# Supplementary Materials

## Vince Buffalo and Andrew Kern

### July 5, 2023

## Contents

# 1 Quantitative Genetic Linked Selection Model Theory

## 1.1 Stochastic Sources of Neutral Allele Frequency Change

Here, we step through a quick derivation of **Santiago1995-hx**. Throughout, we assume random mating, hermaphroditic individuals, and a constant population size. The change in a neutral allele's frequency in one generation can be partitioned into the three sources of stochasticity: the random associations with fitness backgrounds (i.e. *draft*), the non-heritable randomness in family size, and the Mendelian noise from heterozygotes segregating. If we let $x_{0,i} \in \{0, 1/2, 1\}$ be the frequency of neutral alleles individual $i$ in generation 0 carries, we can partition the random neutral allele frequency of the population ($x_t$ without the individual index) into the underlying stochastic causes,

$$x_1 = \frac{1}{2N} \sum_{i=1}^{N} \left( x_{0,i} k_{0,i} + \sum_{j=1}^{k_{0,i}} \delta_{0,i,j} \right) \tag{1}$$

2

where $k_{0,i}$ is the number of surviving gametes parent $i$ passes on, and $\delta_{0,i,j}$ is a random term that encapsulates the noise due to Mendelian segregation of heterozygotes. If parent $i$ is a homozygote ($x_{0,i} \in \{0,1\}$), then $\delta_{0,i,j} = 0$, whereas if $x_{0,i} = 1/2$ then $\delta_{0,i,j} = \pm 1/2$ with equal probability. This is because a heterozygous parent will transmit half a neutral allele in expectation, but each round of Mendelian segregation must pass on either 0 or 1 alleles, which the random $\pm 1/2$ term imposes. The factor of $1/2$ is due to the fact that we're summing over $N$ diploids, but considering the number of gametes they transmit. Since each diploid parent must have two offspring to maintain a constant population size, $1/N \sum_i k_{0,i} = 2$.

The frequency in the initial generation is $x_0 = 1/N \sum_{i=1}^{N} x_{0,i}$, though to indicate that we treat this as fixed rather than random, we use $p_0 := x_0$. Then, the allele frequency change is,

$$\Delta x_1 = x_1 - p_0 = \frac{1}{2N} \sum_{i=1}^{N} x_{0,i}(k_{0,i} - 1) + \frac{1}{2N} \sum_{i=1}^{N} \sum_{j=1}^{k_{0,i}} \delta_{0,i,j}$$

$$\Delta x_1 = K_1 + H_1 \tag{2}$$

where $K_1$ and $H_1$ are the random change in neutral allele frequency change due to offspring number (including heritable and non-heritable components), and Mendelian segregation in generation 1.

## 1.2 The Variance in Neutral Allele Frequency Change

Now, let us look the variance of $V(\Delta x_1)$ over evolutionary replicates. Since the Mendelian segregation and the offspring random process are independent,

$$V(\Delta x_1) = V(K_1) + V(H_1). \tag{3}$$

Looking at each term,

$$V(K_1) \approx \frac{1}{4N^2} \sum_{i=1}^{N} V\left(x_{0,i}(k_{0,i} - 1)\right) \tag{4}$$

where we ignore the covariance terms due to the sum, since these are on order $1/N^3$. In the first generation from an arbitrary starting point, the neutral alleles assort independently into diploids with respect to their fitness, so we can simply the variance as

$$V(K_1) \approx \frac{1}{4N^2} \sum_{i=1}^{N} V(x_{0,i}) V(k_{0,i} - 1)$$

$$\approx \frac{1}{4N} V(x_{0,i}) V(k_{0,i}). \tag{5}$$

Assuming no correlation between parental gametes (e.g. no inbreeding), $V(x_{0,i})$ is the binomial variance in individual allele frequency, or $p_0(1 - p_0)/2$, and $V(k) := V(k_{0,i})$ is the offspring variance of an individual given by the reproduction process. For example, if the reproduction process is a neutral multinomial Wright–Fisher, $V(k) \approx 2$. Then, the variance in allele frequency change due to non-heritable offspring number variation is,

$$V(K_1) \approx \frac{p_0(1-p_0)}{2N}\frac{V(k)}{4}. \tag{6}$$

The Mendelian noise variance term can be derived similarly. First, the sum over each parent's transmitted gametes can be simplified by noting that parents are exchangeable over evolutionary replicates with respect to their contribution to this term. In a constant population size, the double summation over parents and their offspring can be replaced by a summation over offspring, since both sum $N$ exchangeable terms. Then, note that $V(\delta_{i,j}) = 1/4\,p_0(1-p_0)$, so

$$V(H_1) = \frac{p_0(1-p_0)}{2N}\frac{1}{2}. \tag{7}$$

Finally, we have

$$V(\Delta x_1) = V(K_1) + V(H_1) \tag{8}$$
$$\approx \frac{p_0(1-p_0)}{2N}\frac{V(k)}{4} + \frac{p_0(1-p_0)}{2N}\frac{1}{2}$$
$$\approx \frac{p_0(1-p_0)}{2N}\left(\frac{V(k)}{4} + \frac{1}{2}\right) \tag{9}$$

(c.f. **Santiago1995-hx** equation 2 and **Buffalo2019-qs** equation 30). Note that if the reproduction process is a neutral multinomial Wright–Fisher, $V(k) \approx 2$, and this simplifies to the expected Wright–Fisher variance. If we were to define a variance *drift-effective* population size $N_e$ by setting this variance in neutral allele frequency change to the expected variance under a Wright–Fisher model ($V_{\mathrm{WF}}$), we'd have

$$V_{\mathrm{WF}} := \frac{p_0(1-p_0)}{2N_e}$$
$$V_{\mathrm{WF}} = V(\Delta p_1)$$
$$N_e = \frac{4N}{V(k)+2} \tag{10}$$

(c.f. **Wright1938-tv**).

Now, we look at the variance in allele frequency change across two generations for a neutral allele. The change in neutral allele frequency, regardless of the selective system, is directionless in expectation. This is because whichever allele we track is arbitrary by the definition of neutrality, which imposes a symmetry, $\mathbb{E}(\Delta x) = 0$. Then, the variance of change in the first generation is,

$$V(\Delta x_1) = \mathbb{E}_1\left[(x_1 - x_0)^2\right] = \frac{x_0(1-x_0)}{2N}, \tag{11}$$

and across two generations,

$$\begin{aligned}
V(x_2 - x_0) &= \mathbb{E}\left[(x_2 - x_0)^2\right] \\
&= \mathbb{E}_1\left[\mathbb{E}_2\left[((x_2 - x_1) + (x_1 - x_0))^2 | x_1\right]\right] \\
&= \mathbb{E}_1\left[\mathbb{E}_2\left[(\Delta x_2 + \Delta x_1)^2 | x_1\right]\right] \\
&= \mathbb{E}_1\left[\mathbb{E}_2\left[\Delta x_2^2 | x_1\right]\right] + \mathbb{E}_1\left[\Delta x_1^2\right] + 2\mathbb{E}_1\left[\mathbb{E}_2[\Delta x_2 \Delta x_1 | x_1]\right] \\
&= \mathbb{E}_1\left[\mathbb{E}_2\left[\Delta x_2^2 | x_1\right]\right] + \mathbb{E}_1\left[\Delta x_1^2\right] + 2\mathcal{C}_{1,2} \\
&= \frac{p_0(1 - p_0)}{2N}\left(1 - \frac{1}{2N}\right) + \frac{p_0(1 - p_0)}{2N} + 2\mathcal{C}_{1,2}
\end{aligned} \tag{12}$$

Consequently, the variance of the neutral allele frequency changes each generation according to the probability of failing to coalescence each generation and the pairwise covariance terms $\mathcal{C}_{i,j}$ that build up due to associations with heritable fitness backgrounds. Note that the covariance term $\mathcal{C}_{1,2} = \mathbb{E}_1\left[\mathbb{E}_2[\Delta x_2 \Delta x_1 | x_1]\right] = 0$ when there is no heritable fitness variation, and this simplifies to the well-known equation for variance in a Wright–Fisher population.

## 1.3 Heritable Fitness and the Accumulation of Autocovariance

Next, we turn our attention to the case where there is heritable fitness variation, which leads to non-zero covariance terms $\mathcal{C}_{i,j}$. These terms emerge when $V(k)$ has a heritable component of fitness that can be transmitted along with the neutral allele, thereby affecting the neutral allele's trajectory in later generations. If we partition the offspring number into heritable and non-heritable components, $k_i = 2f_i + \varepsilon_i$, then $V(k_i) = 4V_A + V_n$. When there is heritable fitness variation across individuals, $V_A > 0$. Because the population size is assumed to be constant, the mean heritable fitness across individuals is constrained to be $\mathbb{E}_i(f_i) = 1$; this implies that the individual fitness values $f_i$ are relative fitnesses as is standard in population genetics, in this case to the population mean. Thus $V_A$ is the squared coefficient of heritable fitness variation (this is often denoted as $C^2 = V_A/\bar{w}^2$, see **Crow1958-pc**; **Charlesworth1987-ab**; **Houle1992-ur**).

We can further decompose the $K_1$ term into $K_1 = S_1 + D_1$,

$$\begin{aligned}
V(\Delta x_1) &= V(S_1) + V(D_1) + V(H_1) \\
&\approx \frac{p_0(1 - p_0)}{2N}V_A + \frac{p_0(1 - p_0)}{2N}\frac{V_n}{4} + \frac{p_0(1 - p_0)}{2N}\frac{1}{2}
\end{aligned} \tag{13}$$

(c.f. **Santiago1995-hx** equation 11).

To see how the covariance terms accumulate, consider $V(x_3 - p_0)$. Note that $\mathbb{E}(S_t) = \mathbb{E}(D_t) = \mathbb{E}(H_3) = 0$, since which neutral allele we track is arbitrary, so by symmetry the expected change is zero. Then,

$$\begin{aligned}
V(x_3 - p_0) &= \mathbb{E}\left[(S_1 + D_1 + H_1 + S_2 + D_2 + H_2 + S_3 + D_3 + H_3)^2\right] \\
&= \mathbb{E}(S_1^2) + \mathbb{E}(S_2^2) + \mathbb{E}(S_3^2) \\
&\quad + \mathbb{E}(S_1 S_2) + \mathbb{E}(S_1 S_3) + \mathbb{E}(S_2 S_3) \\
&\quad + \mathbb{E}(D_1^2) + \mathbb{E}(H_1^2) + \mathbb{E}(D_2^2) + \mathbb{E}(H_2^2) + \mathbb{E}(D_3^2) + \mathbb{E}(H_3^2).
\end{aligned} \tag{14}$$

The covariance terms $\mathbb{E}(S_i S_j)$ for $j > i$ represent the expected neutral allele frequency change from a neutral allele becoming associated with a fitness background in generation $i$ and that fitness association persisting until generation $j$. However, the covariances terms $\mathbb{E}(S_j S_i)$ for $j > i$ are zero since associations in the future cannot affect the past (see p. 1041 of **Buffalo2019-qs**). Let us consider $S_1$ and $S_2$. We have,

$$S_1 = \frac{1}{2N} \sum_{i=1}^{N} x_{0,i}(f_{0,i} - 1) = \text{cov}(x_{0,i}, f_{0,i})$$

$$S_2 = \frac{1}{2N} \sum_{i=1}^{N} x_{1,i}(f_{1,i} - 1) = \text{cov}(x_{1,i}, f_{0,i}) \tag{15}$$

which are chance covariances (across the population, *not* evolutionary replicates) created by the random sorting of neutral alleles and fitness into individuals, *each generation*. These covariances are equivalent to the Robertson-Price equation (**Robertson1966-fs**; **Price1970-si**), which predict the change in neutral frequency due to heritable fitness (and likewise with the change due to non-heritable factors).

Considering the underlying haplotypes that lead to $x_{0,i}$ and $f_{0,i}$ can give us an equation for the dynamics of the $\mathbb{E}(S_i S_j)$ (for $j > i$) terms over evolutionary replicates. Let us partition the allele frequency and fitness by into the average of the paternal contributions. The neutral allele frequency per gamete is either 0 or 1, and we assume the fitnesses across gametes are additive. Then, as long as there is random mating, we can simplify the associations in a diploid by looking at a single gamete,

$$\begin{aligned}
S_1 = \text{cov}(x_{0,i}, f_{0,i}) &= \text{cov}\left(\frac{x'_{0,i} + x''_{0,i}}{2}, f'_{0,i} + f''_{0,i}\right) \\
&= \frac{1}{2}\left(\text{cov}(x'_{0,i}, f'_{0,i} + f''_{0,i}) + \text{cov}(x''_{0,i}, f'_{0,i} + f''_{0,i})\right) \\
&= \text{cov}(x'_{0,i}, f'_{0,i} + f''_{0,i}) \\
&= \text{cov}(x'_{0,i}, f'_{0,i}) + \text{cov}(x'_{0,i}, f''_{0,i}) \\
&= S'_1 + S''_1
\end{aligned} \tag{16}$$

which follows from symmetry again, since it does not matter which gamete carrying the neutral allele we track (c.f. **Santiago1998-bs** p. 2107). The primes now indicate whether the selected locus is on the same gamete as the neutral allele (single prime), or the homologous gamete (double prime).

After meiosis, a fraction $1 - r$ of these associations between the tracked neutral neutral allele $x'_{0,1}$ and fitness background $f'_{0,i}$ persist, and a fraction $r$ disassociate its currently coupled background and create an association with the homologous fitness background, $f''_{0,i}$. When linkage is tight, this latter term can be ignored as we will do here, though it does impact background levels of neutral diversity (**Santiago1995-hx**). Additionally, the fitness effects of the selected background may change in later generations, in complex ways (**Barton1986-yh**; **Turelli1990-kd**). **Santiago1995-hx** assume an equilibrium level of fitness variation $V_A$, with the fitness variance associated with a particular background decaying at a simple geometric decay at rate $1 - \kappa$ each

generation, which is sufficient for background selection. An approximate rate of the decay can be worked out for the particular selective system. In general, this can be a much more complicated function; **Buffalo2019-qs** show that even fluctuating selection can be accommodated.

Now, the associations created in generation 1 and that persist in future generations as (ignoring $S_i''$ terms) follow the pattern

$$
\begin{aligned}
S_2' &= S_1(1-\kappa)(1-r) \\
S_3' &= S_1(1-\kappa)^2(1-r)^2 \\
S_4' &= S_1(1-\kappa)^3(1-r)^3.
\end{aligned}
\tag{17}
$$

Then, the cumulative effect of the fitness associations created in generation 1 can be written as,

$$
S_1 Q_t = S_1 + S_2' + S_3' + \ldots + S_t'
$$
$$
Q_t = 1 + \sum_{i=1}^{t}(1-r)^i(1-\kappa)^i
\tag{18}
$$

or for continuous $t$,

$$
Q_t = 1 + \frac{(1-\kappa)(1-r)\left(1-(1-\kappa)^t(1-r)^t\right)}{\kappa+(1-\kappa)r}
\tag{19}
$$

which as $t \to \infty$, converges to,

$$
Q_\infty = \frac{1}{\kappa+r(1-\kappa)}.
\tag{20}
$$

This shows the long reach of heritable fitness factors in altering allele frequency through the generations. In terms like $\mathbb{E}(S_1 S_2)$, a fraction $S_1'$ of the heritable allele frequency change $S_2$ were *previously* built associations between the neutral allele and a fitness background from $S_1$. As long as the direction of selection is the same, the expected change of this fraction of associations formed in the first generation to later generations 2 and 3 would be,

$$
\begin{aligned}
\mathbb{E}(S_1 S_2) &= \mathbb{E}(S_1^2)(1-\kappa)(1-r) = \mathbb{E}(S_1^2)Q_2^2 \\
\mathbb{E}(S_1 S_3) &= \mathbb{E}(S_1^2)(1-\kappa)^2(1-r)^2 = \mathbb{E}(S_1^2)Q_3^2.
\end{aligned}
\tag{21}
$$

The $Q_t^2$ terms represent the inflation of variance due to autocovariance in frequency change across generations due to weak draft.

With these covariance terms, we now have a full model for $V(p_t - p_0)$. We will look at the case for $t = 2$. As shown in Supplementary Equation (13), the second moment of every term is a function of the variance in neutral allele frequency in that generation, $x_t(1 - x_t)$. This variance does not stay its initial value of $x_0(1 - x_0)$; it decays each generation due to coalescence from the drift and selection processes modeled by this approach. The rate at which this decay happens is the effective population size each generation implied by this model. We can write,

$$\begin{aligned}
V(x_2 - p_0) &= \mathbb{E}(S_1{}^2) + \mathbb{E}(S_1 S_2) + \mathbb{E}(S_2{}^2) \\
&\quad + \mathbb{E}(D_1{}^2) + \mathbb{E}(H_1{}^2) + \mathbb{E}(D_2{}^2) + \mathbb{E}(H_2{}^2) \\
\frac{V(x_2 - p_0)}{p_0(1 - p_0)} &= \frac{V_A}{2N} + \frac{V_A Q_2^2}{2N}\left(1 - \frac{1}{2N}\right) + \frac{V_A}{2N}\left(1 - \frac{1}{2N}\right) \\
&\quad + \frac{V_n}{8N} + \frac{1}{4N} + \frac{V_n}{8N}\left(1 - \frac{1}{2N}\right) + \frac{1}{4N}\left(1 - \frac{1}{2N}\right).
\end{aligned}$$

(22)

In general, the variance in allele frequency change after $t$ generations in a system with heritable fitness due to a single locus $r$ recombination fraction away from the neutral site is given by,

$$\frac{V(x_t - p_0)}{p_0(1 - p_0)} = \sum_{i=1}^{t} \frac{1}{2N}\left(V_A Q_i^2 + \frac{V_n}{4} + \frac{1}{2}\right)\left(1 - \frac{1}{2N}\right)^{i-1}.$$

(23)

We will see later that with the appropriate form for $Q_t^2$, this equation holds for polygenic selection (see Supplementary Section 1.5). Additionally, we note that since the probability of coalescence (or, equivalently, identity by descent) is proportional to the variance in allele frequency change (**Barton2000-zg**), this equation for the variance encodes the pairwise coalescent rate of the population through time under weak draft. Others have found that the genealogies under purifying selection can be characterized by such a time-dependent effective population size (**Nicolaisen2013-gv**).

## 1.4 The Draft-Effective Population Size

Now, we consider how to turn this expression for the variance in neutral allele frequency at time $t$ under weak draft into an expression for the *draft-effective* population size analogous to drift-effective population size given in Supplementary Equation (10). In an ideal Wright–Fisher population, the instantaneous variance is $V(\Delta x_t) = p_{t-1}(1-p_{t-1})/2N_{d,t-1}$. Santiago and Caballero consider the allele frequency change at some $t$ to define their effective population size at time $t$ (which we here are calling draft-effective population size, $N_{d,t}$),

$$N_{d,t} := \frac{p_{t-1}(1 - p_{t-1})}{2V(x_t) - V(x_{t-1})}.$$

(24)

The time-dependency is due to $Q_i^2$, which reflects the fact that autocovariances build up, but at a decreasing rate. Note that levels of pairwise diversity depend on the full distribution of pairwise coalescence times, or the $N_{d,t}$ (e.g. see equation 4 of **Santiago2016-mu** and equation 16 of **Santiago1998-bs**).

Using Supplementary Equation (23) we have,

$$N_{d,t} \approx \frac{4N}{4V_A Q_t^2 + V_n + 2}$$

(25)

(cf. equation 18, **Santiago1995-hx**). Note if $V_A = 0$, this equation reduces to Wright's (**Wright1938-tv**) equation.

In many cases, we care about the asymptotic $N_d$,

$$N_d := \lim_{t \to \infty} N_{d,t} \approx \frac{4N}{4V_A Q_\infty^2 + V_n + 2}. \tag{26}$$

Now, we turn to different expressions for $Q_t^2$ and $V_A$ under different types of polygenic selection.

## 1.5 The Draft-Effective Population Size Under Polygenic Selection

The results of the previous section assume only a single locus a recombination fraction $r$ apart from the neutral locus is determining the coalescence rate. The model of **Santiago1998-bs** extends this to the case where fitness is polygenic. As they show in their original paper, this system can accommodate a variety of different equilibrium selection systems as long as expressions for $V_A$ and $1 - \kappa$ can be worked out. Other authors have extended similar models to more elaborate bbreeding structures (**Wray1990-zf**; **Woolliams1993-qo**).

First, we extend Supplementary Equation (25) to a polygenic system. Throughout, we assume a multiplicative fitness model (i.e. independent effects, no epistasis), where the fitness of individual $i$ is the product of their fitness contributions from each locus $l$, $w_{i,l}$, giving $w_i = \prod_{l=1}^n w_{i,l}$. Then, assuming mean fitness is one and independence between sites, the total fitness variation is multiplicative across the locus-level fitnesses,

$$V_A = \prod_{l=1}^n (1 + V_{h,l}) - 1 \tag{27}$$

Similarly, the $Q_t^2 V_A$ term in the single-locus model is multiplicative under the polygenic model, as the change in site-specific fitness variances are assumed independent,

$$Q_t^2 V_A = \prod_{l=1}^n \left(1 + \frac{V_{h,l} Q_{t,l}^2}{2}\right) - 1 \tag{28}$$

where $Q_{t,l}$ is the cumulative impact of selection due to the selected site $l$ (c.f. equation 4 **Santiago1998-bs**). The factor of $1/2$ is due to the fact that we are ignoring the chance that the fitness background on the homologous chromosome recombines onto the haplotype with the tracked neutral allele (i.e. due to the $Q_t''$ associations). In other words, only half the fitness variation in a diploid can stay associated with the neutral allele through the generations.

