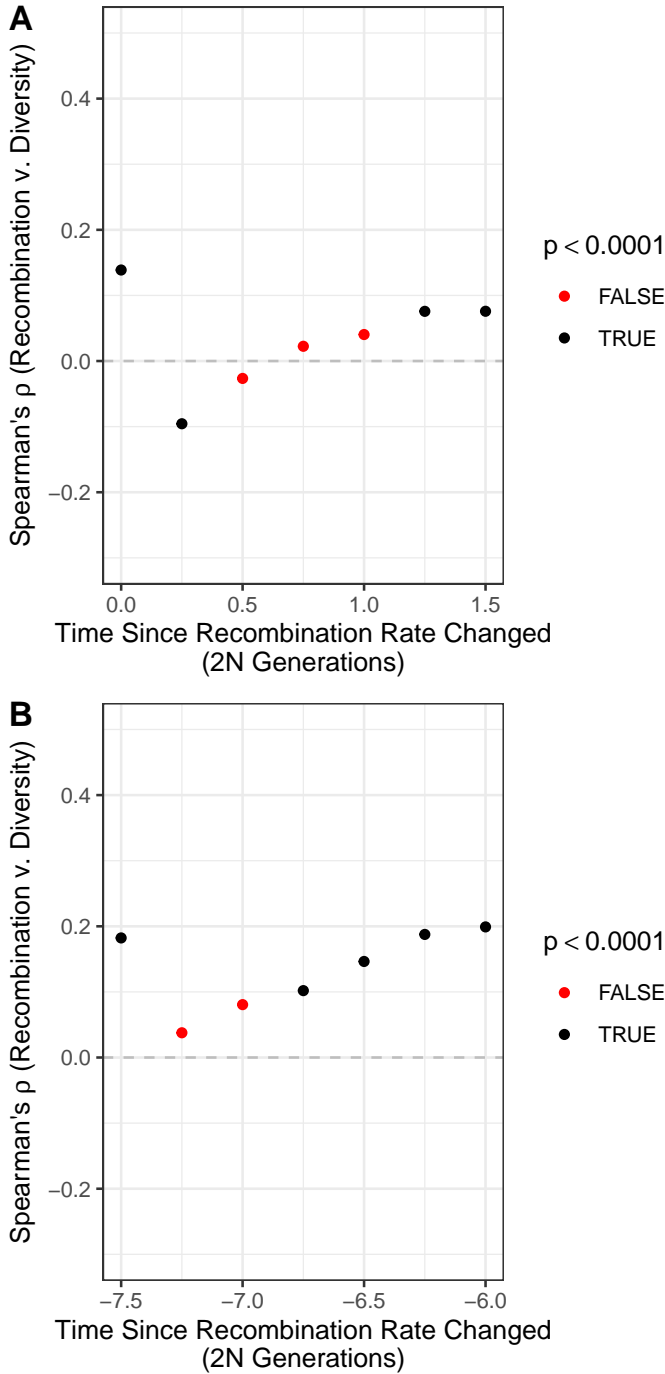


Figure 2 Spearman's correlation between nucleotide diversity (π) and recombination rate (r) over time after recombination rates evolve. Results are shown for 100 Kbp analysis windows.

Numerous species that have been examined do not exhibit a positive correlation between diversity and recombination rate (Cutter and Payseur). For example, wild and domesticated rice species (*Oryza rufipogon* and *Oryza sativa*, respectively) exhibit negative correlations between diversity and recombination rate, potentially due to a positive correlation between functional density and re-

combination rate in those species (Flowers et al 2011).

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

We have restricted our analysis to what happens when fitness affecting mutations are strictly deleterious and generate BGS. However, other modes of selection and their effects on linked variation may also affect nucleotide diversity. For example, there is evidence that recurrent selective sweeps influence diversity patterns in numerous species (Nam et al; Booker et al 2021; Campos et al; Elyashiv et al). It seems reasonable to expect that recombination rate evolution would also influence inferences about selective sweeps.

In mice, nucleotide divergence between *Peromyscus eremicus* and *P. maniculatus* exhibits a clear negative correlation with chromosome size, but for *Mus pahari*, sister species to *M. musculus*, there is no such correlation (Tigano et al 2021). Tigano et al (2021) performed analyses that suggest that karyotype evolution caused gave rise to shifts in the contribution of ancestral diversity to divergence.

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:

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

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

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).

2014).

When the recombination rate landscape evolves it may cause a change in the effects of BGS in particular genomic regions. We modelled the time it takes for coalescence times, or patterns of nucleotide diversity, to reflect background selection expected under a new recombination rate regime using expressions formulated to describe coalescence times after a population size change. 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 $B_{\nu,1}$. We modified an expression for coalescence times after an instantaneous population size change from Johri et al (2020), to obtain the following equation,

$$B_{\nu,\Delta r} = B_{\nu,1} \left(1 + \left(\frac{B_{\nu,0}}{B_{\nu,1}} - 1 \right) e^{-T_0} \right) \quad (2)$$

Note that Equation X from Pool and Nielsen (2009) provided similar expressions to those given in Johri et al (2020).