To simplify the derivation, assume $V_n = 2$ as under a Wright–Fisher model. Then, the effective population size is,

$$\begin{aligned} N_{d,t} &\approx \frac{N}{V_A Q_t^2 + 1} \\ &\approx \frac{N}{\prod_{l=1}^n \left(1 + \frac{V_{h,l} Q_{t,l}^2}{2}\right)} \\ &\approx N \exp\left(-\sum_{l=1}^n \frac{V_{h,l} Q_{t,l}^2}{2}\right) \end{aligned} \tag{29}$$

9

(c.f. **Santiago1998-bs** equation 4b). If we write each site $l$'s contribution to the variance as a deviation from the average contribution, $V_{h,l}^2 = 1/nV_A + \varepsilon_l$, then

$$N_{d,t} \approx N \exp\left(-\frac{1}{2}\sum_{l=1}^{n}\left(\frac{1}{n}V_h + \varepsilon_l\right)Q_{t,l}^2\right)$$

$$\approx N \exp\left(-\frac{V_h}{2n}\sum_{l=1}^{n}Q_{t,l}^2 + \sum_{l=1}^{n}\varepsilon_l Q_{t,l}^2\right) \tag{30}$$

Santiago and Caballero assume that over evolutionary replicates, as selected sites are randomly distributed, there is not systematic covariance between the fitness variation at a site and the decay of its association with the neutral site, e.g. $\mathbb{E}(\varepsilon_l Q_{t,l}^2) = 0$, and thus ignore the second sum term, leading to

$$N_{d,t} \approx N \exp\left(-\frac{V_h}{2n}\sum_{l=1}^{n}Q_{t,l}^2\right). \tag{31}$$

## 1.6   Continuous Approximation to the Segment Under Selection

Our model differs from that of **Santiago1998-bs** in that we do not integrate over the entire genome. In their 1998 model, Santiago and Caballero consider the impact of both sites under selection on the same chromosome, as well as sites on independently assorting chromosomes. We only consider the contribution of a single segment under purifying selection, and the reduction from additional segments accumulate multiplicatively. Like their model, we imagine the reduction experienced by a focal neutral site in the middle of the segment.

The $Q_{t,l}$ terms depend on $l$ only through the recombination fraction $r(l)$ between the neutral allele and the selected site. Assuming a constant per-basepair recombination rate of $r_{\mathrm{BP}}$, $r(l) = r_{\mathrm{BP}}l$ we can approximate the sum with an integral for the asymptotic $Q_\infty$. The reduction experienced in the middle of a chromosome is twice the reduction experienced from each half,

$$N_{d,\infty} \approx N \exp\left(-\frac{V_h}{2L}\sum_{l=1}^{L}Q_{\infty,l}^2\right)$$

$$\approx N \exp\left(-\frac{V_h}{M}\int_{0}^{M/2}Q_\infty(r)^2 dr\right) \tag{32}$$

where $M = nr_{\mathrm{BP}}$ is the total segment length in Morgans. Then, the total asymptotic effect of associations from all sites as $Q_\infty^2$. Then,

$$Q_\infty^2 = \frac{2}{M}\int_{0}^{M/2}Q_\infty(r)^2 dr$$

$$= \frac{2}{M}\int_{0}^{M/2}\left(\frac{1}{1-(1-r)Z}\right)^2 dr.$$

$$= \frac{2}{(1-Z)(2-(2-M)Z)} \tag{33}$$

10

Santiago and Caballero further approximate this; we can derive their approximation by setting $Z = 1 - \kappa$ and doing a first-order Taylor series expansion around $\kappa$ (since the loss in variance is presumed to be small). Then,

$$Q_\infty^2 = \frac{2}{M(1-Z)} + \frac{2(M-2)}{M^2} + \frac{2\left(M^2 - 4M + 4\right)(1-Z)}{M^3} + \mathcal{O}(\kappa^2) \tag{34}$$

Santiago and Caballero's approximation only keeps this first term, e.g. their $Q_\infty^2 \approx 2/(M(1-Z))$, which is only accurate in the domain $M > 0.2$.

$$N_{d,\infty}^{(SC98)} \approx N \exp\left(-\frac{V_A}{(1-Z)M}\right) \tag{35}$$

(c.f. **Santiago1998-bs** equation 8). For small $M$, the small $\kappa$ approximation strongly deviates from the full equation; thus, throughout, we use the unapproximated version,

$$N_{d,\infty} \approx N \exp\left(-\frac{V_A}{(1-Z)(2-(2-M)Z)}\right). \tag{36}$$

For the reasons described earlier, a single asymptotic $N_{d,\infty}$ cannot fully characterize pairwise coalescent rates in all cases, and thus neutral diversity levels depend on the full $N_{d,t}$, for which we integrate equation (19) to get,

$$N_{d,t} \approx N \exp\left(-\frac{V_A}{M}\int_0^{M/2}\left(1 + \frac{(1-k)(1-r)\left(1-(1-k)^t(1-r)^t\right)}{k+(1-k)r}\right)dr\right). \tag{37}$$

## 1.7 The Variance Dynamics

The heritable variance in fitness $V_A$ associated with the tracked neutral allele does not remain constant through the generations, but decays due to selection and drift, and is maintained by the input of new mutations. For an arbitrary polygenic selective system, the dynamics are extraordinarily complicated due to the complexity of multilocus selection, and the dynamics of the trait response to selection depend on the variance, which in turn depends on higher moments of the fitness distribution (the "moment-closure problem"). We discuss the simplification of Santiago and Caballero's model here.

Each generation, the genetic variance changes due to changes in the allele frequencies at selected sites and linkage disequilibria between these sites. These are typically expressed as recursions, but certain models permit these changes to be expressed as a proportional reduction (e.g. truncation selection and stabilizing selection models; **Keightley1988-eq** p. 36, and **Walsh2018-bt** p. 557). Santiago and Caballero's model follows this approach, assuming the variance decays due to selection and drift at a constant rate $Z := 1 - \kappa$ through the generations, such that the variance in the next generation excluding mutation is $V_A' = (1-\kappa)V_A$ and $\Delta V_A = -\kappa V_A$.

With mutation, $\Delta V_A = V_m - \kappa V_A$. The model of Santiago and Caballero assumes that under the long-run equilibrium, $\Delta V_A = 0$, such that $\kappa V_A = V_m$, and thus the rate of loss is $\kappa = {}^{V_m}/_{V_A}$. Then the variance for the dynamics simplify to,

$$V'_A = \left(1 - \frac{V_m}{V_A}\right) V_A$$

$$Z := \frac{V'_A}{V_A} = 1 - \kappa = 1 - \frac{V_m}{V_A} \tag{38}$$

(c.f. **Santiago1998-bs** equation 9).

## 1.8 The Draft-Effective Population Size Under Strong Background Selection

Under strong background selection, multiple segregating deleterious sites contribute to fitness variation. BGS models assume a multiplicative fitness model, such that the fitness of an individual is

$$w = (1 - s)^n \tag{39}$$

where $n$ is the number of deleterious mutations this individual carries.

Classic background selection theory assumes an infinite population size, such that mutation frequencies are at their mutation-selection equilibrium and the expected number of mutations across individuals is $\mathbb{E}(n) = \mu/sL$ where $L$ is the total number of basepairs that can be mutated (the mutational target size). Under a multiplicative fitness function in an infinite population, selection cannot generate linkage disequilibria (**Turelli1990-kd**). Thus the classic strong background selection model assumes additive genetic fitness is equal to the additive genic fitness (which excludes the contribution of covariance between selected sites),

$$V_{\text{BGS}} := \text{V}(w) = 2s^2 \sum_{l=1}^{L} p_l(1 - p_l)$$

$$= 2Ls^2 \frac{\mu}{s}\left(1 - \frac{\mu}{s}\right) = Us + \mathcal{O}(\mu^2) \approx Us. \tag{40}$$

since under mutation-selection balance $p_l = \mu/s$ and $U$ is defined $U := 2L\mu$, the deleterious mutation rate per diploid genome, per generation. They then set $V_A = V_{\text{BGS}}$, and note that $V_m \approx Us^2$, since mutations have frequency of $1/2N$ and each increase the variance by a factor of $s^2$. Note that because the classic BGS model assumes deterministic selection dynamics (i.e. an infinite population), the equilibrium $Z$ given by equation (38) excludes the reduction due to drift. The reduction factor under the strong BGS model is,

$$Z_{\text{BGS}} := 1 - \frac{Vm}{V_{\text{BGS}}} = 1 - \frac{Us^2}{Us} = 1 - s \tag{41}$$

(c.f. **Santiago1998-bs** equation 10). An alternate derivation can be found by working out the deterministic reduction in variance from the single-locus dynamics.

Now, we can use these values for $Z_{\text{BGS}}$, $V_m$, and $V$ in equation (36). Through this alternate quantitative genetics derivation, we arrive at the same equation as classic BGS theory,

$$N_{d,\infty}^{\mathrm{BGS}} = N \exp\left(-\frac{U}{2s + M - Ms}\right). \tag{42}$$

This matches equation 8 of **Hudson1995-xc** except for the $Ms$ term and equation 10 of **Nordborg1996-nq** except for the $2s$ term.

## 1.9 The Draft-Effective Population Size Under Weak and Strong Negative Selection

With the possibility of weak selection, deleterious alleles can drift up in frequency and fix. Consequently, the fitness distribution is no longer stationary, as the set of haplotypes without any deleterious mutations can be lost due to drift. This changes the shape of the fitness distribution, and thus the variance parameter that our model relies upon (**Gessler1995-hz**; **OFallon2010-my**; **Good2013-lp**; **Haigh1978-gt**; **Higgs1995-xc**). So far, finding an explicit equation for the variance has been difficult due to the "moment-closure problem" previously mentioned. In their approximation, **Santiago2016-mu** consider a modified additive genic variance that is below $V_{\mathrm{BGS}}$ due to the loss of variation from each fixation of a weakly deleterious mutation. This fixation process is related to the Muller's Ratchet, the rate $R$ of fixation of deleterious mutations per generation per region, since each fixation directly reduces the additive genic variance of the fitness distribution. However, in general, the rate of this ratchet is unknown since it depends on the fitness variance (and thus higher moments of the fitness distribution).

In their 2016 paper, Santiago and Caballero derive use an approach based on Fisher's Fundamental Theorem of Natural Selection (equation 2, **Santiago2016-mu**, equation 15A **Garcia-Dorado2007-jj**). A more formal derivation follows from **Higgs1995-xc** (though they assume a haploid asexual system, the end result is the same up to a factor of two). In their work, they derive equations for the change per generation for the moments and cross-moments of the fitness distribution under multiplicative selection in a haploid model. They find, the haploid rate of the ratchet is $r = u - sv_n$ (equation 13.3, **Higgs1995-xc**), where we use lowercase to distinguish between haploid and diploid models. Here, $v_n$ is the variance in the *number* of deleterious mutations; as long as selection is not too strong, the haploid additive variance is $v_A \approx s^2 v_n$. Substituting this for $v_n$ and rearranging, the haploid fitness variance is related to the $v_A = us - rs$, which is identical to Santiago and Caballero's (**Santiago2016-mu**) equation 2, noting that the waiting time between fixations (i.e. ratchet clicks) is $T = 1/r$. Intuitively, the fitness variance $V_{\mathrm{BGS}} = Us$ is reduced by the rate at which deleterious mutations fix and remove fitness variation from the population.

Now, note that under a diploid model, the fitness variance is twice that of the haploid model, so $V_A = 2v_A$. Then, the fitness variance is,

$$V_A = Us - 2Rs. \tag{43}$$

where $U = 2u$ and $R = r$ is the diploid rate of the ratchet, which depends on higher moments of the fitness distribution, but only differs from $r$ through a rescaling of the population size. We also find that Equation (6) and Supplementary Equation 43 above can also be found through Kimura's diffusion models with a flux of mutations into infinite, discrete sites (**Kimura1969-jw**). Note that if selected sites are sufficiently deleterious that they cannot fix (or the population is infinite),

$R = 0$ and the additive fitness variance is identical to that under strong background selection, $V_A = Us = V_{\text{BGS}}$ (equation 40). We note that while we use $V_A$ here following **Santiago2016-mu**, this is only the *additive genic* component of additive fitness variation, as our simulations confirm in the main text.

Next, we turn to the fixation probability of a new deleterious mutation with additive effects ($1 - s$ in the heterozygote, $1 - 2s$ in the homozygote). In a diploid population with drift-effective population size $N_e$ this is,

$$p_{\text{fix}} := \frac{1 - \exp(2N_e s/N)}{1 - \exp(4N_e s)} \approx \frac{2N_e s}{N\left(\exp(4N_e s) - 1\right)} \tag{44}$$

(c.f. **Durrett2008-ql**, equation 7.21; **Kimura1957-rk**, equation 5.6). Then, Santiago and Caballero argue that the rate of the ratchet is the inverse of the averaging waiting time until a fixation, $T \approx 1/(UN p_{\text{fix}})$ (note that since $U$ is the *diploid* mutation rate, the total population mutation rate is $UN$ per generation), so

$$T \approx \frac{\exp(4N_e s) - 1}{2UsN_e}. \tag{45}$$

Since the ratchet rate depends on $N_e$, which in turn depends on the rate of the ratchet, we write the ratchet, or substitution, rate per $L$-basepair region per generation as,

$$R(N_e) := \frac{2UsN_e}{\exp(4N_e s) - 1}. \tag{46}$$

Under weak and strong background selection the variance input of mutations $V_m$ is the same, and using the $V_A$ derived above, the decay in variance is

$$Z_{\text{WS}} = 1 - \frac{V_m}{V_A} = 1 - \frac{Us^2}{(U - R(N_e))s} = 1 - \frac{Us}{U - R(N_e)}. \tag{47}$$

At equilibrium, the $N_e$ that determines fixation probability is the asymptotic $N_{d,\infty}$. Consequently, Santiago and Caballero argue that the asymptotic reduction in draft-effective population size uses the $Q_\infty^2$ and is given by the following system of non-linear equations,

$$R(N_{d,\infty}) = \frac{4UsN_{d,\infty}}{\exp(4N_{d,\infty}s) - 1} \tag{48}$$

$$N_{d,\infty} \approx N \exp\left(-\frac{V_A}{(1 - Z)(2 - (2 - M)Z)}\right) \tag{49}$$

The solution to these equations is the asymptotic equilibrium $\widetilde{N}_d$ and equilibrium ratchet rate $\widetilde{R}$. This sets the equilibrium variance, $\widetilde{V} = Us - 2\widetilde{R}s$, or $\widetilde{V} = (U - 2\widetilde{R})s$. Note that because $p_{\text{fix}} < 1/2N$ for deleterious variation, $2R < U$, or the diploid mutation rate must be greater than or equal to twice the per-basepair per-generation substitution rate for equilibrium variance $\widetilde{V} > 0$ for $s > 0$.

# 2 Statistical Methods

The previous section outlined the basis of quantitative genetic linked selection models that predict the draft-effective population size under polygenic selection and linkage. Our work here only considers the contribution of strongly and weakly deleterious mutations to fitness variation. In this section, we outline more details about the statistical models used to construct the genome-wide B and B' maps that are used in model fitting.

## 2.1 Parametric Forms of Negative Linked Selection Models

Both the classic BGS and the SC16 model we extend have the same parametric form, since both (1) model the impact of linked selection as a rescaling of effective population size, and (2) both assume multiplicative fitness across segments. Thus, the total reduction experienced by a neutral site at position $x$ due to segments across the genome is the product of the individual reductions,

$$
\begin{aligned}
B(x, \Psi) &= \prod_{g \in G} \exp \left( \int_0^1 b(\mu(s, k(g)), s, L_g, r_g, d(x, g)) ds \right) \\
&= \exp \left( \sum_{g \in G} \int_0^1 b(\mu(s, k(g)), s, L_g, r_g, d(x, g)) ds \right)
\end{aligned}
\tag{50}
$$

where, $b(\mu(s, k(g)), s, L_g, r_g, d(x, g))$ is the log reduction due to segment $g$, and there are $S$ total segments. Here, $\mu(s, k(g))$ is the per-basepair per-genome rate that mutations with selection coefficient $s$ enter segments with feature class $k(g)$ (e.g. it incorporates the distribution of fitness effects and mutation rate), $L_g$ and $r_g$ are the length in basepairs and recombination rate per basepair of segment $g$, and the recombination distance between focal site $x$ and segment $g$ is $d(x, g)$. The functional form of $b$ varies depending on whether the classic BGS model is used (Supplementary Section 2.3), or our form of Santiago and Caballero's (**Santiago2016-mu**) equations (see Supplementary Section 2.4).

Under both negative selection models implemented in our software, the diversity in window $v$ is $\bar{\pi}(v, \Psi) = \bar{B}(v, \Psi) \pi_0$. Here, $\bar{B}(v|\Psi)$ is the predicted reduction in diversity due to BGS in window $v$, given background selection parameters $\Psi$. In practice, we pre-calculate $B(x|\Psi)$ at fixed sites $x$ across the genome, and take the average of the fixed sites $B$ values within window $v$ for the average $\bar{B}(v|\Psi)$.

## 2.2 Discretization of Parameters and Annotation Feature Classes

In practice, to fit negative selection models, we must discretize both $\mu(s, k(g))$ and $s$ in Supplementary Equation (50). We define $\mu_{i,k(g)}$ to be the rate of mutations entering the population per basepair per generation with selection coefficient $s_i$ for the feature class $k(g)$ of segment $g$. This $\mu_{i,k(g)}$ term reflects the product of the mutation rate and the distribution of deleterious fitness effects for selection coefficient $s_i$ and feature class $k(g)$. Then, the reduction at position $x$ is,

$$
B(x, \Psi) = \exp \left( \sum_g^S \sum_i^{m_s} b(\mu_{i, k(g)}, s_i, L_g, r_g, d(x, g)) \right).
\tag{51}
$$

15

Because it is computationally-intensive to numerically solve the system of two non-linear equations to calculate $b$ at each segment, these are pre-computed for an $n_s$-element grid of selection coefficients and an $n_\mu$-element grid of mutation rates–DFE products. Likewise, when we calculate $b$ under classic BGS models, we pre-calculate segment components across these same grids, and combine them for each $x$.

The number of annotation features $K$ is variable and set by the annotation data specified. There are two ways to parameterize the mutation rates and DFE for all annotation classes in background selection models. First, for the *free-mutation* parameterization, each annotation feature class has a free mutation rate parameter for each of these selection coefficients (as in Supplementary Equation 50). We experimented with this approach, but ultimately found its results less intuitive and it has more degrees of freedom.

Instead, we opted for a *simplex* parameterization, which has a single shared mutation rate across all features, so the mutation parameter $\mu_{i,\,k(g)}$ above is given by the product of the mutation rate and the DFE weight matrix. For example, for a $\log_{10}$ selection grid and three features, CDS, UTRs, and PhastCons (PC), we'd have a DFE matrix like,

$$\mathbf{M} = \mu\mathbf{W} = \begin{bmatrix} w_{10^{-6},\text{CDS}} & w_{10^{-6},\text{UTR}} & w_{10^{-6},\text{PC}} \\ w_{10^{-5},\text{CDS}} & w_{10^{-5},\text{UTR}} & w_{10^{-5},\text{PC}} \\ \vdots & \vdots & \vdots \\ w_{10^{-1},\text{CDS}} & w_{10^{-1},\text{UTR}} & w_{10^{-1},\text{PC}} \end{bmatrix} \tag{52}$$

which has $2 + K \times (n_s - 1)$ free parameters, since each column must sum to one. This imposes the constraint that $\sum_i w_{i,k} = 1$, requiring constrained MLE optimization. We implement this using softmax, which describe further in Supplementary Section 5.1.

## 2.3 Reductions Under the Classic BGS Model

The classic BGS model, which assumes deleterious mutations cannot fix, expresses the reduction for a focal site in the middle of a segment (**Hudson1995-xc**; **Hudson1994-oh**; **Nordborg1996-nq**); here we consider a focal site directly to an arbitrary side of a segment under a deleterious mutations, and extend the model to handle a focal segment arbitrarily far from the segment. To simplify notation, we will only consider a single feature class in this and the next section. The log reduction under the classic mutation-selection-balance BGS model is,

$$b(\mu, s, L, r, 0) = -\int_0^L \frac{\mu(l)}{s(1 + (1-s)r(l)/s)^2} dl \tag{53}$$

(c.f. **McVicker2009-ax** p. 11, **Hudson1995-xc** equation 5,), where $\mu(l)$ is the *haploid* per-basepair mutation rate at $l$ and $r(l)$ is the recombination fraction between the focal site (at position 0) and basepair $l$. Assuming constant per-basepair recombination and mutation rates, $r(l) = r_{\text{BP}}l$ and $\mu(l) = \mu$,

$$b(\mu, s, L, r_{\text{BP}}, 0) = -\frac{\mu L}{s + (1-s)r_{\text{BP}}L} \tag{54}$$

16

(c.f. **Hudson1995-xc** equation 8; their model considers a neutral site in the middle of the segment and drops small terms). Often per-region rates are defined $U = 2\mu L$ and $M = r_{\mathrm{BP}}L$ are the segment-wide mutation diploid mutation rates (mutations per diploid segment per generation) and recombination length (Morgans).

For computational efficiency, we adjust this model by setting $r(l) = d + r_{\mathrm{BP}}l$ in Supplementary Equation (53) and integrating, where $d$ is the recombination fraction between the focal site $x$ and the start of segment $g$. This allows us to pre-compute the local effects of the segment, and substitute in $b$ as the focal site changes. This integral substitution has a closed form,

$$b(\mu, s, L, r_{\mathrm{BP}}, d) = -\frac{\mu L}{(d(1-s)+s)(s+(d+r_{\mathrm{BP}}L)(1-s))}. \tag{55}$$

Then, terms can be collected in powers of $d$ and pre-computed for all segments for the grid of $\mu$ and $s$. Our implementation of classic BGS in `bgspy` pre-computes these segment components, and then calculates the final $b$ value for an $x$ and $g$ by calculating the distance $d(x,g) = |M(x) - M(g)|$, where $M(x)$ is the cumulative recombination map length at position $x$ in Morgans. Since $r(l)$ could vary by position, segments with differing recombination rates are split in two of the same annotation class by their recombination rate. This has no effect on the calculation due to the multiplicative fitness, but ensures more accurate estimates of the reduction factor $B$.

## 2.4 Reduction Under the Modified Santiago and Caballero Model

Under Santiago and Caballero's model, the correct effective population size to gauge reduction in heterozygosity or coalescent times is given by their equation 4 (**Santiago2016-mu**). This equation considers how the inflation factor $Q_t^2$ that sets the draft-effective population size varies over $t$. This is because selective processes can lead to non-constant pairwise coalescent rates, which impact pairwise coalescent rates and thus the reduction factor. We experimented with calculating this full sum and integral approximations to it in our methods, but it seemed to make little difference in our application. Additionally, it was too costly to calculate computationally for genome-wide calculations of the B' map. Instead, we solve for the asymptotic $\widetilde{N}_{e,\infty}$ for each segment, and use that to calculate the log reduction factor under the SC16 model, $b'$.

Note that the asymptotic model above determines the reduction in effective population size experienced by a focal site in the middle of a $L$-basepair segment, but B maps require the reduction $b'$ experienced by a focal site $x$ at an arbitrary recombination distance apart. For each segment, we pre-compute the solution to the system of two equations in Supplementary Equation (48) for the mutation and selection coefficient grids, giving us $\widetilde{V}_a$ and $\widetilde{Q}_\infty^2$ for each set of parameters. Then,

$$b'(\mu, s, L, r, d) = -\frac{\widetilde{V}_a(\mu, s, r)}{2} Q^2(\mu, s, r) \tag{56}$$

$$Q^2(\mu, s, r) = \left(\frac{1}{(1-(1-d)\widetilde{Z})}\right)^2 \tag{57}$$

where $\widetilde{Z} = 1 - Us/(U - \widetilde{R})$ and $\widetilde{V}_a = Us - 2\widetilde{R}s$. As before, $d$ is the recombination fraction between the segment and the focal neutral site.

## 2.5 Data Summary Matrices

Our underlying data for all likelihood and pairwise diversity estimates is the allele count matrix $\mathbf{C}$ with dimensions $L \times 2$, where $L$ is the chromosome length. This is transformed to a pairwise summary matrix with identical dimensions, $\mathbf{Y}$. The first column of $\mathbf{Y}$ is the number of pairwise comparisons between chromosomes that are identical, and the second column is the number that are different. Both of these columns are combinatoric summaries of the raw allele counts matrix needed for the binomial likelihood and pairwise diversity estimates. Let $[c_1, c_2]$ be a row of $\mathbf{C}$ for basepair $l$ (the $l$ index is omitted for clarity), and $n = c_1 + c_2$. Then, the (1) total number of pairwise combinations of chromosomes $n_T$, (2) the number of pairwise with identical alleles $n_S$, and (3) the number of pairwise combinations with differing alleles $n_D$ are respectively,

$$n_T = \frac{n(n-1)}{2}$$
$$n_S = \binom{c_1}{2} + \binom{c_2}{2}$$
$$n_D = n_T - n_S$$

which would be stored in row $\mathbf{Y}_l = [n_S, n_D]$. Note that the per-site $\mathbf{Y}$ handles non-polymorphic sites and missing data. Non-polymorphic sites have $n_S = \binom{n}{2}$ and $n_D = 0$, and missing data has $n_S = n_D = 0$.

## 2.6 Pairwise Diversity Estimates

The pairwise diversity at site $l$ across the $n$ sampled chromosomes can be calculated from row $l$ of the $\mathbf{Y}$ matrix as follows,

$$\pi_l = \frac{n_D}{n_T} \tag{58}$$

which is identical to the more common expression of this estimator,

$$\pi_l = \frac{2}{n(n-1)} \sum_{i<j}^{n} k_{i,j} \tag{59}$$

where $k_{i,j}$ is 1 if the alleles at this site differ at site $l$, and 0 otherwise.

There are three ways to aggregate $\pi_l$ across all sites. The first is,

$$\pi^{(1)} = \frac{1}{L} \sum_{i=1}^{L} \frac{n_{D,i}}{n_{T,i}} \tag{60}$$

which if the number of samples across loci is constant, simplifies to an unweighted average across sites. Second, one can take a weighted average, with weights determined by the total number of samples present at a site,

$$\pi^{(2)} = \frac{1}{\sum_{i=1}^{L} n_i} \sum_{i=1}^{L} n_i \frac{n_{D,i}}{n_{T,i}} \tag{61}$$

Third, one can weight by the number of pairwise comparisons at a site, $n_{T,i}$, rather than total number of samples, $n_i$, which leads to a ratio of sums,

$$\pi^{(3)} = \frac{\sum_{i=1}^{L} n_{D,i}}{\sum_{i=1}^{L} n_{T,i}}. \tag{62}$$

We predominantly use the estimator $\pi^{(3)}$, as it corresponds to how we summarize the matrix $\mathbf{Y}$ across windows for our likelihood. All methods have mean squared errors and biases very close to one another.

Note, however, that estimates of pairwise diversity often condition on the accessible bases, and thus treat this as fixed. However, the number of accessible bases varies across the chromosome; this can lead to a source of apparent bias during block-bootstrap estimates of uncertainty. In this case, pairwise diversity is a ratio estimator, and is thus biased, since by Jensen's inequality $\mathbb{E}(y/x) \geq \mathbb{E}(y)/\mathbb{E}(x)$ for random variables $x$ and $y$.

## 2.7 Ratchet Rate Prediction

After rescaling each segment $N_e$ by the local predicted reduction $\hat{B}(x)$, the ratchet rates and $B'$ values are re-calculated.

The rescaled $B'$ calculation outputs a $n_\mu \times n_s \times S$ multidimensional array $\mathbf{R}$ of ratchet rates per segment. These are rescaled by the segment lengths $\mathbf{l}$, giving the per-basepair mutation rate array. The per-basepair ratchet rate given maximum likelihood estimates of $\hat{\mu}$ and $\widehat{\mathbf{W}}$ is used to predicted the ratchet rate for each segment $g$. The DFE estimate for feature $k$ is a column vector of $\widehat{\mathbf{W}}$, i.e. $\widehat{\mathbf{w}}_k$

$$\lambda_D(g) = \tag{63}$$

The predicted ratchet rate for segment $g$ $\lambda_d(g)$ is based on these values.

## 2.8 Numeric Optimization

mu bounds
  bounds and time, bounds and convergence
  Same optima, but took longer

## 2.9 Numeric methods for calculating $B$

Depending on the number of features, this can be incredibly memory intensive.

## 2.10 Optimization

Optimization in non-linear regression is notoriously difficult both in general (Bates) and in the specific problem of estimating the effects of genome-wide selection (**Murphy2022-sj**). We experimented with a variety of global and local optimization approaches and parameterizations, but found that

# 3 Human Genomic Data

In order to try to ensure accurate estimation of pairwise diversity, which is a ratio estimator that is sensitivity to its denominator, we use the complete per-basepair genotype calls (gVCF) produced by Illumina's DRAGEN pipeline (**Illumina˙Inc2020-dk**). The original samples were sequenced to 30x by the New York Genome Center (**Byrska-Bishop2022-tn**). This allows for filtering to be applied to the entire genome at once, rather than just variants, so the denominator does not need to be estimated separately.

The full list of samples is available in TSV format in the GitHub repository (`data/h1kg/yri_samples.tsv`).

XXX

## 3.1 Filtering gVCFs

gVCFs were filtered using a custom Python tool, `gvcf2counts.py` (in `tools/gvcf2counts.py`), which reads the gVCFs, filters them according to the criteria below, and outputs a Numpy `.npy` file of reference and alternative allele counts for each chromosome (hereafter, "allele counts matrices").

Genotypes are included in the allele count if and only if:

1. The variant call is set to `PASS` in the VCF.

2. The `QUAL > 50`.

3. The `GQ > 30` (or `RGQ > 30` for invariant sites).

Because the data underlying the counts files are per-basepair resolution gVCFs, each chromosome's allele counts matrix is of size $l \times 2$, where $l$ is the total chromosome length. Basepairs that fail these filtering requirements lead to a row of zero counts, e.g. no observed reference and alternative allele counts, and thus do not effect the data that goes into the binomial likelihood or $\pi$ estimates used in figures.

## 3.2 Non-accessible and Masked Regions

The allele counts matrices include many basepairs that may have allele counts that pass the genotype call filters, but are still need to be filtered out because the region of the genome may produce unreliable estimates. The following filters are applied based on masking regions:

1. **Non-accessible regions**: masks out centromeres (`acen` entries in `cytoBand.txt`), with 5 Mbp padding on either side. The file of passing masks is `data/annotation/no_centro.bed`.

2. **Reference masking**: soft and hard-masked regions in the human GRCh38 reference genome are also masked. Soft-masked regions were determined by Ensembl (**Cunningham2022-vk**), which uses Repeat Masker (**Smit2015-nm**) and Dust (**Morgulis2006-uo**). We also only used sites that pass the 1000 Genomes strict filter (**1000˙Genomes˙Project˙Consortium2015-mi**).

## 3.3 Putatively Neutral Tracks
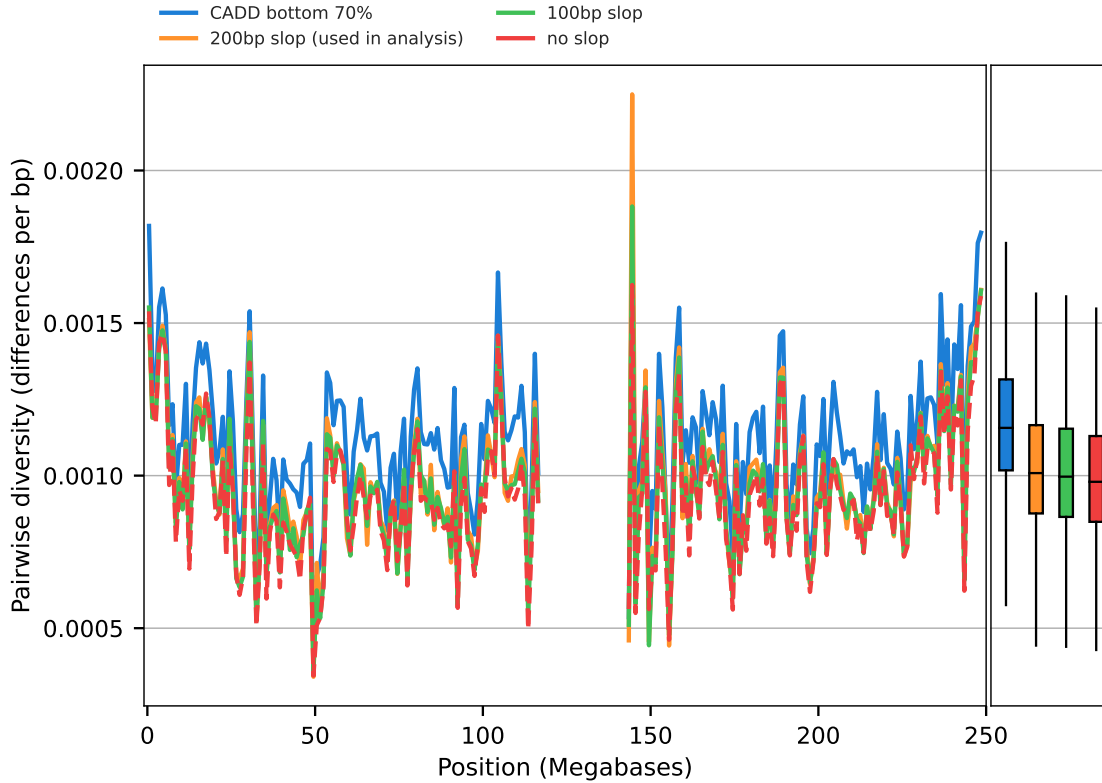


Figure 1: Estimates of chromosome 1 YRI diversity in non-overlapping megabase windows, under different neutral tracks (colors) and the filtering criteria described in Supplementary Section 3.2. To the right of the chromosome subfigure are the genome-wide boxplots of megabase-scale diversity.

Linked selection theory typically assumes that focal sites are not under purifying selection themselves. Consequently, linked selection statistical methods often only consider the fraction of the genome not *a priori* under selective constraint (**Murphy2022-sj**). Our methods follow suit, and first create a track consisting of the union (i.e. `bedtools merge`) of: exons and coding sequences from Ensembl Version 107 (**Cunningham2022-vk**) and PhastCons regions (**Siepel2005-wh**). Then, we remove 200 basepairs of "slop" around each of these regions (i.e. `bedtools slop`). We show in Supplementary Figure 1 that average diversity (we only consider YRI samples for this analysis) is fairly insensitive to the slop width, but increasing slop widths around putatively conserved segments does lead to higher diversity estimates. This is expected, since increasing the slop width considers a smaller subset of sites, further away from putatively conserved regions. For

our analyses we choose the 200bp slop version, which covers 63.66% of the autosomal reference sequence and consists of 1.26 million contiguous genomic regions.

Additionally, we compared our neutral track (and the slop width alternatives) against the bottom 70% of CADD site scores (this works out to be 64.70% of the autosomal reference sequence). The correlation between these two tracks is extraordinarily high (Pearson $r = 0.98$, p-value $= 0$), but the average pairwise diversity is 13% higher (two-tailed t-test p-value $1.29 \times 10^{-71}$) in the CADD bottom 70% neutral track. However, we did not use the CADD bottom 70% neutral track in our analysis of human population genomic data. Our concern was that because CADD models are trained by differentiating high-frequency derived SNPs in humans from simulated variants at every basepair, there could be circularity in basing our neutral regions on predictions based on modeling frequency data. Moreover, this neutral track would be close to a complement of the CADD 6% track, which we use as one of our annotation models, and might lead to issues. Additionally, the CADD bottom 70% neutral track is much less sparse than the track we used in our analysis, with a total of $\approx 20$ times more contiguous regions than the neutral track used in analysis (25.38 million total). Many of these ranges are at very small spatial scales (average: $\approx 7$ basepairs).

Additionally, we ran one test model with the CADD bottom 70% as the neutral track as a test. Using the CADD bottom 70% neutral track with the CADD 6% annotation model, the in-sample was $R^2 = 63.99\%$; by comparison, the in-sample $R^2 = 68.03\%$ using the constructed neutral track with 200bp slop. Overall, future work calibrating the neutral track is needed.

These neutral tracks are all produced by the Snakemake file `data/fit_annotation/Snakefile` in this study's GitHub repository (`http://github.com/vsbuffalo/bprime`). We summarize the percentage of basepairs after masking per chromosome in Supplementary Table **??**.

## 3.4 Window-based Summaries and filtering

Based on exploratory data analyses, we noticed some rare windows that were outliers. Given that maximum likelihood estimation can be easily biased by outliers, we took a conservative approach and removed the top 0.5% of high-diversity windows. The results of this can be seen in Supplementary Figure 2. We also remove 1cM from each end of the recombination map. These window-based filters are in addition to the accessible and neutral sequence masked described in Supplementary Sections 3.1 and 3.2.

## 3.5 Annotation Models

CADD 6%: 6.49% of autosomal sequence. CADD 8%: 8.56% of autosomal sequence.

## 3.6 Model Comparisons

## 3.7 Table of $R^2$ Values

## 3.8 Derivation of Approximate Coalescence $R^2_{\text{coal}}$

Our approach models the observed megabase-scale diversity $y_i$ in window $i$ as the predicted level under our negative selection model, $\pi_0 B_i$, plus some random residual error $\varepsilon_i$,

$$y_i = \pi_0 B_i + \varepsilon_i \tag{64}$$

Table 1: Accessible and putatively neutral sequence mask statistics, for the criteria used in the main model fits.

| Chromosome | Accessible (%) | Neutral (%) | Intersection (%) | Total Basepairs ($\times 10^6$) |
|---|---|---|---|---|
| 1 | 39.7 | 62.0 | 15.7 | 39.09 |
| 2 | 45.3 | 60.1 | 18.9 | 45.77 |
| 3 | 43.2 | 59.6 | 16.9 | 33.51 |
| 4 | 42.7 | 65.9 | 20.2 | 38.42 |
| 5 | 42.4 | 62.2 | 18.1 | 32.86 |
| 6 | 43.8 | 61.4 | 18.4 | 31.43 |
| 7 | 41.9 | 64.0 | 18.6 | 29.64 |
| 8 | 42.3 | 66.1 | 20.2 | 29.32 |
| 9 | 36.2 | 65.4 | 14.6 | 20.21 |
| 10 | 42.9 | 62.2 | 18.8 | 25.15 |
| 11 | 41.3 | 61.5 | 17.0 | 22.96 |
| 12 | 40.6 | 60.4 | 15.9 | 21.19 |
| 13 | 38.8 | 70.7 | 18.8 | 21.50 |
| 14 | 36.5 | 64.6 | 13.8 | 14.77 |
| 15 | 34.4 | 66.3 | 13.5 | 13.77 |
| 16 | 36.0 | 66.0 | 15.5 | 14.00 |
| 17 | 37.2 | 58.8 | 14.2 | 11.82 |
| 18 | 40.3 | 66.1 | 19.6 | 15.75 |
| 19 | 28.1 | 67.5 | 11.5 | 6.74 |
| 20 | 36.1 | 62.5 | 15.4 | 9.92 |
| 21 | 29.8 | 76.9 | 17.3 | 8.08 |
| 22 | 25.5 | 75.5 | 12.6 | 6.40 |
| weighted average | 40.1 | 63.6 | 17.1 | - |

Table 2: $R^2$ and mutation rate estimates for all primary models (top table) and models fit on a smaller DFE grid (excluding the possibility of strong selection $s = 10^{-1}$). Note that the repeated values in the last rows are correct and due to rounding; see `notebooks/main_fits.ipynb` in the GitHub repository for more information.

| Model | Track type | Pop. | B' $R^2_{\text{LOCO}}$ | B' $R^2_{\text{IS}}$ | B $R^2_{\text{IS}}$ | B' $\hat{\mu} \times 10^{-8}$ | B $\hat{\mu} \times 10^{-8}$ |
|---|---|---|---|---|---|---|---|
| **phastcons>CDS>genes** | sparse | YRI | **68.45** | 68.16 | 65.23 | 1.8636 | 0.2975 |
| phastcons>CDS>genes | full | YRI | 68.19 | 68.12 | 63.67 | 1.7082 | 0.1666 |
| CADD 6% | full | YRI | 68.08 | 68.08 | 62.96 | 1.4612 | 0.1721 |
| CADD 6% | sparse | YRI | 68.04 | 68.03 | 68.01 | 2.0912 | 2.0716 |
| CADD 8% | full | YRI | 66.98 | 67.11 | 63.61 | 1.0126 | 0.1713 |
| CADD 8% | sparse | YRI | 66.67 | 67.01 | 66.98 | 1.5688 | 1.5473 |
| CDS>genes>phastcons | full | YRI | 64.90 | 65.51 | 61.23 | 4.1951 | 0.1620 |
| CDS>genes>phastcons | sparse | YRI | 64.90 | 65.51 | 63.16 | 4.1795 | 0.3070 |
| **phastcons>CDS>genes** | sparse | CEU | **61.98** | 63.02 | 60.02 | 2.0194 | 0.3036 |
| phastcons>CDS>genes | full | CEU | 61.78 | 63.01 | 58.57 | 2.3499 | 0.1679 |
| CADD 6% | full | CEU | 61.74 | 62.94 | 57.42 | 1.5365 | 0.1739 |
| CADD 6% | sparse | CEU | 61.66 | 62.86 | 62.84 | 2.1215 | 2.1084 |
| CADD 8% | full | CEU | 60.80 | 62.12 | 57.95 | 1.1572 | 0.1742 |
| CADD 8% | sparse | CEU | 60.59 | 62.06 | 62.04 | 1.6030 | 1.5876 |
| CADD 6% | full | CHB | 59.62 | 61.31 | 55.31 | 1.5083 | 0.1727 |
| CADD 6% | sparse | CHB | 59.54 | 61.21 | 61.20 | 2.1242 | 2.1170 |
| **phastcons>CDS>genes** | sparse | CHB | **59.44** | 61.08 | 57.88 | 2.1982 | 0.3060 |
| phastcons>CDS>genes | full | CHB | 59.29 | 61.08 | 56.71 | 2.5573 | 0.1694 |
| CADD 8% | full | CHB | 58.88 | 60.63 | 55.86 | 1.1128 | 0.1733 |
| CADD 8% | sparse | CHB | 58.65 | 60.57 | 60.55 | 1.6098 | 1.6001 |
| CDS>genes>phastcons | sparse | CEU | 58.58 | 60.30 | 58.41 | 4.1041 | 0.3110 |
| CDS>genes>phastcons | full | CEU | 58.58 | 60.30 | 56.75 | 4.1001 | 0.1648 |
| CDS>genes>phastcons | sparse | CHB | 56.48 | 58.73 | 56.60 | 3.7195 | 0.3136 |
| CDS>genes>phastcons | full | CHB | 56.48 | 58.73 | 54.95 | 3.7552 | 0.1654 |
| *constrained DFE grid* | | | | | | | |
| **phastcons>CDS>genes** | full | YRI | **67.27** | 67.36 | | 2.0333 | |
| CADD 6% | sparse | YRI | 66.66 | 66.81 | | 1.6565 | |
| CADD 6% | full | YRI | 66.50 | 66.68 | | 1.5556 | |
| CDS>genes>phastcons | full | YRI | 63.61 | 64.28 | | 3.3493 | |
| **phastcons>CDS>genes** | full | CEU | **61.04** | 62.25 | | 1.9128 | |
| CADD 6% | sparse | CEU | 60.37 | 61.80 | | 1.6797 | |
| CADD 6% | full | CEU | 60.24 | 61.77 | | 1.6347 | |
| **phastcons>CDS>genes** | full | CHB | **58.76** | 60.45 | | 1.8691 | |
| CADD 6% | sparse | CHB | 58.67 | 60.49 | | 1.7454 | |
| CADD 6% | full | CHB | 58.48 | 60.40 | | 1.5981 | |
| CDS>genes>phastcons | full | CEU | 57.24 | 59.11 | | 3.0461 | |
| CDS>genes>phastcons | full | CHB | 55.63 | 57.88 | | 3.1723 | |
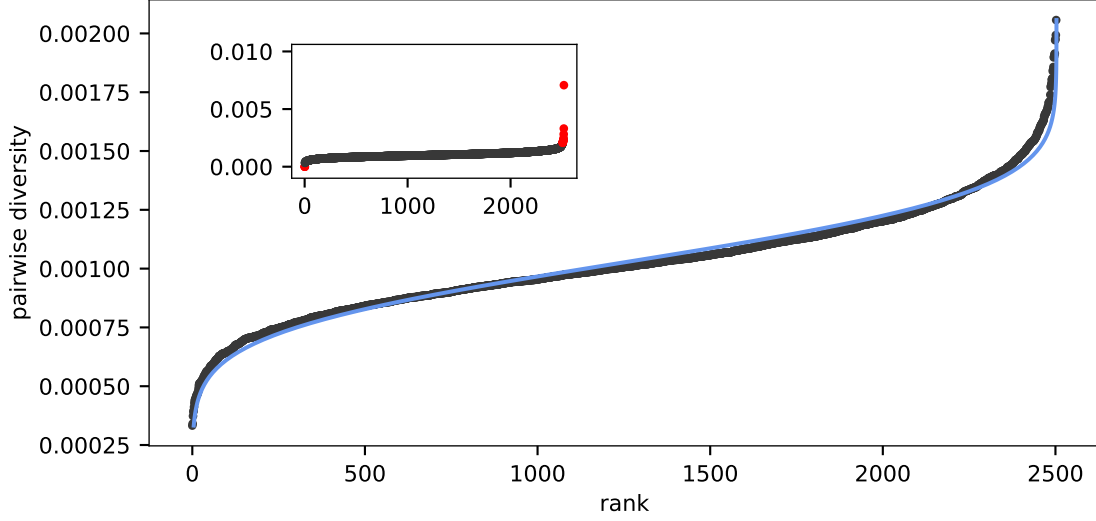
Figure 2: The distribution of diversity across the genome at the megabase scale, with outliers trimmed. The blue line is the best-fitting normal CDF. We note that this implies the distribution of megabase-scale diversity is slightly fat-tailed compared to a normal distribution. The inset figure is the untrimmed data, with trimmed points shown in red.

where each $B_i = f(X_i|\Psi)$ for some data $X_i$ (i.e. the recombinaton map and annotation model) and parameters $\Psi$. To compare our models and assess model goodness-of-fit, we use the out-sample $R^2_{\text{LOCO}}$ which is calculated by fitting the genome-wide model leaving one chromosome out, and calculating the residual variance between predictions and observed values for this out-sample chromosome.