Simulations

We simulated BGS under recombination rate evolution using two types of simulations in *SLiM* v3.2 (Haller et al 2018). In all cases, diploid populations of $N = 1,000$ individuals were simulated.

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 -0.05. 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. Re-

combination 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. Deleterious mutations occurred at random across the length of the sequence at a rate of 1×10^{-7} 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. Populations evolved under background selection for 15,000 generations

We modelled the evolution of hotspots in the following way. At the beginning of a simulation, a Poisson number of hotspots was sampled with an expectation of 60, which was based on the average number of double-strand break hotspots observed by Smagulova et al (2015). Locations for 10,000bp hotspots were then sampled across the simulated chromosome. Recombination occurred at a uniform rate of $r = 2.08 \times 10^{-7}$ except in hotspots where it occurred at a rate of $r = 2.08 \times 10^{-5}$. These rates were chosen to give an overall recombination rate similar to that of a chromosome that recombined at a uniform rate of $r = 2.5 \times 10^{-6}$.

We modelled recombination rate evolution at broad scales in the following way. The genome-wide average recombination rate in *Mus musculus* is XXX cM/Mbp. The distribution of recombination rates is approximately

normal with mean XXX cM/Mbp and variance YYY. We scaled recombination in these simulations to simulate a large natural population using a comparatively small number of individuals in *SLiM*. At the beginning of a simulation, we sampled 10 recombination rates from a normal distribution (mean scaled_XXX, variance scaled_YYY). As above, we generated a new recombination map after 15,000 generations of evolution, sampled the population then continued to sample the population every we generated a new set of recombination hotspots as above and sampled the population every 500 generations for a further 3,000 generations.

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 2.5×10^{-6} using PySLiM and msprime (version XX, REF). Nucleotide diversity (π) was calculated in windows of varying size using sci-kit-allel (version X.Y, REF) and Spearman's ρ between π and r was calculated using R (version X.Y, REF)

Acknowledgements

I would like to extend thanks to Nadia Singh and Judith Mank for inviting me to present this work at vSMBE 2021.

Supplementary Material

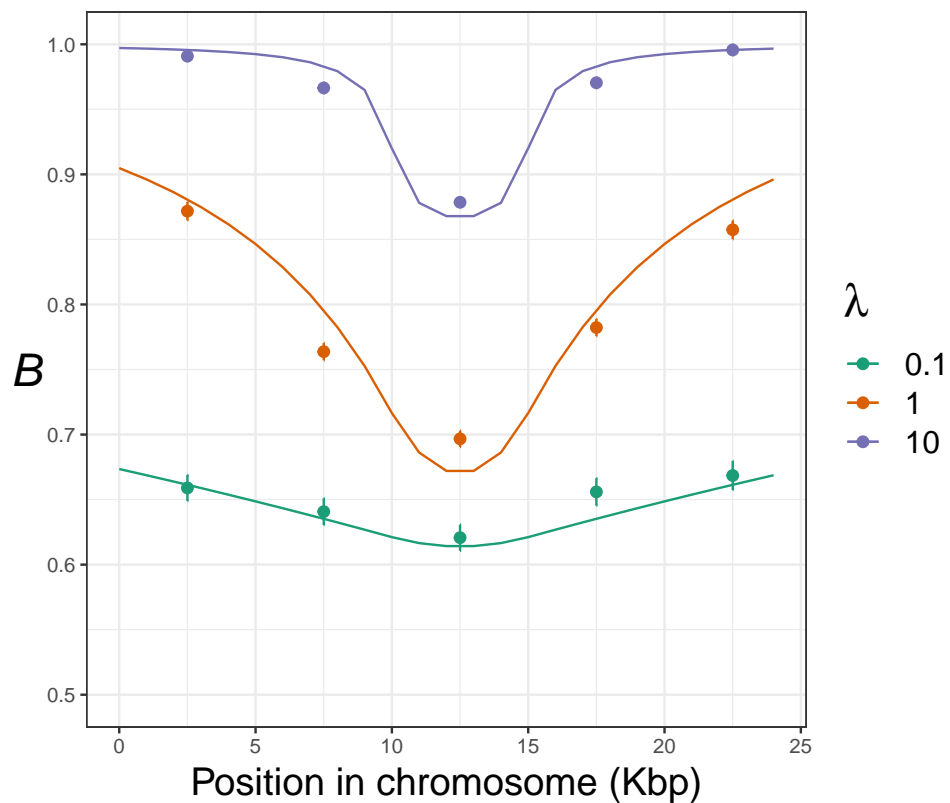


Figure S1 The effects of background selection across simulated chromosomes. B was calculated for simulated data by comparing observed π to the neutral expectation of $4N_e\mu = 0.01$. The lines show the theoretical expectation calculated using formulae from Nordborg et al (1996).