In general, $R^2$ is calculated for a model fit across windows as,

$$R^2 = 1 - \frac{V_{\text{res}}}{V_{\text{tot}}} \tag{65}$$

where $V_{\text{res}}$ is the mean squared error across windows and $V_{\text{tot}}$ is the total variance in the observed values. Suppose we have estimates $\widehat{\pi}_0\widehat{B}_i$ for $\pi_0 B_i$ from our model. Then,

$$V_{\text{res}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{\pi}_0\widehat{B}_i)^2 \tag{66}$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2}_{\text{irreducible error}} + \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}(\pi_0 B_i - \widehat{\pi}_0\widehat{B}_i)\right)^2}_{\text{bias squared}}. \tag{67}$$

The total variance among the observed diversity values is,

$$V_{\text{tot}} = V_{\text{res}} + V_{\text{model}} + 2\widehat{\pi}_0\text{cov}_i(\varepsilon_i, \widehat{B}_i) \tag{68}$$

25

where

$$V_{\text{model}} = \frac{\widehat{\pi}_0^2}{n} \sum_{i=1}^{n} \left( \widehat{B}_i - \frac{1}{n} \sum_{i=1}^{n} \widehat{B}_i \right)^2. \tag{69}$$

Generally, fitting procedures act to minimize the squared deviation between predictors and observed values, which drives $\text{cov}_i(\varepsilon_i, \widehat{B}_i)$ to zero XXX. **Murphy2022-sj** suggest XXX

Thus, our $R^2$ depends on (1) the bias of our predictors, (2) the covariance between residuals and predictors, and (3) the irreducible error due to variance in diversity within windows. However, it is of interest to approximate the irreducible error under theoretic models that assume all irreducible error is due entirely to the variance of coalescence times in a window, assuming that all of the effects of negative selection on the variance can be thought of as a local reduction in the effective population size to $B_i N_e$ and demographic effects amount to a simple rescaling of $N_e$. This also assumes mutation rates are constant across the genome, and do not contribute to the irreducible variance. We call this theoretic goodness-of-fit under drift $R^2_{\text{drift}}$, and it provides a rough approximation of the irreducible error in our model due to "coalescence noise" around the modeled reductions due to negative selection.

Then,

$$\mathbb{E}(\varepsilon_i^2) = \text{V}(\pi_i) \tag{70}$$

since the mean residual is zero. The variance in pairwise diversity $\text{V}(\pi_i)$ is equivalent to the variance in the number of segregating sites for a sample of two, $\text{V}(\pi) = \text{V}(S_2)$. The variance in coalescence times in a window of width $w$ basepairs with population-scaled recombination rate $\rho = 4B_i N_e r w$ is,

$$\text{V}(S_2) = \theta + \theta^2 \frac{2}{\rho^2} \int_0^\rho (\rho - x) f_2(x) dx \tag{71}$$

(**Wakeley2009-ua**) where,

$$f_2(\rho) = \frac{\rho + 18}{\rho^2 + 13\rho + 18}. \tag{72}$$

Since this is a rough approximation, we set $\theta = w\pi_0 B_i$ and fix $\rho = w\gamma\pi_0 B_i$ where $\gamma$ is the ratio of recombination to mutation rates, $r/\mu$. In humans, $\gamma \approx 1$, but we explore different ratios to assess sensitivity of this calculation (Supplementary Figure **??**).

This back-of-the-envelope calculation finds that the upper bound of $R^2$ would be around 67% for Yoruba, and 64% for European and Han Chinese models (Figure 3 This rough approximation assumes constant demography, which for bottlenecked out-of-Africa populations assumes that the variance in coalescence times is determined entirely by reducing the effective population size. However, the variance in coalescence times is *higher* in bottlenecked populations (CEU and CHB) compared to Yoruba, which we confirm with simple simulations using the `OutOfAfrica_3G09` model from `stdpopsim` (**Gutenkunst2009-pg**; **Adrion2020-cf**). This higher than expected variance would inflate $SS_{\text{res}}$, which would decrease $R^2_{\text{drift}}$, bringing it closer to the observed $R^2_{\text{LOCO}}$. A more accurate approximation of the $R^2_{\text{drift}}$ could be found with simulations with more realistic demography and varying coalescence rates along the genome.
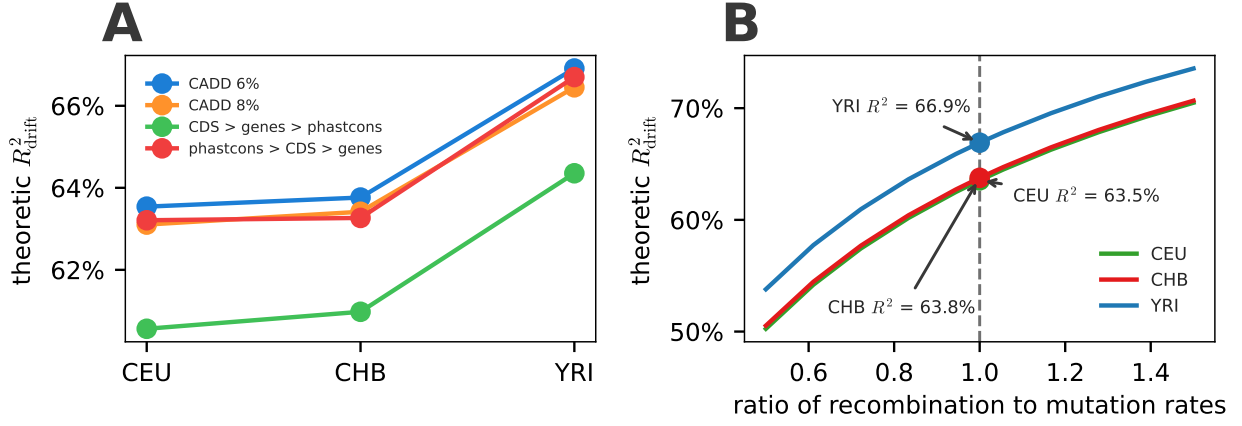
Figure 3: (A) the theoretic $R^2_{\text{drift}}$ using different plugin estimators for $B_i$ under different models. This shows that among the best-fitting models, the predicted $R^2_{\text{drift}}$ varies little; most variation is among populations due to differing $\pi_0$. (B) The $R^2_{\text{drift}}$ across different populations, for varying ratio of recombination to mutation rates (x-axis).
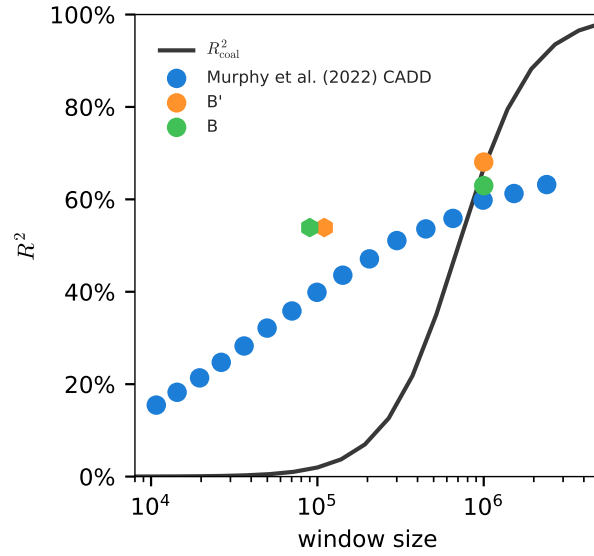


Figure 4: XXX

## 3.9 Predicted Rate of Fitness Change

Each fixation of a mutant with fitness effect $-s$ reduces the population fitness to $1 - 2s$ since all individuals are homozygotes for the deleterious mutation and we assume additive effects at a site. For each segment $l$, the maximum likelihood parameter estimates of our model imply a prediction of the deleterious substitution rate per segment $\widehat{R}_{l,i}$ for each selection coefficient $s_i$ in the $m_s$-element DFE grid (according to which feature class it belongs to).

Since we assume multiplicative fitness effects, the log population fitness (considering only fixed sites) for segment $l$ in the next generation is,

$$\log(L_l) = \sum_{i=1}^{m_s} \log(1 - 2s_i)\widehat{R}_{l,i}. \tag{73}$$

Since we assume multiplicative fitness, the total genome-wide log population fitness in the next generation is

$$\log(L) = \sum_l \log L_l. \tag{74}$$

This amounts to a predicted population fitness loss per generation of

$$\widehat{\Delta}_W = 1 - L \tag{75}$$
$$= 1 - \exp\left(\sum_l \sum_{i=1}^{m_s} \log(1 - 2s_i)\widehat{R}_{l,i}\right).$$

We calculate this value for all segments within a feature class, and then sum over all features to arrive at our genome-wide estimate. We use a parametric bootstrap to carry forward the estimation uncertainty to the predicted fitness loss per generation. Briefly, we do this by sampling new estimates of the DFE per selection coefficient and feature from a normal distribution centered on the ML point estimate and with the jackknife standard deviation.

## 3.10 Likelihood

Our model is essentially a generalized nonlinear model with a binomial link function. This is the form used by **Elyashiv2016-vt** and **Murphy2022-sj**. These models fit the observed number of pairwise nucleotide differences in genomic windows to the expected pairwise diversity under some evolutionary model. We only consider BGS models here, so the mean function for position $x$ is the product of the proportion by which BGS reduces diversity at position $x$, $B(x)$, and genome-wide neutral diversity $\pi_0$,

$$\pi(x) = B(x, \Phi)\pi_0 \tag{76}$$

$\Phi$ is the set of BGS parameters (i.e. the DFE matrix $\mathbf{W}$ and mutation rate $\mu$), and $\pi_0 = 4N_e\mu$ is determined by the genome-wide drift-effective population size $N_e$, set by only reproductive and

demographic processes. Regional mutation rate heterogeneity can be accounted for with a regional mutation rate scaling function $m(x)$, $\pi(x) = B(x, \Phi)m(x)\pi_0$, but we do not explore mutation-rate heterogeneity, as it is unclear how to differentiate variance in mutation rate from the substitution rate heterogeneity we find across the genome.

The likelihood of the set of parameters $\Psi = \{\pi_0, \Phi\}$ can be written in the form of,

$$\log \mathcal{L}(\Psi) = \sum_{v \in \mathcal{V}} \sum_{i \neq j \in \mathcal{S}} \log(P(O_{i,j}(v)|\Psi)) \tag{77}$$

(c.f. **McVicker2009-ax**; **Elyashiv2016-vt**; **Murphy2022-sj**) where $\mathcal{V}$ is the set of putatively neutral sites, $\mathcal{S}$ is the set of samples, and $\Psi$ are the BGS parameters. The indicator variable $O_{i,j}(v)$ is 1 if samples $i$ and $j$ are different at putatively neutral site $v$, and zero otherwise. While the theoretic $\pi(v)$ gives the average number of pairwise differences, for small values, this is approximately the heterozygosity probability, so we can write

$$P(O_{i,j}(v)|\Psi) = \begin{cases} \pi(v), & O_{i,j}(v, \Psi) = 1 \\ 1 - \pi(v), & O_{i,j}(v, \Psi) = 0 \end{cases}. \tag{78}$$

(c.f. **Elyashiv2016-vt**).

As in Section 2.6, the number of pairwise differences and the total number of pairwise comparison at a site are sufficient statistics for the likelihood. Then, the binomial log-likelihood for the data at site $v$ is,

$$\ell_v(\Psi) = \log(\pi(v, \Psi))n_{\mathrm{D},v} + \log(1 - \pi(v, \Psi))n_{\mathrm{S},v} \tag{79}$$

## 3.11 The scale of processes

We can observe $\widehat{\pi}(x)$ at a per-basepair resolution. However, for a variety of reasons, we do not want to fit the composite likelihood model to the per-basepair scale of data. First, this would be computationally infeasible. Second, the mean function $\pi(x)$ varies on a natural scale that is itself a free parameter of the model. Our model can be written as,

$$\ell(\Psi, h) = \sum_b \sum_{v \in \mathcal{V}_b} \ell_v(\Psi) \tag{80}$$

$$= \sum_b \left[ \log(\bar{\pi}(b, \Psi)) \sum_{v \in \mathcal{V}_b} n_{\mathrm{D},v} + \log(1 - \bar{\pi}(b, \Psi)) \sum_{v \in \mathcal{V}_b} n_{\mathrm{S},v} \right] \tag{81}$$

$$= \sum_b \left[ \log(\bar{\pi}(b, \Psi))Y_{\mathrm{D},b} + \log(1 - \bar{\pi}(b, \Psi))Y_{\mathrm{S},b} \right] \tag{82}$$

$$\tag{83}$$

where $h$ is the bandwidth or window size, $b$ the bin index for windows of width $h$, $\mathcal{V}_b$ is the set of putatively neutral sites in bin $b$, $\bar{\pi}(b|\Psi)$ are the average diversity in bin $b$, and $Y_{\mathrm{S},b}$ and $Y_{\mathrm{S},b}$ are the sums across putatively neutral sites within a bin. Note that by binning, we sum the pairwise

summaries of the data **Y** across sites, so the likelihood across bins is naturally weighted by the quantity of observed data.

This model corresponds to a binomial likelihood for the observed data summarized at genomic scale $h$. Thus an alternative way to express this model is as,

$$Y_{\mathrm{D},b} \sim \mathrm{Binom}(\bar{\pi}(b, \boldsymbol{\Psi}), Y_{\mathrm{D},b} + Y_{\mathrm{S},b}). \tag{84}$$

Here, $\bar{\pi}(b, \boldsymbol{\Psi})$ is assumed to be the *probability* of sampling two different alleles, rather than the average *number* of pairwise differences; these are approximately equal when $\pi$ is small. This corresponds to an identity link function; one could alternatively use a two-alleles finite sites model link function of the form, $\pi/(1 + 2\pi)$. We experimented with this and found there was little difference between these link functions, so we opted for the simpler identity link function.

## 3.12  Out-Sample Error Estimation and $\widehat{R}^2_{\mathrm{LOCO}}$

To estimate out-sample prediction error, we fit each model leaving one chromosome out (LOCO) and calculating the out-sample prediction error on these excluded chromosomes. Each fit used 10,000 random starting positions as the main fits. The out-sample prediction error (i.e. sum of residuals squared, $\widehat{\mathrm{SSR}}_{\mathrm{LOCO}}$) for all bins in chromosome $C$ left out is,

$$\widehat{\mathrm{SSR}}_{\mathrm{LOCO}}(C) = \sum_{b \in C} (\widehat{\pi}_{b,(C)} - \pi_b)^2 \tag{85}$$

where $\widehat{\pi}_{b,(C)}$ denotes the predictions for window $b$ fit on data excluding chromosome $C$. The average out-sample error $\widehat{\mathrm{SSR}}_{\mathrm{LOCO}}$ across chromosomes is then the average of $\widehat{\mathrm{SSR}}_{\mathrm{LOCO}}(C)$ across chromosomes, weighted by the number of bins per chromosome. We then calculate the out-sample coefficient of determination $\widehat{R}^2_{\mathrm{LOCO}}$

$$\widehat{R}^2_{\mathrm{LOCO}} = 1 - \frac{\widehat{\mathrm{SSR}}_{\mathrm{LOCO}}}{\widehat{\mathrm{SST}}_{\mathrm{LOCO}}} \tag{86}$$

where,

$$\widehat{\mathrm{SST}}_{\mathrm{LOCO}}(C) = \sum_{b \in C} (\widehat{\pi}_{b,(C)} - \bar{\pi})^2 \tag{87}$$

and $\bar{\pi}$ is the average diversity across all windows.

## 3.13  Jackknife Procedures

We use a block jackknife procedure to calculate the uncertainty of our estimates. This approach is unfortunately computationally intensive, since a 10Mbp block width requires $\approx 280$ blocks to cover 2.8Gbp of autosomal sequence and each fit requires thousands of random starts to reliably find consistent optima. We use a parallelization strategy where a set number of evenly-spaced positions along the genome that are chosen to approximately cover all 2.8Gbps of autosomal sequence. Each
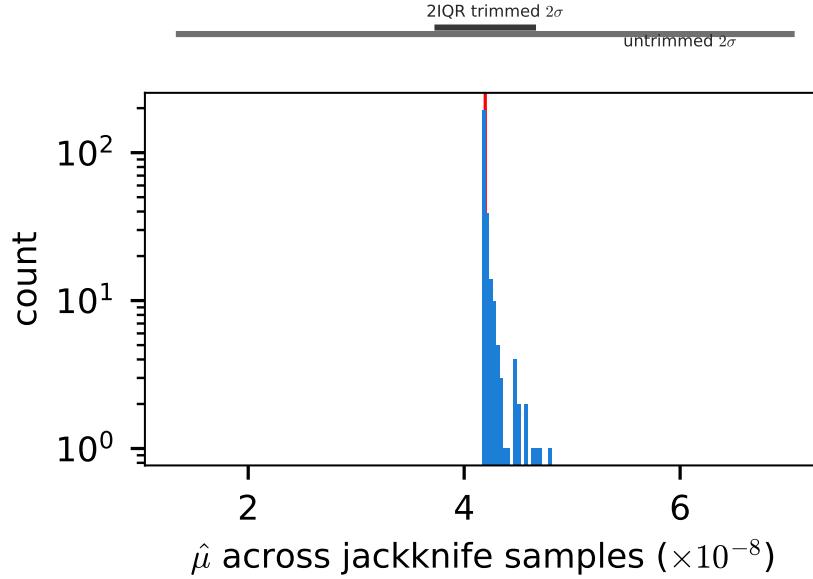
Figure 5: The distribution of block jackknife samples for the Feature Priority model fit on YRI data. The ranges at the top of the figure indicate the untrimmed and the 2 IQR-trimmed jackknife intervals. Trimming off the numeric optimization outliers leads to more realistic estimates of uncertainty.

of these positions is sent out across a computing cluster, which excludes the 10Mbp block there and fits on the remaining data. Then, using these jackknife samples we calculated the jackknife standard error for each of parameters (**Wasserman2004-ip**). In some cases, our jackknife standard error estimates were very large. Upon inspection of optimization diagnostic plots (see Supplementary Materials Section **??**) and histograms of jackknife fits (Supplementary Figure **??**), these were due to rare aberrant optima. We removed these using a trimmed-interquartile range, which removes points that are 2 interquartile ranges below the first quartile or above the third quartile (1.5 is standard, but we chose to be more conservative to not underestimate uncertainty).

## 4   Simulations

All of our forward simulations used SLiM version 4.0.1 (**Haller2019-vu**; **Haller2023-uk**) and all Eidos scripts are available in the `slim_sims/` directory in the GitHub repository (`https://github.com/vsbuffalo/bprime`). These Eidos scripts take command line arguments for the parameters. Our simulation pipelines use YAML configuration files that specify the parameter grids and other settings, which are read and dispatched across a cluster using Snakemake (**Koster2012-iv**).

### 4.1   Segment Simulations

To confirm the SC16 theory, we simulated a 100 kbp segment in a population of $N = 1000$ diploids across grids of mutation and selection parameters. The Eidos script for these simulations can be found at `slim_sims/region/region.slim` and the analysis notebook at `notebooks/region_`

`simulations.ipynb` in the GitHub repository (`https://github.com/vsbuffalo/bprime`). We simulated over grids of mutation rates ($\mu \in \{10^{-9.5}, 10^{-9}, 10^{-8.5}, 10^{-8}\}$) and selection coefficients ($s \in \{10^{-1}, 10^{-1.5}, 10^{-2}, \ldots, 10^{-4.5}\}$). Selection was additive within a site and multiplicative across sites, and recombination was fixed at $10^{-8}$ per basepair.

We replicated $10{,}000$ simulations for each parameter combination to average and compare against theory. Each simulation recorded the number of fixations and the fitness variance per generation. At the end $10N$ generations, the ancestral recombination graph was recorded. Then, downstream Python scripts calculate branch-length diversity using `tskit` (**Kelleher2018-qu**), and we estimate the reduction factor in a window centered in the region with $B = \pi/4N$ (since branch length diversity implies $\mu = 1$).

## 4.2   Whole-chromosome Simulations

Our whole-chromosome simulations use realistic human recombination maps and putatively conserved segments. These annotation differ slightly from our primary model fits, since these simulations were written before the method was finalized. We use the HapMap recombination map (**International·HapMap·Consortium2007-nu**), PhastCons regions, and UTRs as an approximate set of putatively conserved segments.

Our first set of simulations were used to validate that our methods were producing accurate classic B and B' maps. We replicated 100 whole-chromosome simulations over grids of mutation rates ($\mu \in \{10^{-10}, 10^{-9.5}, \ldots, 10^{-8}, 10^{-7.5}\}$) and selection coefficients ($s \in \{10^{-1}, 10^{-1.5}, 10^{-2}, \ldots, 10^{-6}\}$). For each parameter combination, we calculate mean branch diversity in 10 kbp windows across the 100 replicates to create the average simulation reduction map.

We then compare this average simulation map to our method's theoretic B' maps (see Figure 1 D and E). We calculate the MSE per 10 kbp window to estimate the total rate across the chromosome. We can assess the accuracy relative to a theoretic lower bound of the MSE if the noise were due to coalescence noise. We approximate this by noting that the MSE can be decomposed into a squared bias term and a variance term. If our methods are unbiased in estimating the true reduction at a position, the MSE is just the stochastic variance due to the $n = 100$ replicates we simulate and the coalescence noise in each one. **Tajima1983-gu** showed that variance in diversity due to coalescence noise is much larger than sampling noise, and is given by $V(\pi) \approx 2\theta^2/9$ in an infinite sample. One can derive an analogous equation for the variance of average simulation $\bar{B}$ across $n$ replicate simulations as, $V(\bar{B}) \approx 2B^2/9n$. We note that this assumes no recombination but this approximation is fine over short scales (such as the 10 kbp scale used here).

# 5   Additional Results and Figures

## 5.1   Optimization and Diagnostics

We only show one optimization diagnostic plot here (Supplementary Figure **??**), but the can be easily calculated from the model pickle files in the XXX repository (XXX). In one case, our optimization diagnostics indicated that the best optima across 10,000 random starts was one of few of the top 100 hits that were hitting either an alternate likelihood mode, or the optimization routine had numeric issues. In this one case, we removed the top hit and used the second, which was the consensus among the other top 100 optima.
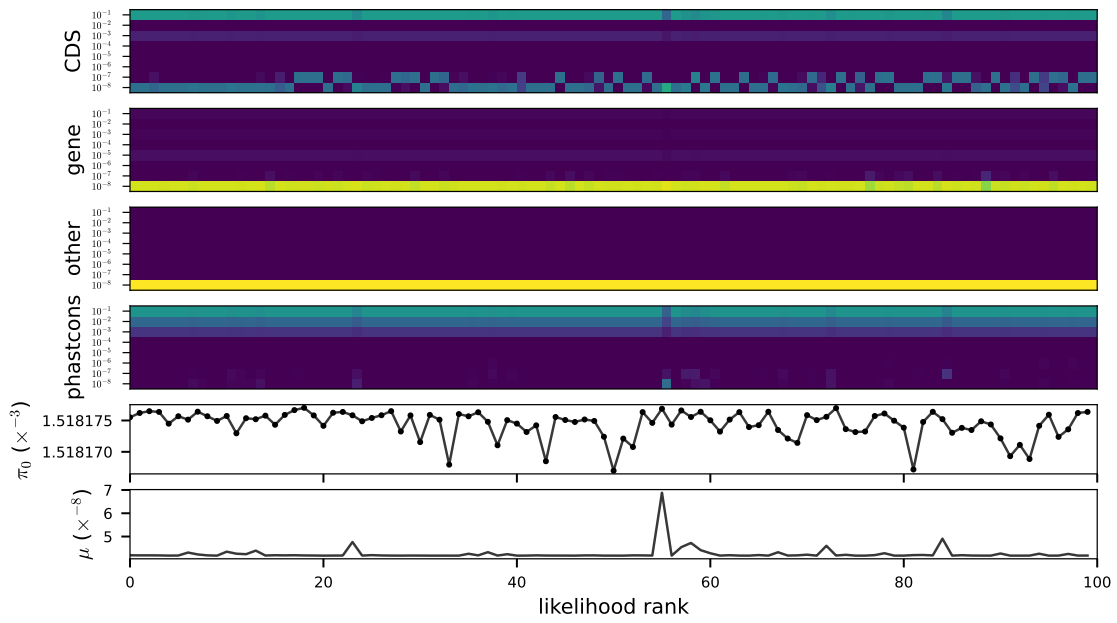
Figure 6: An optimization diagnostic plot showing the 100 top optima. Heatmap colors indicate probability mass for the DFE parameters (top four rows). Estimates of $\pi_0$ and $\mu$ are shown in the bottom two rows. This is the optimization plot for YRI Feature Priority model.

## 5.2 Demographic Expansion and Recessive Simulation Tests
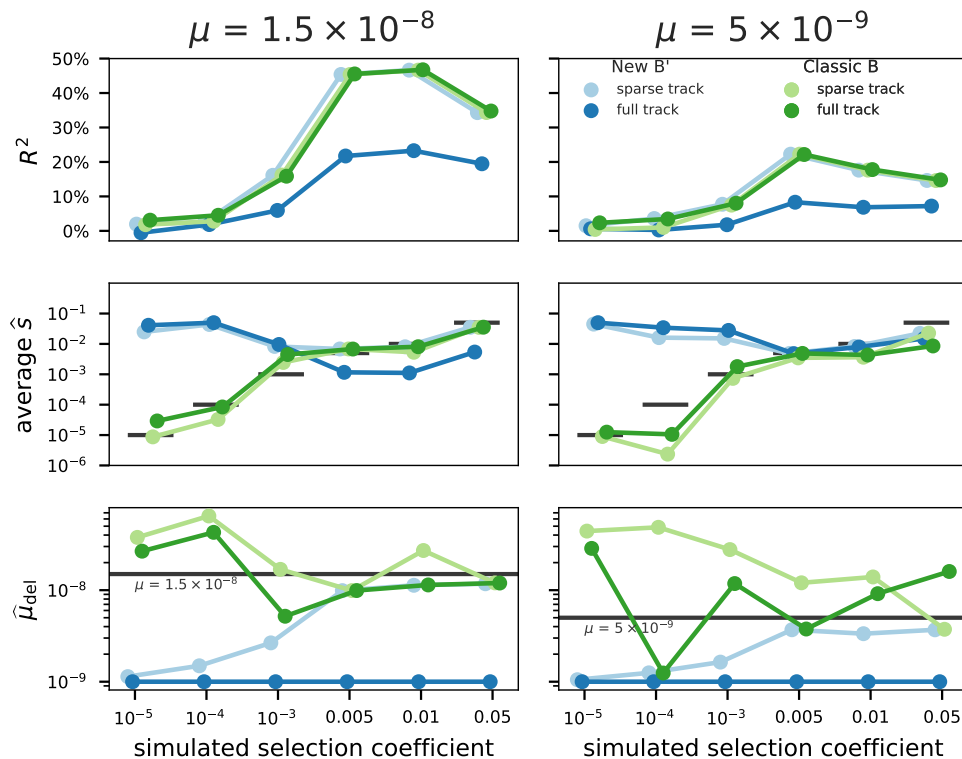


Figure 7: Simulations to diagnose the effect of population exponential expansion on our composite likelihood estimates. The simulated population expansion occurs at a rate of 0.4% a generation, starting $9.3N$ (out of $10N$) generations in. These parameters were meant to approximate the European expansion parameters estimated by **Gutenkunst2009-pg**. Overall, we see little impact on estimated parameters.

## 5.3 Residual Covariate Diagnostics

## 5.4 Additional DFE Figures

## 5.5 Spatial Residuals

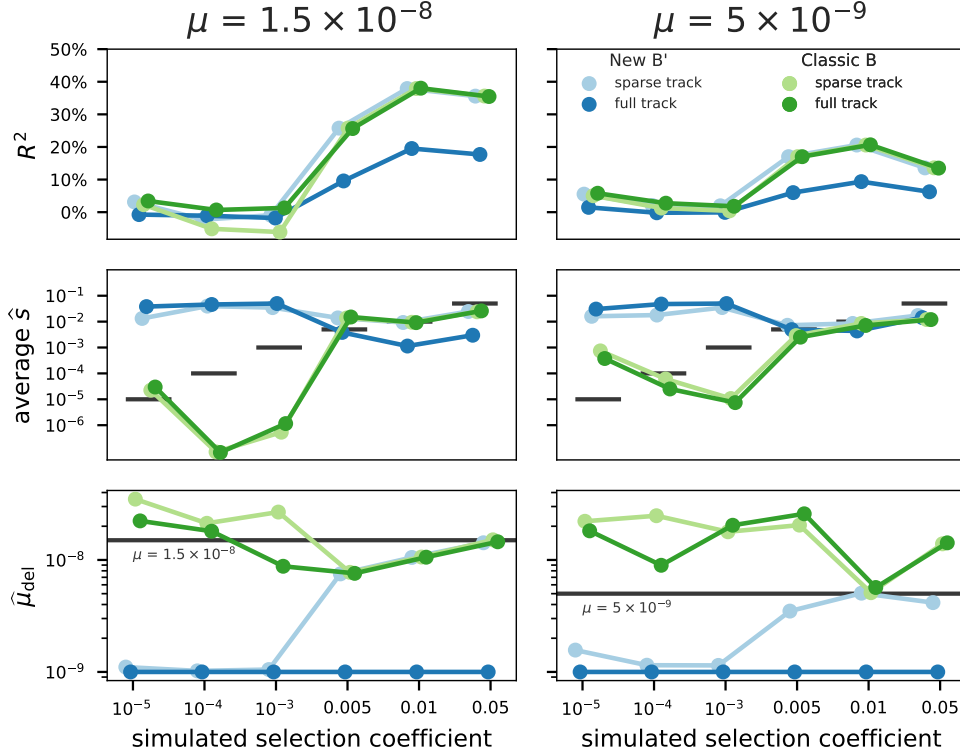Figure 8: Simulations to diagnose the effect of non-additive mutations. These simulations use a dominance coefficient of $h = 0.3$, i.e. the selective effects are $1$ ancestral state, $1 + hs$ for heterozygotes, and $1 + s$ for homozygotes. We find that slightly recessive mutations have little effect on the estimated selection coefficient when selection is strong, consistent with the fact that only the selective effect in heterozygotes matters when mutations are kept at low frequency by strong selection. However, our B' method underestimates weak selection coefficients when mutation effects are recessive. This is because recessive mutations are removed far more efficiently once they reach appreciable frequencies and are found in heterozygotes. Since our theory assumes additive effects, it accommodates by estimating weaker selection coefficients.
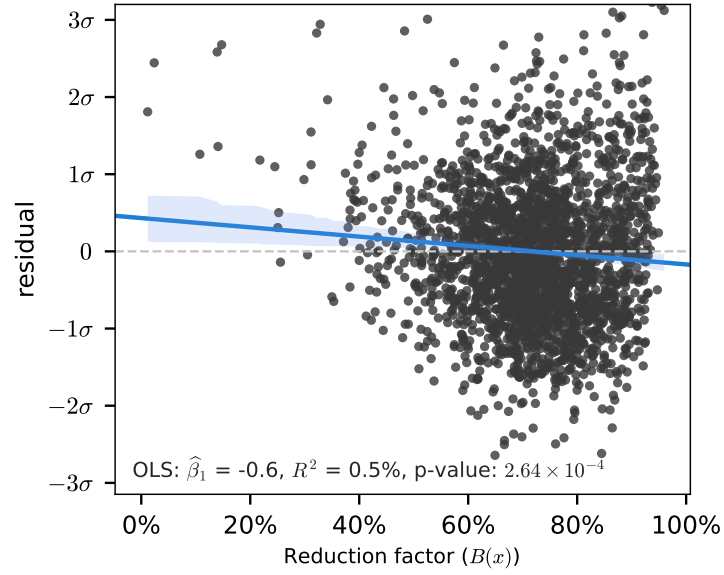
Figure 9: Residuals from the CADD 6% model plotted against the predicted $B(x)$ value.
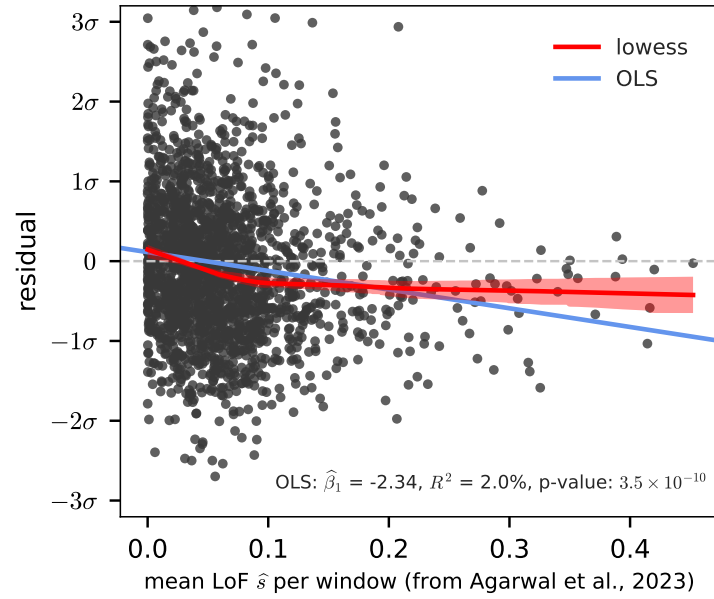


Figure 10: Residuals from the PhastCons Priority model plotted against the average LoF selection coefficient across genes in a megabase window (estimated from **Agarwal2023-un**). These results show that the remaining likely selection signal is not dependent on the model used (PhastCons Priority versus CADD 6%).
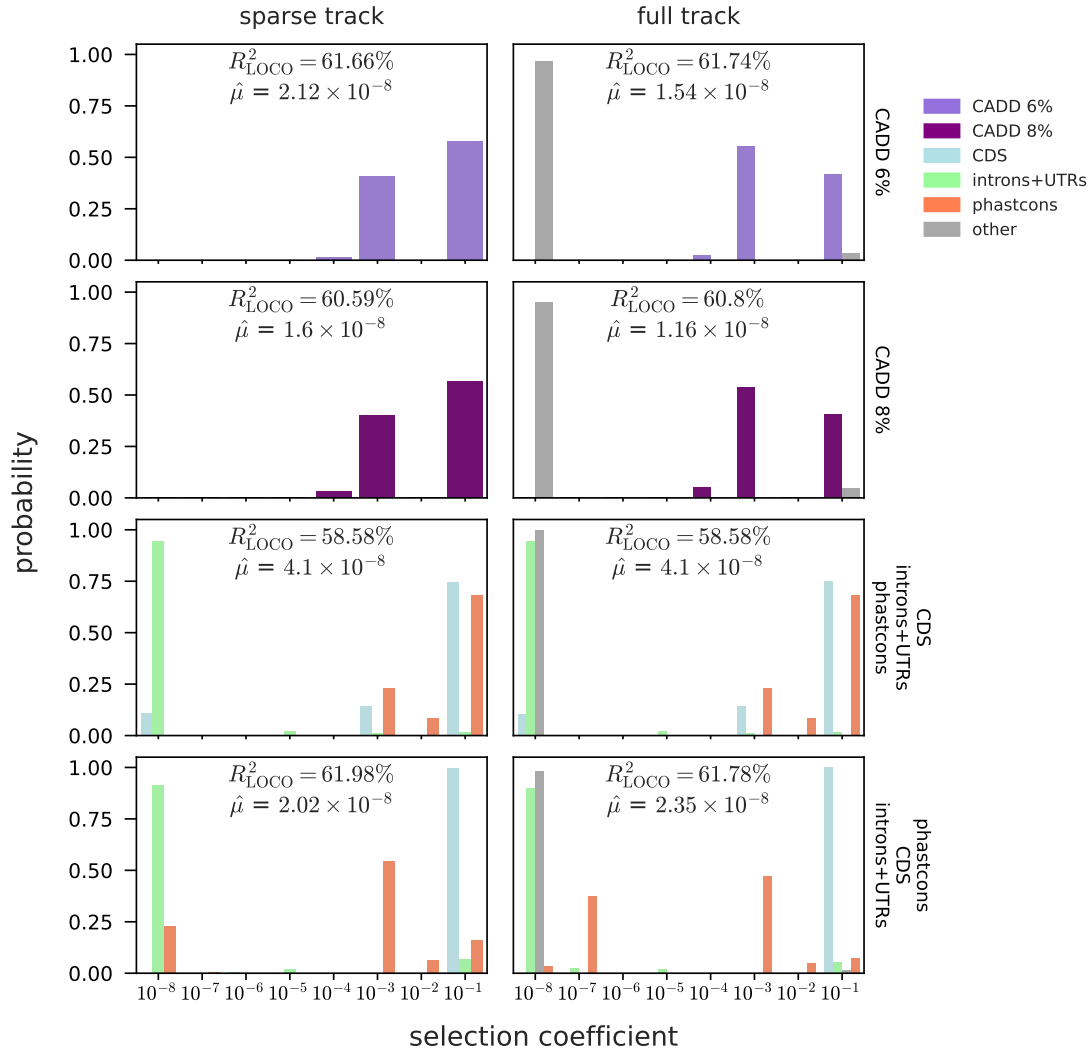
Figure 11: The DFE across models for the CEU reference sample (analogous to the DFE for YRI samples in Figure 3.
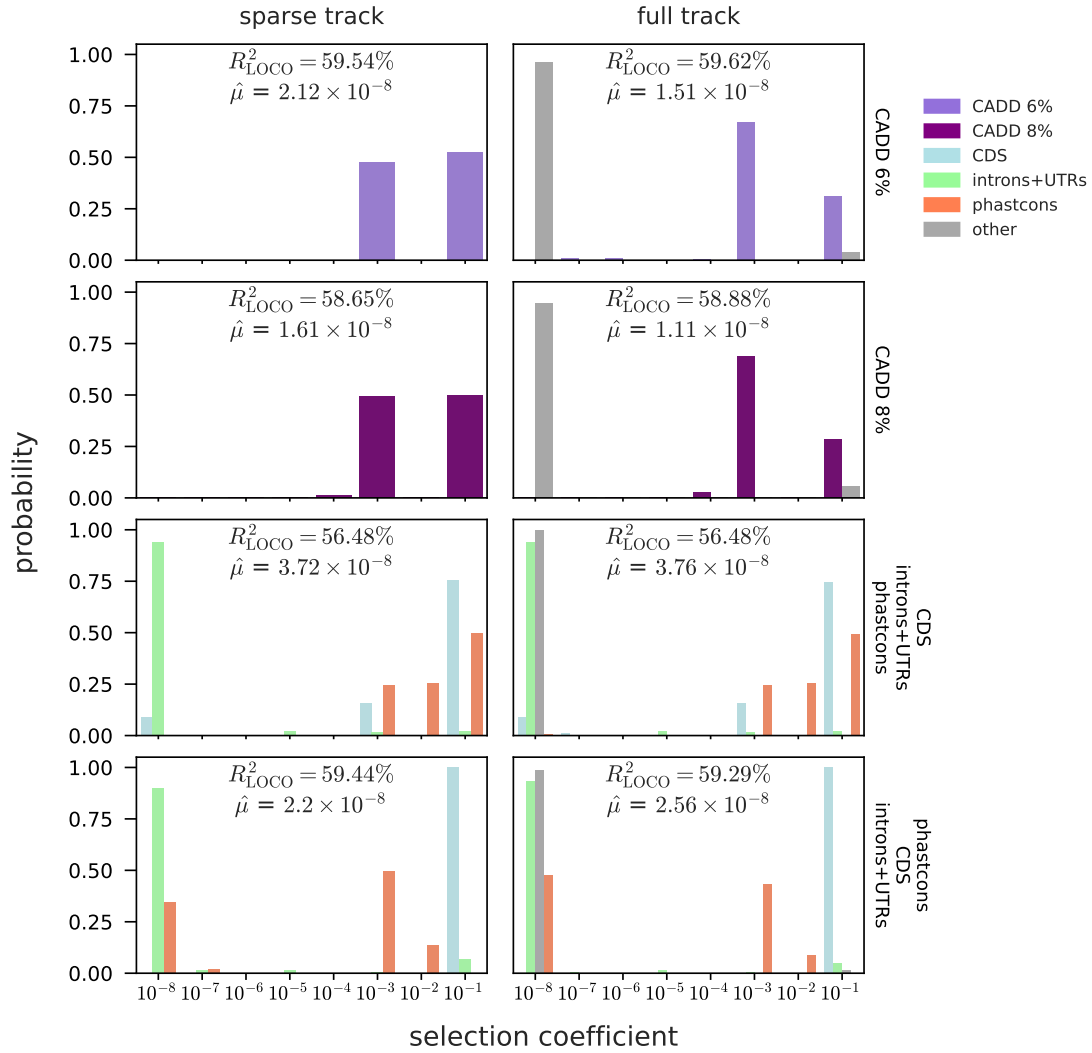
Figure 12: The DFE across models for the CHB reference sample (analogous to the DFE for YRI samples in Figure 3.
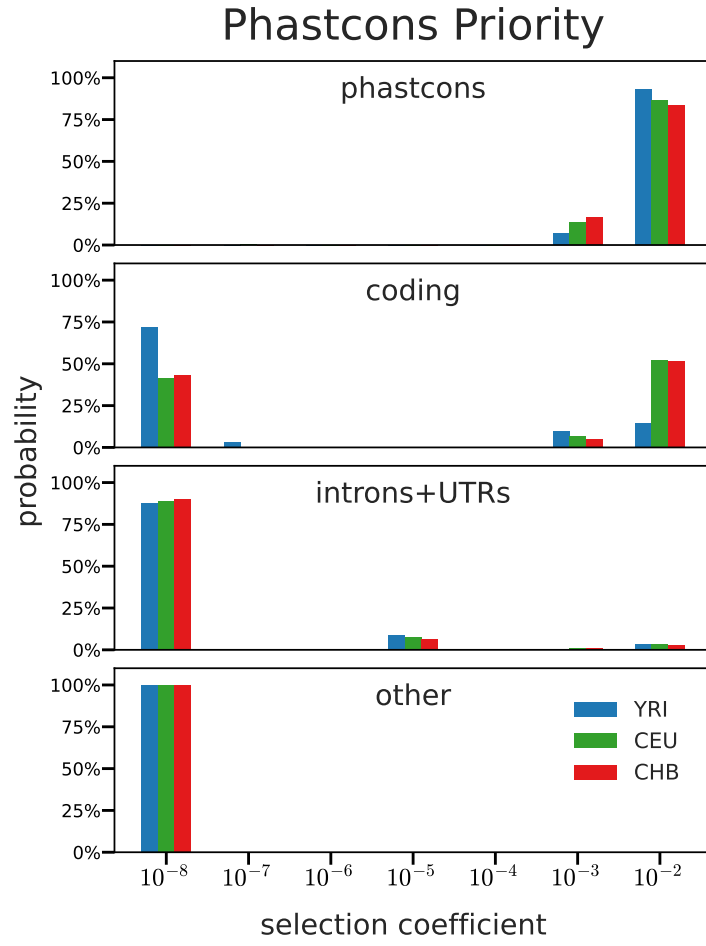
Figure 13: The DFE for the Feature Priority model, fit with a smaller DFE grid from $\mu \in \{10^{-8}, 10^{-7}, \ldots, 10^{-2}\}$.
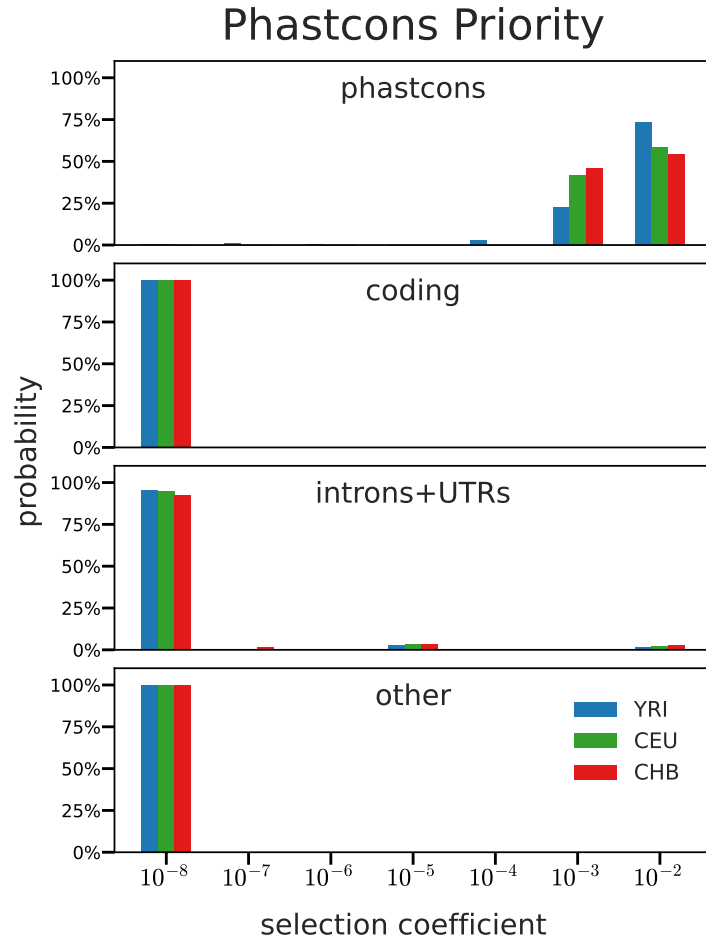
Figure 14: The DFE for the PhastCons Priority model, fit with a smaller DFE grid from $\mu \in \{10^{-8}, 10^{-7}, \ldots, 10^{-2}\}$.
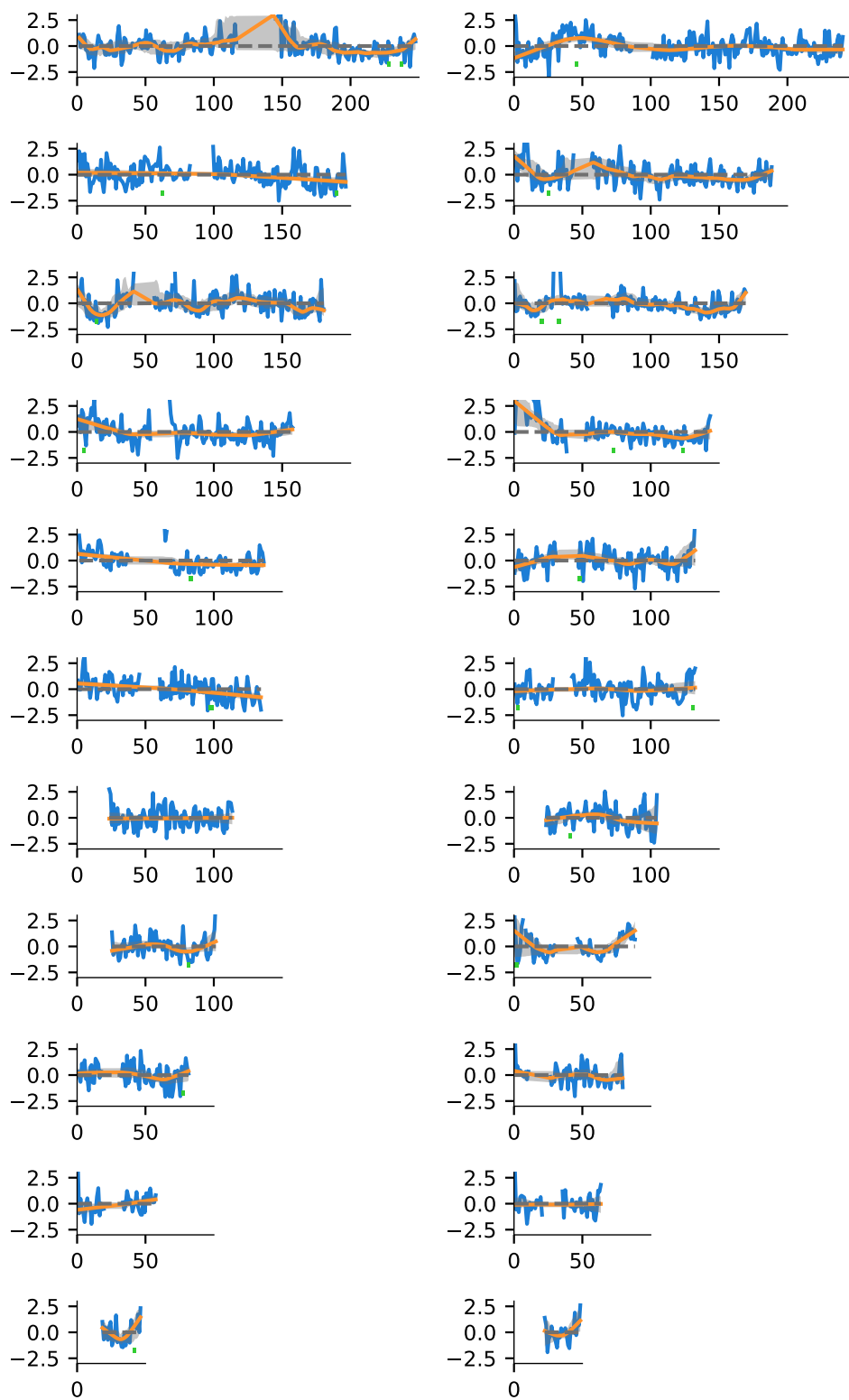
## 5.6 Chromosome Fit Figures
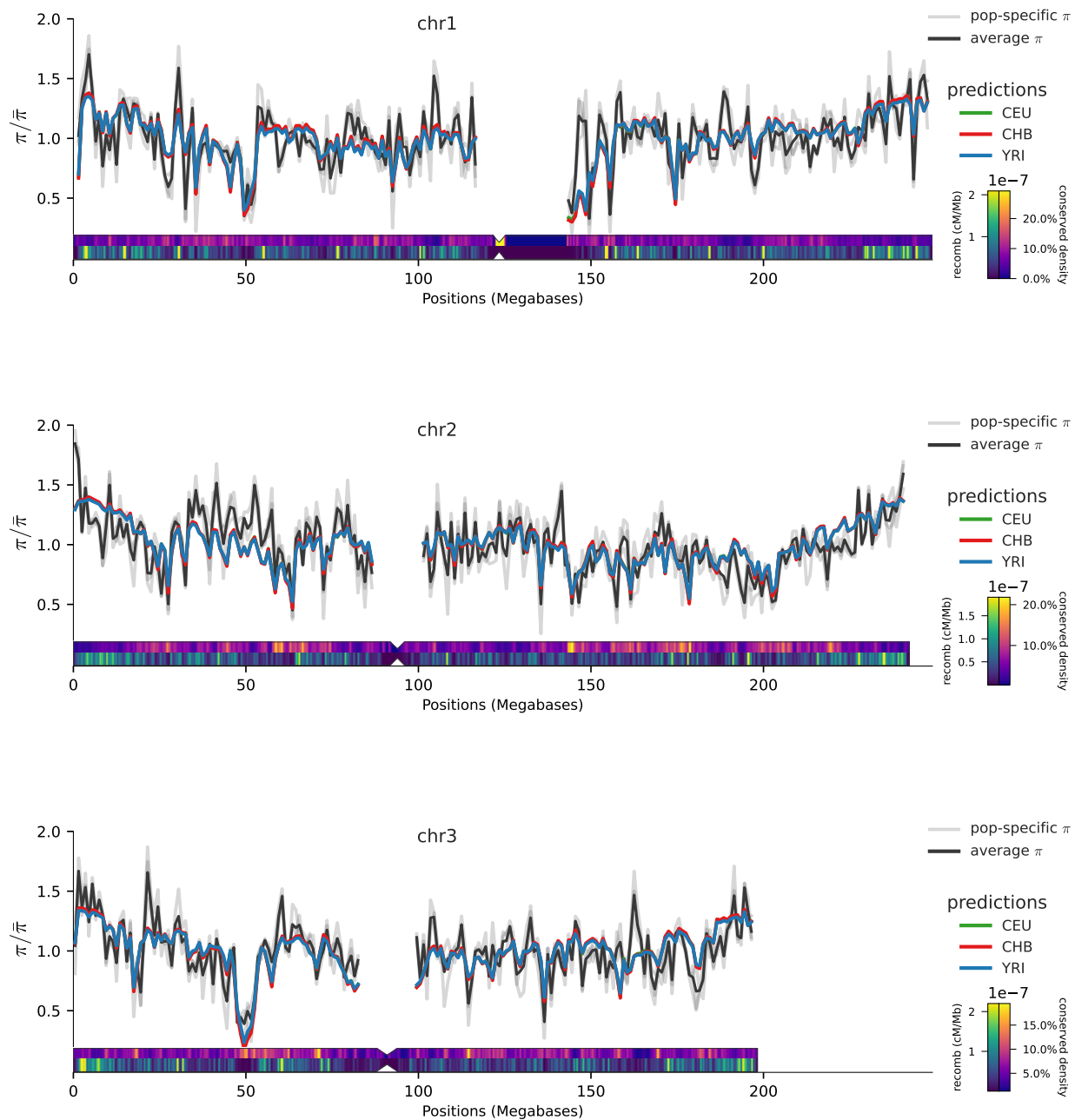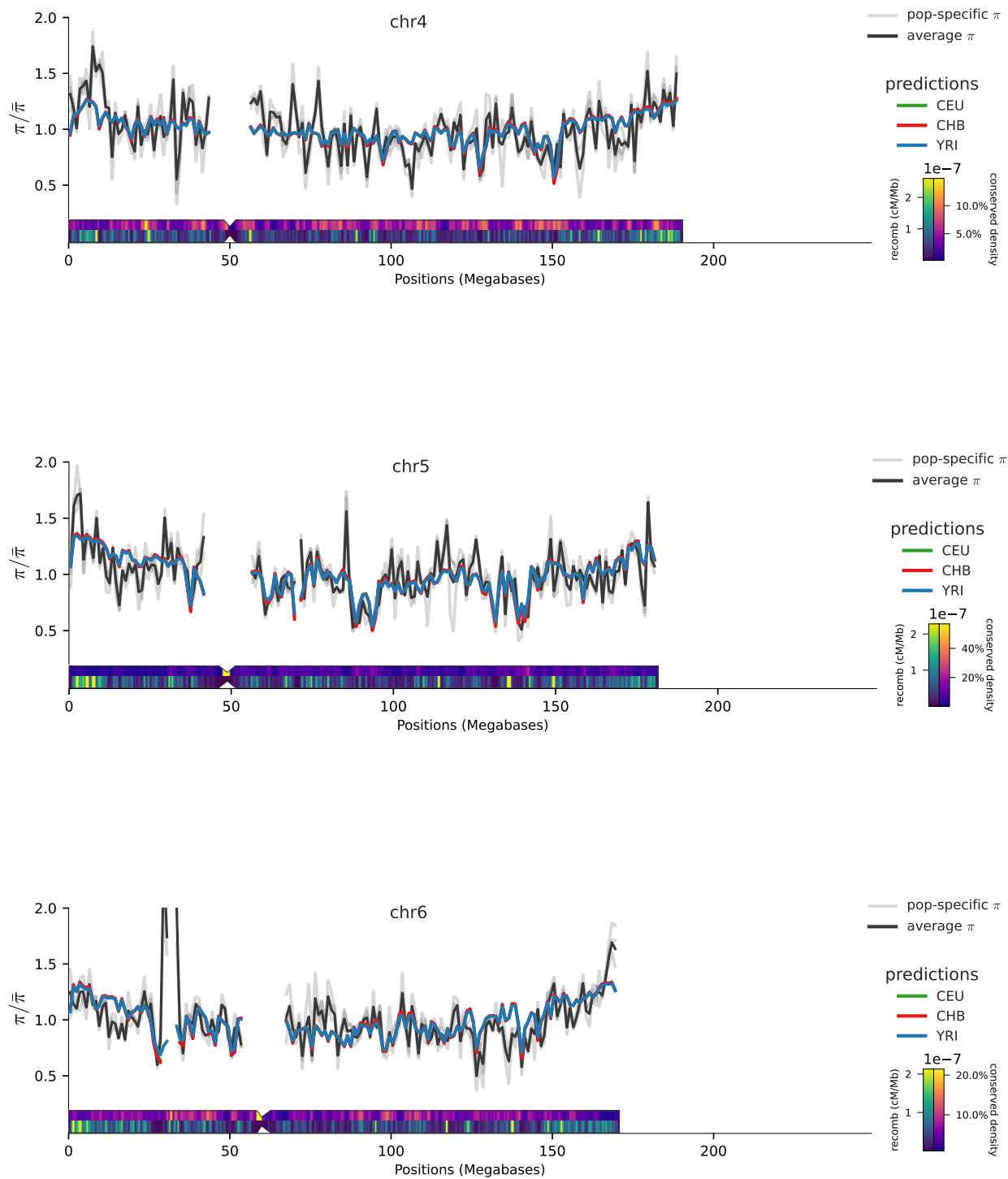
Figure 15: Note: we mask the MHC region during model fit.

chr19

chr20

chr21

chr22
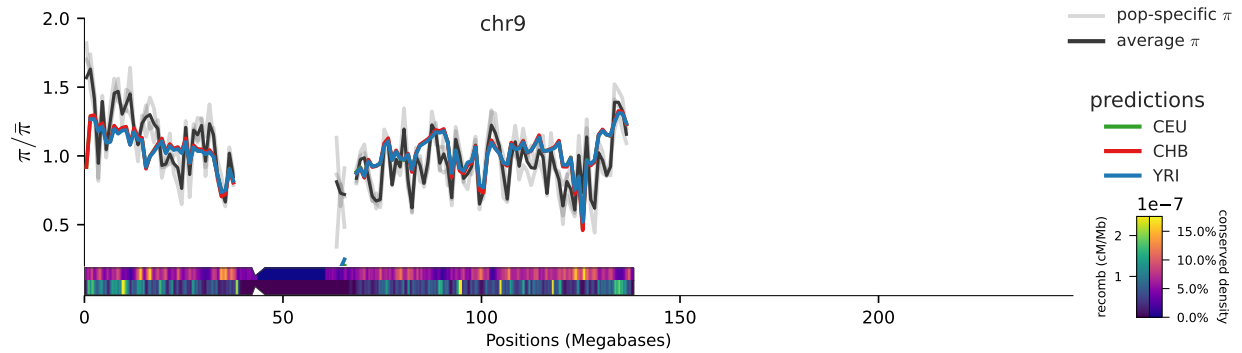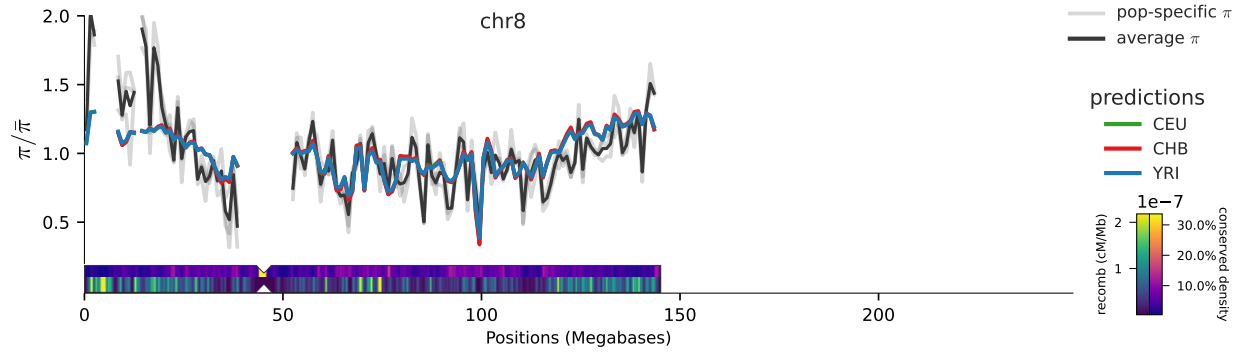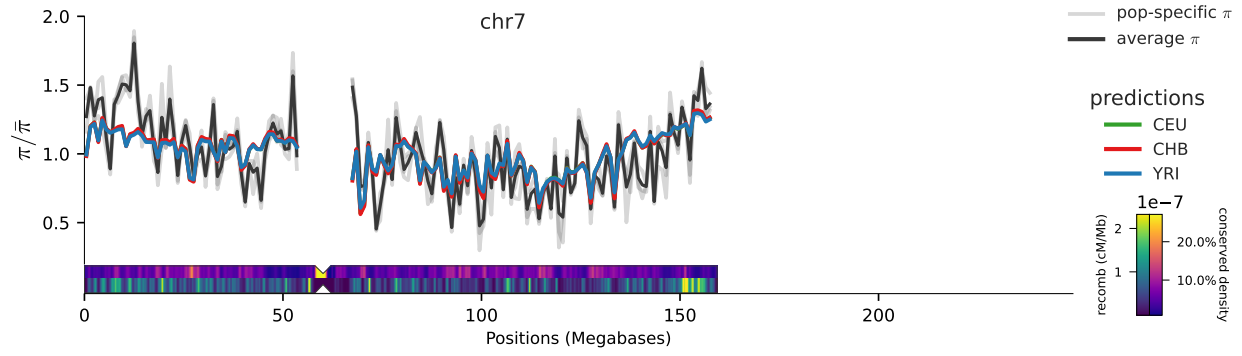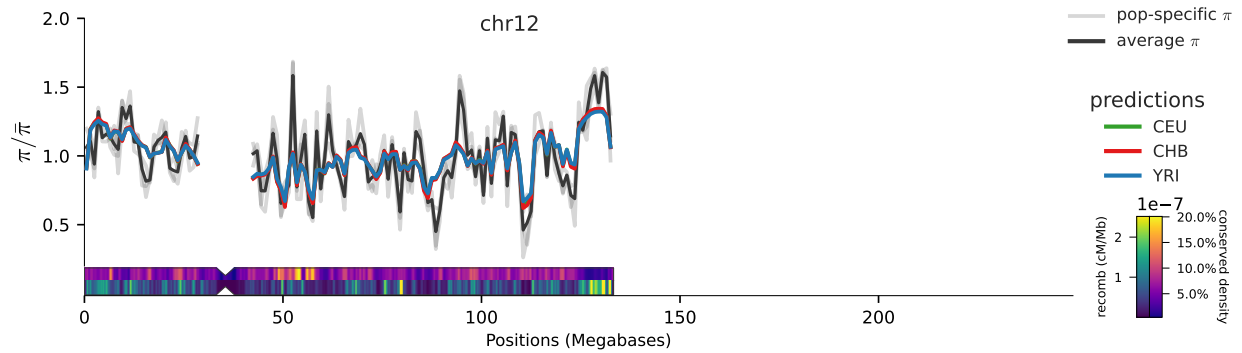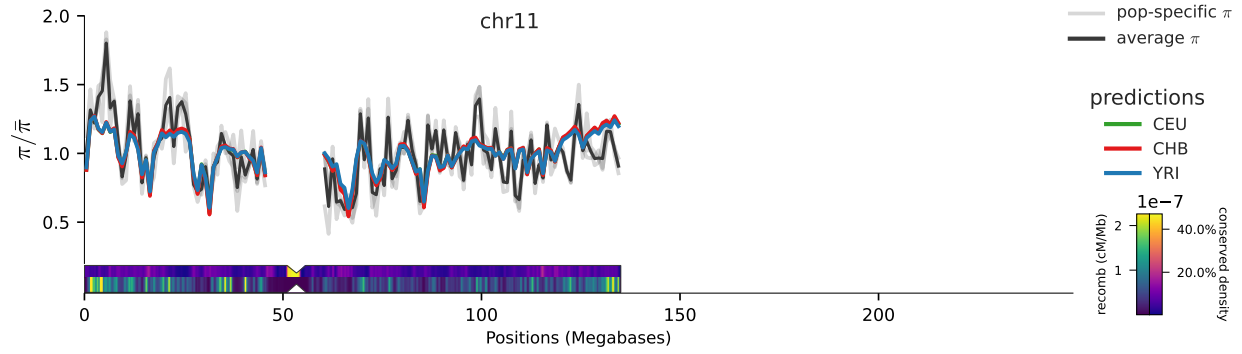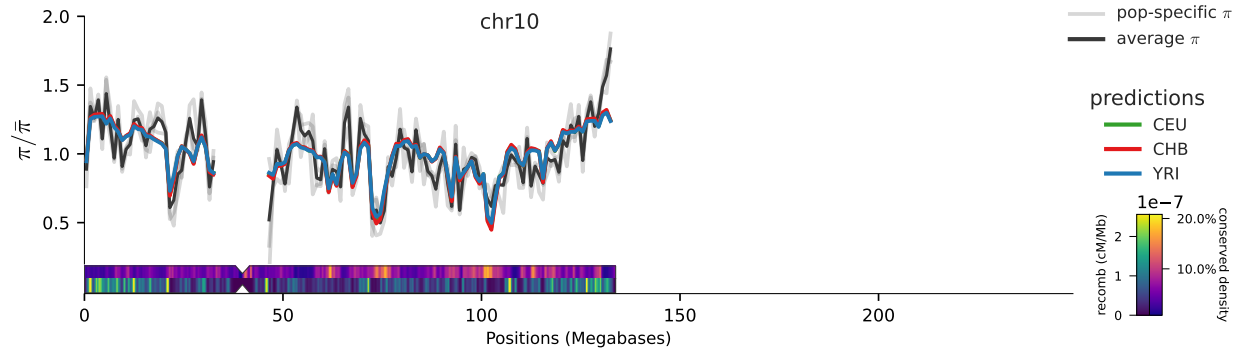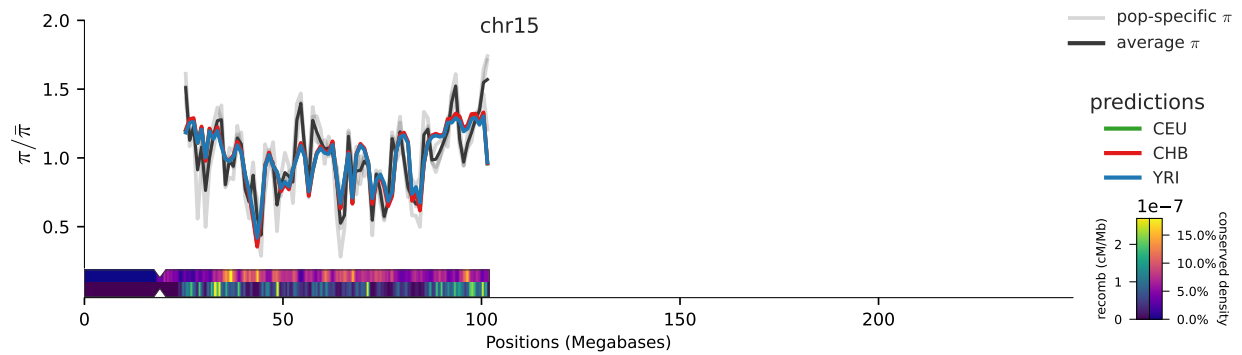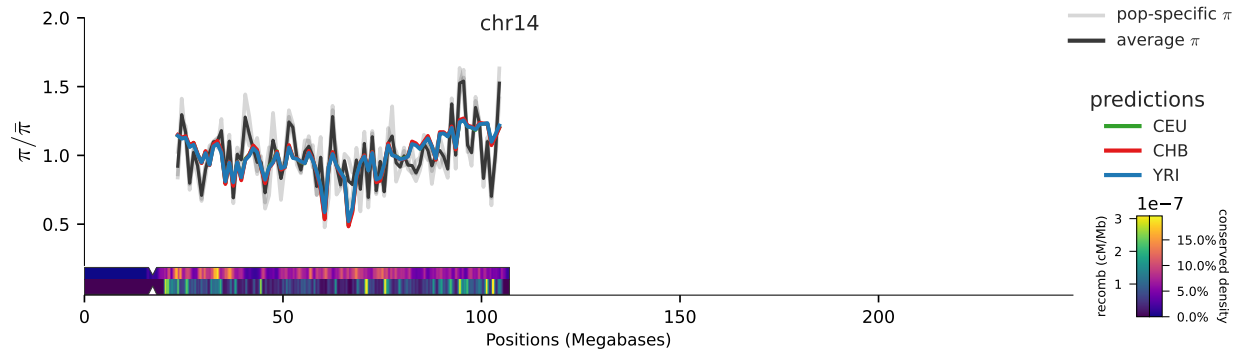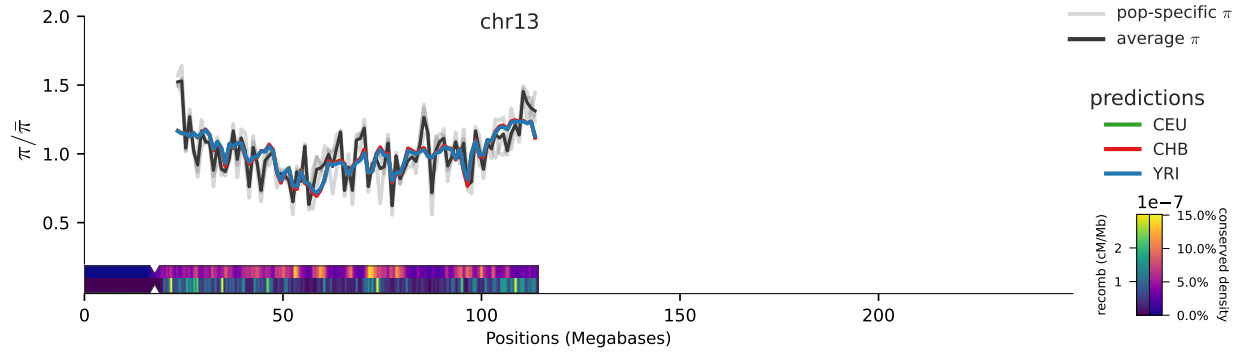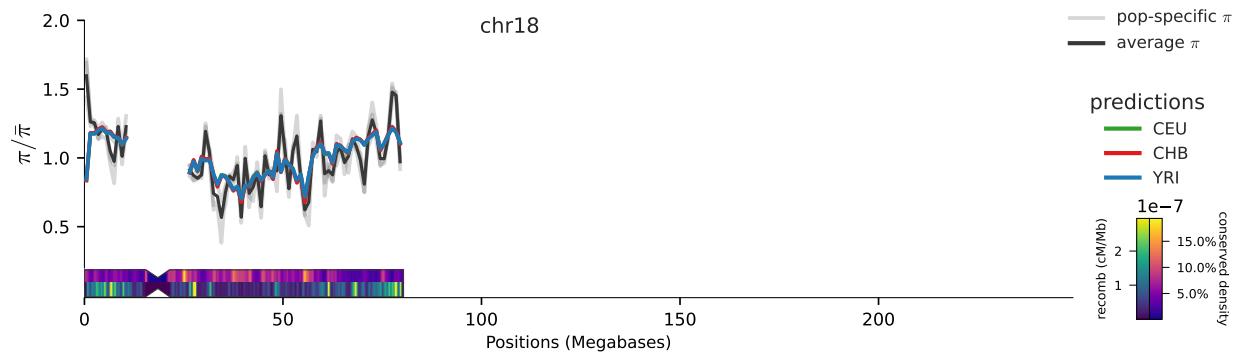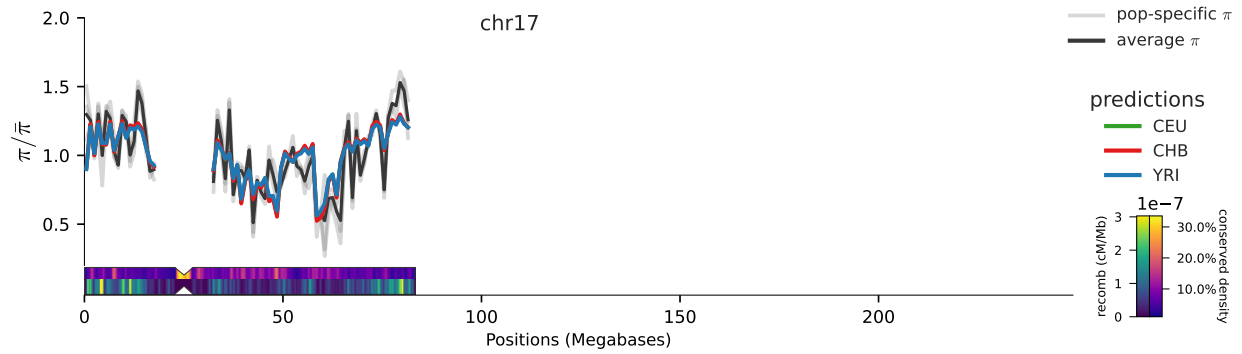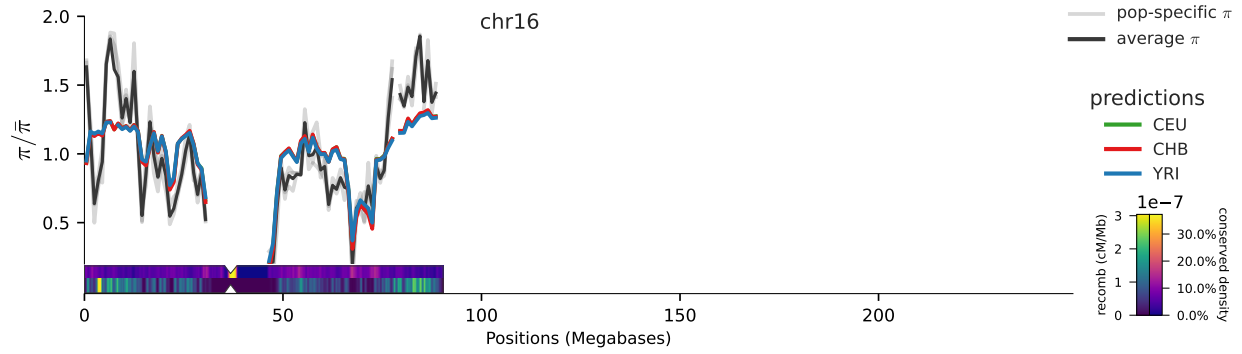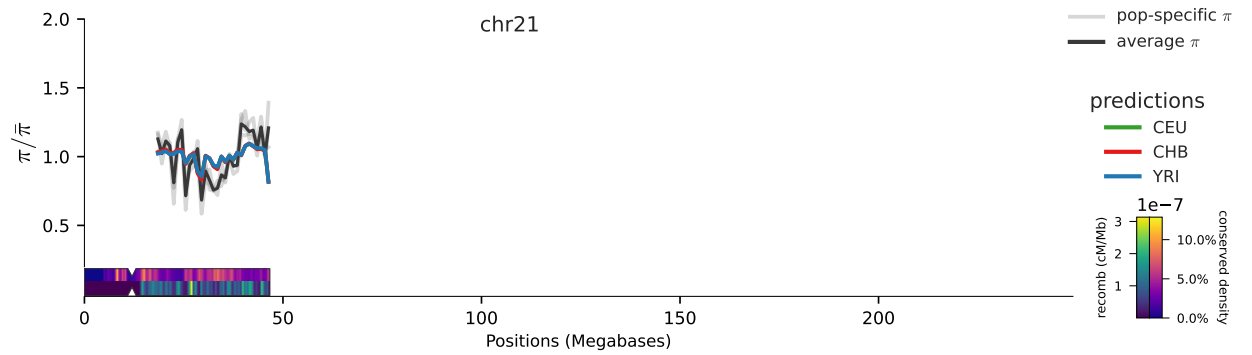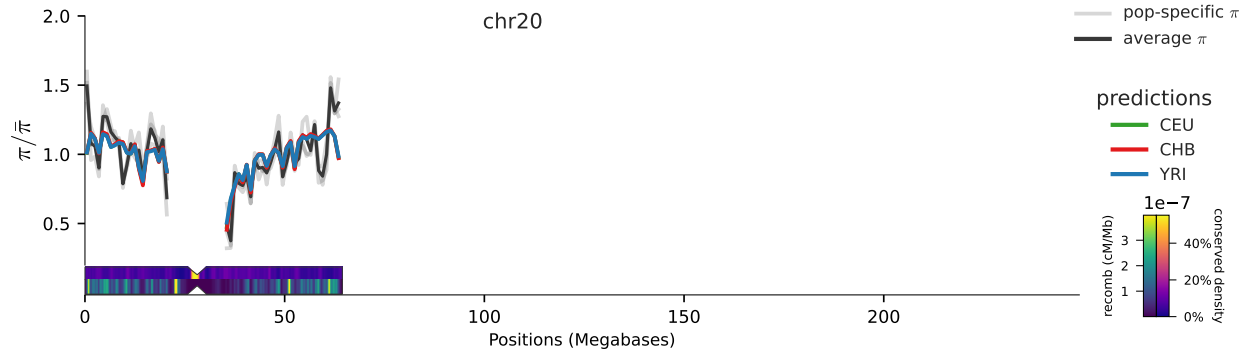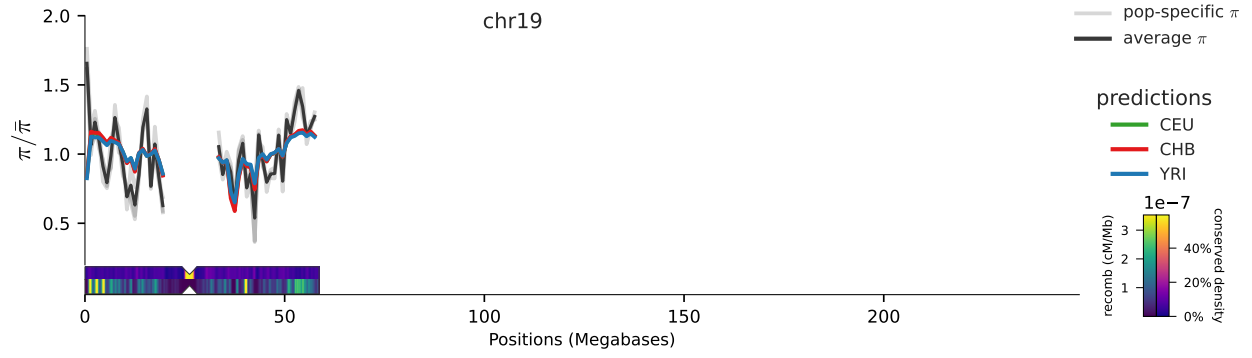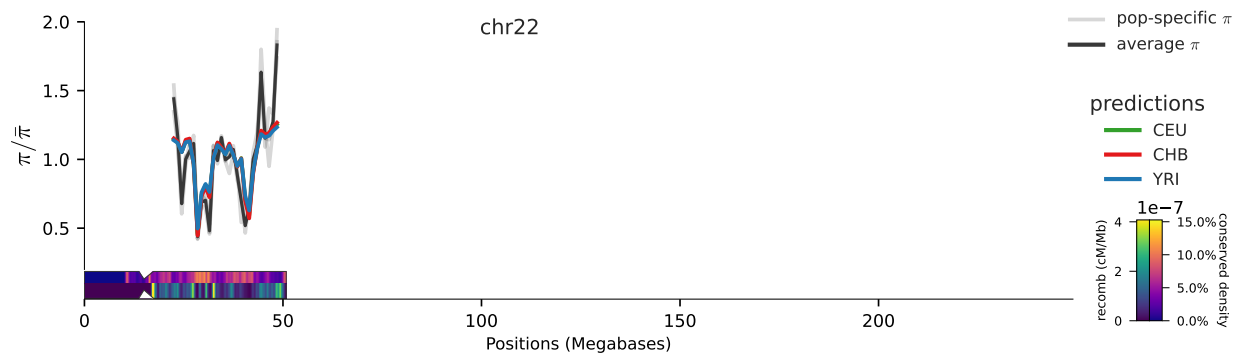
# 6 All Model Fits

## 6.1 Phastcons Priority, Full Track

**CEU**

```
ML estimates: (whole genome)
number jackknife samples: 280
standard error method: moving block jackknife
negative log-likelihood: 197825873893.4953
number of successful starts: 10000 (100.0% total)
pi_0 = 0.0011279 (0.00113, 0.00113)
pi  = 0.000801495
mut. rate = 2.35e-08  (1.98e-08, 2.72e-08)
Ne = 28,197 (if mu=1e-8), Ne = 14,098 (if mu=2e-8)
R2 = 63.0106% (in-sample)  61.7759% (out-sample)
W =
```

| | CDS | gene | other | phastcons |
|------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1e-08 | 0.000 (−0.0410, 0.0410) | 0.901 (0.6521, 1.1503) | 0.984 (0.9739, 0.9938) | 0.034 (−2.8217, 2.8898) |
| 1e-07 | 0.000 (−0.1314, 0.1314) | 0.026 (−0.2223, 0.2745) | 0.000 (−0.0096, 0.0096) | 0.376 (−2.4935, 3.2456) |
| 1e-06 | 0.000 (−0.0088, 0.0088) | 0.000 (−0.0018, 0.0027) | 0.000 (−0.0001, 0.0001) | 0.001 (−0.0308, 0.0318) |
| 1e-05 | 0.000 (−0.0002, 0.0002) | 0.019 (0.0133, 0.0249) | 0.000 (−0.0000, 0.0000) | 0.000 (−0.0003, 0.0003) |
| 0.0001 | 0.000 (−0.0000, 0.0000) | 0.000 (−0.0000, 0.0000) | 0.000 (−0.0000, 0.0000) | 0.000 (−0.0000, 0.0000) |
| 0.001 | 0.000 (−0.0000, 0.0000) | 0.000 (−0.0000, 0.0000) | 0.000 (−0.0000, 0.0000) | 0.472 (0.4050, 0.5388) |
| 0.01 | 0.000 (−0.0001, 0.0001) | 0.000 (−0.0000, 0.0000) | 0.000 (−0.0000, 0.0000) | 0.046 (0.0351, 0.0571) |
| 0.1 | 1.000 (0.8620, 1.1380) | 0.053 (0.0454, 0.0608) | 0.016 (0.0138, 0.0185) | 0.071 (0.0571, 0.0858) |

**CHB**

```
ML estimates: (whole genome)
number jackknife samples: 280
standard error method: moving block jackknife
negative log-likelihood: 61821783504.67356
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00106467 (0.00106, 0.00107)
pi  = 0.000752601
mut. rate = 2.557e-08  (-8.06e-10, 5.2e-08)
Ne = 26,616 (if mu=1e-8), Ne = 13,308 (if mu=2e-8)
R2 = 61.0847% (in-sample)  59.2923% (out-sample)
W =
```

| | CDS | gene | other | phastcons |
|------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1e-08 | 0.000 (−0.5839, 0.5839) | 0.931 (0.6381, 1.2238) | 0.985 (0.9661, 1.0044) | 0.478 (−2.5565, 3.5128) |
| 1e-07 | 0.000 (−0.2506, 0.2506) | 0.005 (−0.2858, 0.2949) | 0.000 (−0.0165, 0.0165) | 0.002 (−3.0192, 3.0225) |
| 1e-06 | 0.000 (−0.0119, 0.0119) | 0.000 (−0.0043, 0.0045) | 0.000 (−0.0002, 0.0002) | 0.000 (−0.0172, 0.0177) |
| 1e-05 | 0.000 (−0.0001, 0.0001) | 0.013 (−0.0013, 0.0268) | 0.000 (−0.0000, 0.0000) | 0.000 (−0.0002, 0.0002) |
| 0.0001 | 0.000 (−0.0000, 0.0000) | 0.000 (−0.0000, 0.0000) | 0.000 (−0.0000, 0.0000) | 0.000 (−0.0000, 0.0000) |
| 0.001 | 0.000 (−0.0000, 0.0000) | 0.003 (0.0005, 0.0049) | 0.000 (−0.0000, 0.0000) | 0.430 (0.1533, 0.7071) |
| 0.01 | 0.000 (−0.0001, 0.0001) | 0.000 (−0.0000, 0.0000) | 0.000 (−0.0000, 0.0000) | 0.090 (0.0316, 0.1477) |
| 0.1 | 1.000 (0.3574, 1.6426) | 0.049 (0.0174, 0.0804) | 0.015 (0.0053, 0.0243) | 0.000 (−0.0005, 0.0005) |

## YRI

```
ML estimates: (whole genome)
number jackknife samples: 280
standard error method: moving block jackknife
negative log-likelihood: 250656989872.19766
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00147277 (0.00147, 0.00147)
pi  = 0.00106447
mut. rate = 1.708e-08  (-4.08e-08, 7.49e-08)
Ne = 36,819 (if mu=1e-8), Ne = 18,409 (if mu=2e-8)
R2 = 68.1177% (in-sample)  68.1947% (out-sample)
W =
        CDS                   gene                   other                  phastcons
------  --------------------- ---------------------  ---------------------  ---------------------
1e-08   1.000 (-3.4930, 5.4930) 0.902 (0.6716, 1.1325) 0.979 (0.9416, 1.0156) 0.000 (-1.2400, 1.2407)
1e-07   0.000 (-4.4936, 4.4936) 0.000 (-0.1820, 0.1820) 0.000 (-0.0194, 0.0194) 0.004 (-0.5459, 0.5537)
1e-06   0.000 (-0.0107, 0.0107) 0.000 (-0.0024, 0.0024) 0.000 (-0.0003, 0.0003) 0.001 (-0.0353, 0.0365)
1e-05   0.000 (-0.0001, 0.0001) 0.042 (-0.0177, 0.1024) 0.000 (-0.0000, 0.0000) 0.000 (-0.0005, 0.0006)
0.0001  0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000) 0.000 (-0.0001, 0.0001)
0.001   0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000) 0.523 (-0.2146, 1.2611)
0.01    0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000) 0.212 (-0.0866, 0.5104)
0.1     0.000 (-0.0006, 0.0006) 0.056 (-0.0228, 0.1339) 0.021 (-0.0088, 0.0516) 0.260 (-0.1070, 0.6270)
```

## 6.2   Phastcons Priority, Sparse Track

### CEU

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 197826612684.87143
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00112308
pi  = 0.000801495
mut. rate = 2.019e-08
Ne = 28,076 (if mu=1e-8), Ne = 14,038 (if mu=2e-8)
R2 = 63.0152% (in-sample)  61.9834% (out-sample)
W =
        CDS    gene    phastcons
------  -----  ------  -----------
1e-08   0      0.913       0.227
1e-07   0      0           0.004
1e-06   0.004  0           0
1e-05   0      0.019       0
0.0001  0      0           0
0.001   0      0           0.546
0.01    0      0           0.061
0.1     0.996  0.069       0.161
```

## CHB

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 61822016128.79776
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00106069
pi  = 0.000752601
mut. rate = 5.887e-08
Ne = 26,517 (if mu=1e-8), Ne = 13,258 (if mu=2e-8)
R2 = 61.0804% (in-sample)  59.4419% (out-sample)
W =
         CDS    gene    phastcons
------   -----  ------  -----------
1e-08    0.617  0.968       0.744
1e-07    0.01   0           0.02
1e-06    0      0           0
1e-05    0      0.005       0
0.0001   0      0           0
0.001    0      0.001       0.185
0.01     0      0           0.051
0.1      0.373  0.026       0
```

## YRI

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 250657908564.79294
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00146801
pi  = 0.00106447
mut. rate = 1.864e-08
Ne = 36,700 (if mu=1e-8), Ne = 18,350 (if mu=2e-8)
R2 = 68.1582% (in-sample)  68.4462% (out-sample)
W =
         CDS    gene    phastcons
------   -----  ------  -----------
1e-08    1      0.907       0
1e-07    0      0.001       0
1e-06    0      0           0
1e-05    0      0.035       0
0.0001   0      0           0
0.001    0      0           0.477
0.01     0      0           0.199
0.1      0      0.057       0.323
```

## 6.3 Feature Priority, Full Track

**CEU**

```
ML estimates: (whole genome)
number jackknife samples: 457
standard error method: moving block jackknife
negative log-likelihood: 197848627167.32043
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00115698 (0.00116, 0.00116)
pi  = 0.000801495
mut. rate = 4.1e-08  (2.2e-08, 6e-08)
Ne = 28,924 (if mu=1e-8), Ne = 14,462 (if mu=2e-8)
R2 = 60.3026% (in-sample)  58.5808% (out-sample)
W =
        CDS                     gene                    other                   phastcons
------  ---------------------   ---------------------   ---------------------   ---------------------
1e-08   0.107 (-1.0244, 1.2375) 0.946 (0.7974, 1.0947)  1.000 (0.9991, 1.0009)  0.000 (-0.2751, 0.2757)
1e-07   0.000 (-1.0319, 1.0319) 0.001 (-0.1448, 0.1458) 0.000 (-0.0009, 0.0009) 0.000 (-0.2482, 0.2482)
1e-06   0.000 (-0.0087, 0.0087) 0.000 (-0.0013, 0.0021) 0.000 (-0.0000, 0.0000) 0.000 (-0.0150, 0.0150)
1e-05   0.000 (-0.0001, 0.0001) 0.023 (0.0088, 0.0368)  0.000 (-0.0000, 0.0000) 0.000 (-0.0003, 0.0003)
0.0001  0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000)
0.001   0.144 (0.0831, 0.2059)  0.014 (0.0080, 0.0196)  0.000 (-0.0000, 0.0000) 0.229 (0.1327, 0.3256)
0.01    0.000 (-0.0012, 0.0012) 0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000) 0.086 (0.0462, 0.1253)
0.1     0.749 (0.4358, 1.0622)  0.016 (0.0095, 0.0233)  0.000 (-0.0000, 0.0000) 0.685 (0.3987, 0.9709)
```

**CHB**

```
ML estimates: (whole genome)
number jackknife samples: 454
standard error method: moving block jackknife
negative log-likelihood: 61827828856.16418
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00108991 (0.00109, 0.00109)
pi  = 0.000752601
mut. rate = 3.755e-08  (1.55e-08, 5.96e-08)
Ne = 27,247 (if mu=1e-8), Ne = 13,623 (if mu=2e-8)
R2 = 58.7328% (in-sample)  56.4805% (out-sample)
W =
        CDS                     gene                    other                   phastcons
------  ---------------------   ---------------------   ---------------------   ---------------------
1e-08   0.088 (-0.9233, 1.0989) 0.940 (0.7665, 1.1139)  1.000 (0.9988, 1.0012)  0.006 (-0.3608, 0.3730)
1e-07   0.010 (-0.8568, 0.8763) 0.000 (-0.1685, 0.1691) 0.000 (-0.0011, 0.0011) 0.003 (-0.2610, 0.2674)
1e-06   0.000 (-0.0075, 0.0078) 0.000 (-0.0021, 0.0021) 0.000 (-0.0000, 0.0000) 0.000 (-0.0145, 0.0151)
1e-05   0.000 (-0.0001, 0.0001) 0.023 (0.0037, 0.0422)  0.000 (-0.0000, 0.0000) 0.000 (-0.0003, 0.0003)
0.0001  0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000)
0.001   0.156 (0.0732, 0.2384)  0.017 (0.0081, 0.0260)  0.000 (-0.0000, 0.0000) 0.244 (0.1185, 0.3696)
0.01    0.000 (-0.0139, 0.0140) 0.000 (-0.0000, 0.0000) 0.000 (-0.0000, 0.0000) 0.254 (0.1215, 0.3871)
0.1     0.746 (0.3590, 1.1338)  0.020 (0.0094, 0.0298)  0.000 (-0.0000, 0.0000) 0.492 (0.2392, 0.7448)
```

## YRI

```
ML estimates: (whole genome)
number jackknife samples: 280
standard error method: moving block jackknife
negative log-likelihood: 250679665159.79324
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00151818 (0.00152, 0.00152)
pi  = 0.00106447
mut. rate = 4.195e-08  (2.76e-08, 5.63e-08)
Ne = 37,954 (if mu=1e-8), Ne = 18,977 (if mu=2e-8)
R2 = 65.5097% (in-sample)  64.9049% (out-sample)
W =
        CDS                    gene                   other                  phastcons
------  ---------------------  ---------------------  ---------------------  ---------------------
1e-08   0.364 (-0.4599, 1.1874)  0.931 (0.8360, 1.0260)  1.000 (0.9993, 1.0007)  0.003 (-0.1973, 0.2040)
1e-07   0.001 (-0.7930, 0.7942)  0.000 (-0.0901, 0.0901)  0.000 (-0.0007, 0.0007)  0.000 (-0.1669, 0.1675)
1e-06   0.000 (-0.0038, 0.0040)  0.000 (-0.0009, 0.0009)  0.000 (-0.0000, 0.0000)  0.000 (-0.0106, 0.0109)
1e-05   0.000 (-0.0001, 0.0001)  0.039 (0.0240, 0.0532)  0.000 (-0.0000, 0.0000)  0.000 (-0.0002, 0.0002)
0.0001  0.000 (-0.0000, 0.0000)  0.000 (-0.0000, 0.0000)  0.000 (-0.0000, 0.0000)  0.000 (-0.0000, 0.0000)
0.001   0.091 (0.0617, 0.1212)  0.011 (0.0077, 0.0150)  0.000 (-0.0000, 0.0000)  0.151 (0.1030, 0.1986)
0.01    0.000 (-0.0001, 0.0001)  0.000 (-0.0001, 0.0001)  0.000 (-0.0000, 0.0000)  0.328 (0.2249, 0.4316)
0.1     0.544 (0.3725, 0.7157)  0.019 (0.0130, 0.0250)  0.000 (-0.0000, 0.0000)  0.517 (0.3548, 0.6796)
```

## 6.4  Feature Priority, Sparse Track

### CEU

```
ML estimates: (whole genome)
number jackknife samples: 27
standard error method: moving block jackknife
negative log-likelihood: 197848626646.91736
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00115698 (0.00116, 0.00116)
pi  = 0.000801495
mut. rate = 4.104e-08  (3.87e-08, 4.34e-08)
Ne = 28,924 (if mu=1e-8), Ne = 14,462 (if mu=2e-8)
R2 = 60.3027% (in-sample)  58.5823% (out-sample)
W =
        CDS                    gene                   phastcons
------  ---------------------  ---------------------  ---------------------
1e-08   0.107 (0.0490, 0.1649)  0.947 (0.9402, 0.9539)  0.001 (-0.0478, 0.0491)
1e-07   0.000 (-0.0273, 0.0279)  0.000 (-0.0052, 0.0052)  0.000 (-0.0073, 0.0075)
1e-06   0.000 (-0.0003, 0.0006)  0.000 (-0.0001, 0.0001)  0.001 (-0.0008, 0.0019)
1e-05   0.000 (-0.0000, 0.0000)  0.023 (0.0187, 0.0268)  0.000 (-0.0000, 0.0000)
0.0001  0.000 (-0.0000, 0.0000)  0.000 (-0.0000, 0.0000)  0.000 (-0.0000, 0.0000)
0.001   0.144 (0.1365, 0.1522)  0.014 (0.0130, 0.0146)  0.229 (0.2150, 0.2428)
0.01    0.000 (-0.0001, 0.0001)  0.000 (-0.0000, 0.0000)  0.086 (0.0733, 0.0981)
0.1     0.748 (0.7078, 0.7887)  0.016 (0.0155, 0.0173)  0.684 (0.6463, 0.7219)
```

## CHB

```
ML estimates: (whole genome)
number jackknife samples: 26
standard error method: moving block jackknife
negative log-likelihood: 61827828699.43107
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00108991 (0.00109, 0.00109)
pi  = 0.000752601
mut. rate = 3.719e-08  (3.2e-08, 4.24e-08)
Ne = 27,247 (if mu=1e-8), Ne = 13,623 (if mu=2e-8)
R2 = 58.7329% (in-sample)  56.4805% (out-sample)
W =
```

| | CDS | gene | phastcons |
|------|---------------------|---------------------|---------------------|
| 1e-08 | 0.089 (-0.0565, 0.2340) | 0.938 (0.9196, 0.9573) | 0.000 (-0.1223, 0.1223) |
| 1e-07 | 0.000 (-0.0861, 0.0861) | 0.001 (-0.0154, 0.0180) | 0.000 (-0.0145, 0.0146) |
| 1e-06 | 0.000 (-0.0005, 0.0010) | 0.000 (-0.0001, 0.0002) | 0.000 (-0.0009, 0.0009) |
| 1e-05 | 0.000 (-0.0000, 0.0000) | 0.023 (0.0203, 0.0260) | 0.000 (-0.0000, 0.0000) |
| 0.0001 | 0.000 (-0.0000, 0.0000) | 0.000 (-0.0000, 0.0000) | 0.000 (-0.0000, 0.0000) |
| 0.001 | 0.157 (0.1380, 0.1766) | 0.017 (0.0151, 0.0194) | 0.246 (0.2161, 0.2767) |
| 0.01 | 0.000 (-0.0004, 0.0006) | 0.000 (-0.0000, 0.0000) | 0.257 (0.2252, 0.2883) |
| 0.1 | 0.754 (0.6611, 0.8461) | 0.020 (0.0173, 0.0222) | 0.497 (0.4358, 0.5578) |

## YRI

```
ML estimates: (whole genome)
number jackknife samples: 17
standard error method: moving block jackknife
negative log-likelihood: 250679664591.4161
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00151817 (0.00152, 0.00152)
pi  = 0.00106447
mut. rate = 4.179e-08  (3.84e-08, 4.52e-08)
Ne = 37,954 (if mu=1e-8), Ne = 18,977 (if mu=2e-8)
R2 = 65.5099% (in-sample)  64.9049% (out-sample)
W =
```

| | CDS | gene | phastcons |
|------|---------------------|---------------------|---------------------|
| 1e-08 | 0.360 (0.3063, 0.4146) | 0.931 (0.9239, 0.9375) | 0.000 (-0.0647, 0.0647) |
| 1e-07 | 0.002 (-0.0234, 0.0265) | 0.000 (-0.0051, 0.0052) | 0.000 (-0.0097, 0.0097) |
| 1e-06 | 0.000 (-0.0003, 0.0003) | 0.000 (-0.0000, 0.0001) | 0.000 (-0.0009, 0.0009) |
| 1e-05 | 0.000 (-0.0000, 0.0000) | 0.039 (0.0359, 0.0417) | 0.000 (-0.0000, 0.0000) |
| 0.0001 | 0.000 (-0.0000, 0.0000) | 0.000 (-0.0000, 0.0000) | 0.000 (-0.0000, 0.0000) |
| 0.001 | 0.092 (0.0849, 0.0987) | 0.011 (0.0105, 0.0122) | 0.151 (0.1401, 0.1627) |
| 0.01 | 0.000 (-0.0000, 0.0000) | 0.000 (-0.0000, 0.0000) | 0.329 (0.3048, 0.3540) |
| 0.1 | 0.546 (0.5054, 0.5870) | 0.019 (0.0176, 0.0205) | 0.519 (0.4804, 0.5580) |

## 6.5   CADD 6%, Full Track

### CEU

```
ML estimates: (whole genome)
number jackknife samples: 280
standard error method: moving block jackknife
negative log-likelihood: 197826953189.02365
number of successful starts: 20000 (100.0% total)
pi_0 = 0.00109999 (0.00110, 0.00110)
pi  = 0.000801495
mut. rate = 1.537e-08  (8e-09, 2.27e-08)
Ne = 27,499 (if mu=1e-8), Ne = 13,749 (if mu=2e-8)
R2 = 62.9394% (in-sample)  61.7433% (out-sample)
W =
        cadd6                   other
------  ----------------------  ----------------------
1e-08   0.000 (-0.2792, 0.2795)  0.965 (0.9473, 0.9832)
1e-07   0.000 (-0.2355, 0.2355)  0.000 (-0.0109, 0.0109)
1e-06   0.000 (-0.0371, 0.0371)  0.000 (-0.0001, 0.0001)
1e-05   0.000 (-0.0011, 0.0013)  0.000 (-0.0000, 0.0000)
0.0001  0.026 (0.0129, 0.0393)   0.000 (-0.0000, 0.0000)
0.001   0.555 (0.3257, 0.7834)   0.000 (-0.0000, 0.0000)
0.01    0.000 (-0.0001, 0.0001)  0.000 (-0.0000, 0.0000)
0.1     0.419 (0.2460, 0.5921)   0.035 (0.0204, 0.0491)
```

### CHB

```
ML estimates: (whole genome)
number jackknife samples: 280
standard error method: moving block jackknife
negative log-likelihood: 61821323176.26884
number of successful starts: 20000 (100.0% total)
pi_0 = 0.00103624 (0.00104, 0.00104)
pi  = 0.000752601
mut. rate = 1.508e-08  (-2.43e-09, 3.26e-08)
Ne = 25,905 (if mu=1e-8), Ne = 12,952 (if mu=2e-8)
R2 = 61.3115% (in-sample)  59.6228% (out-sample)
W =
        cadd6                   other
------  ----------------------  ----------------------
1e-08   0.000 (-0.6278, 0.6278)  0.961 (0.9187, 1.0036)
1e-07   0.007 (-0.4928, 0.5076)  0.000 (-0.0227, 0.0227)
1e-06   0.009 (-0.0393, 0.0570)  0.000 (-0.0002, 0.0002)
1e-05   0.000 (-0.0014, 0.0015)  0.000 (-0.0000, 0.0000)
0.0001  0.004 (-0.0082, 0.0171)  0.000 (-0.0000, 0.0000)
0.001   0.670 (0.0898, 1.2509)   0.000 (-0.0000, 0.0000)
0.01    0.000 (-0.0001, 0.0001)  0.000 (-0.0000, 0.0000)
0.1     0.309 (0.0412, 0.5766)   0.039 (0.0051, 0.0726)
```

## YRI

```
ML estimates: (whole genome)
number jackknife samples: 280
standard error method: moving block jackknife
negative log-likelihood: 250659830283.90668
number of successful starts: 20000 (100.0% total)
pi_0 = 0.00145893 (0.00146, 0.00146)
pi  = 0.00106447
mut. rate = 1.461e-08  (8.61e-09, 2.06e-08)
Ne = 36,473 (if mu=1e-8), Ne = 18,236 (if mu=2e-8)
R2 = 68.0839% (in-sample)  68.0751% (out-sample)
W =
        cadd6                    other
------  ----------------------  ----------------------
1e-08   0.000 (-0.2318, 0.2327)  0.958 (0.9399, 0.9768)
1e-07   0.023 (-0.2328, 0.2797)  0.000 (-0.0083, 0.0089)
1e-06   0.001 (-0.0153, 0.0176)  0.000 (-0.0001, 0.0001)
1e-05   0.000 (-0.0003, 0.0004)  0.000 (-0.0000, 0.0000)
0.0001  0.065 (0.0378, 0.0921)   0.000 (-0.0000, 0.0000)
0.001   0.407 (0.2491, 0.5649)   0.000 (-0.0000, 0.0000)
0.01    0.078 (0.0439, 0.1126)   0.000 (-0.0000, 0.0000)
0.1     0.425 (0.2604, 0.5891)   0.041 (0.0253, 0.0573)
```

## 6.6  CADD 6%, Sparse Track

### CEU

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 197829116669.21103
number of successful starts: 20000 (100.0% total)
pi_0 = 0.00107623
pi  = 0.000801495
mut. rate = 2.121e-08
Ne = 26,905 (if mu=1e-8), Ne = 13,452 (if mu=2e-8)
R2 = 62.8555% (in-sample)  61.6574% (out-sample)
W =
        cadd6
------  -------
1e-08   0
1e-07   0
1e-06   0
1e-05   0
0.0001  0.014
0.001   0.408
0.01    0
0.1     0.578
```

## CHB

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 61822132892.58568
number of successful starts: 20000 (100.0% total)
pi_0 = 0.00101201
pi  = 0.000752601
mut. rate = 2.124e-08
Ne = 25,300 (if mu=1e-8), Ne = 12,650 (if mu=2e-8)
R2 = 61.2064% (in-sample)  59.5382% (out-sample)
W =
         cadd6
------  -------
1e-08     0
1e-07     0
1e-06     0
1e-05     0
0.0001    0
0.001     0.477
0.01      0
0.1       0.523
```

## YRI

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 250663447300.65836
number of successful starts: 20000 (100.0% total)
pi_0 = 0.00142415
pi  = 0.00106447
mut. rate = 2.091e-08
Ne = 35,603 (if mu=1e-8), Ne = 17,801 (if mu=2e-8)
R2 = 68.028% (in-sample)  68.0414% (out-sample)
W =
         cadd6
------  -------
1e-08     0
1e-07     0
1e-06     0
1e-05     0
0.0001    0.038
0.001     0.302
0.01      0.033
0.1       0.627
```

## 6.7 CADD 6%, Full Track

**CEU**

```
ML estimates: (whole genome)
number jackknife samples: 140
standard error method: moving block jackknife
negative log-likelihood: 197834651568.3971
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00109606 (0.00110, 0.00110)
pi  = 0.000801495
mut. rate = 1.157e-08  (9.01e-09, 1.41e-08)
Ne = 27,401 (if mu=1e-8), Ne = 13,700 (if mu=2e-8)
R2 = 62.1245% (in-sample)  60.8009% (out-sample)
W =
        cadd8                    other
------  -----------------------  -----------------------
1e-08   0.000 (-0.0835, 0.0835)  0.952 (0.9415, 0.9624)
1e-07   0.004 (-0.1451, 0.1523)  0.000 (-0.0043, 0.0043)
1e-06   0.000 (-0.0145, 0.0150)  0.000 (-0.0001, 0.0001)
1e-05   0.000 (-0.0005, 0.0005)  0.000 (-0.0000, 0.0000)
0.0001  0.052 (0.0393, 0.0656)   0.000 (-0.0000, 0.0000)
0.001   0.537 (0.4316, 0.6417)   0.000 (-0.0000, 0.0000)
0.01    0.000 (-0.0000, 0.0000)  0.000 (-0.0000, 0.0000)
0.1     0.407 (0.3276, 0.4867)   0.048 (0.0386, 0.0575)
```

**CHB**

```
ML estimates: (whole genome)
number jackknife samples: 140
standard error method: moving block jackknife
negative log-likelihood: 61823378586.32404
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00103245 (0.00103, 0.00103)
pi  = 0.000752601
mut. rate = 1.113e-08  (4.22e-09, 1.8e-08)
Ne = 25,811 (if mu=1e-8), Ne = 12,905 (if mu=2e-8)
R2 = 60.6265% (in-sample)  58.8768% (out-sample)
W =
        cadd8                    other
------  -----------------------  -----------------------
1e-08   0.000 (-0.3716, 0.3716)  0.944 (0.9141, 0.9736)
1e-07   0.000 (-0.2895, 0.2903)  0.001 (-0.0075, 0.0094)
1e-06   0.000 (-0.0376, 0.0378)  0.000 (-0.0001, 0.0001)
1e-05   0.000 (-0.0005, 0.0005)  0.000 (-0.0000, 0.0000)
0.0001  0.026 (0.0069, 0.0447)   0.000 (-0.0000, 0.0000)
0.001   0.690 (0.3451, 1.0351)   0.000 (-0.0000, 0.0000)
0.01    0.000 (-0.0001, 0.0001)  0.000 (-0.0000, 0.0000)
0.1     0.284 (0.1418, 0.4254)   0.055 (0.0275, 0.0829)
```

## YRI

```
ML estimates: (whole genome)
number jackknife samples: 140
standard error method: moving block jackknife
negative log-likelihood: 250669695762.61414
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00145867 (0.00146, 0.00146)
pi  = 0.00106447
mut. rate = 1.013e-08  (5.65e-09, 1.46e-08)
Ne = 36,466 (if mu=1e-8), Ne = 18,233 (if mu=2e-8)
R2 = 67.1076% (in-sample)  66.9835% (out-sample)
W =
        cadd8                   other
------  ----------------------  ----------------------
1e-08   0.000 (-0.3206, 0.3212)  0.932 (0.9054, 0.9583)
1e-07   0.002 (-0.1725, 0.1762)  0.000 (-0.0062, 0.0062)
1e-06   0.000 (-0.0130, 0.0130)  0.000 (-0.0001, 0.0001)
1e-05   0.000 (-0.0004, 0.0004)  0.000 (-0.0000, 0.0000)
0.0001  0.103 (0.0627, 0.1425)  0.000 (-0.0000, 0.0000)
0.001   0.438 (0.2744, 0.6008)  0.000 (-0.0000, 0.0000)
0.01    0.000 (-0.0003, 0.0003)  0.000 (-0.0000, 0.0000)
0.1     0.458 (0.2870, 0.6282)  0.068 (0.0427, 0.0936)
```

## 6.8   CADD 6%, Sparse Track

### CEU

```
ML estimates: (whole genome)
number jackknife samples: 140
standard error method: moving block jackknife
negative log-likelihood: 197836763801.90414
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00107057 (0.00107, 0.00107)
pi  = 0.000801495
mut. rate = 1.603e-08  (1.57e-08, 1.63e-08)
Ne = 26,764 (if mu=1e-8), Ne = 13,382 (if mu=2e-8)
R2 = 62.061% (in-sample)  60.5937% (out-sample)
W =
        cadd8
------  ----------------------
1e-08   0.000 (-0.0171, 0.0171)
1e-07   0.000 (-0.0007, 0.0007)
1e-06   0.000 (-0.0000, 0.0000)
1e-05   0.000 (-0.0000, 0.0000)
0.0001  0.032 (0.0268, 0.0372)
0.001   0.400 (0.3894, 0.4110)
0.01    0.000 (-0.0000, 0.0000)
0.1     0.568 (0.5575, 0.5781)
```

## CHB

```
ML estimates: (whole genome)
number jackknife samples: 140
standard error method: moving block jackknife
negative log-likelihood: 61824178668.93152
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00100597 (0.00101, 0.00101)
pi  = 0.000752601
mut. rate = 1.61e-08  (-4.15e-08, 7.37e-08)
Ne = 25,149 (if mu=1e-8), Ne = 12,574 (if mu=2e-8)
R2 = 60.5692% (in-sample)  58.6497% (out-sample)
W =
        cadd8
------  ----------------------
1e-08   0.000 (-0.7779, 0.7779)
1e-07   0.000 (-0.0032, 0.0033)
1e-06   0.000 (-0.0008, 0.0008)
1e-05   0.000 (-0.0000, 0.0000)
0.0001  0.011 (-0.0013, 0.0236)
0.001   0.492 (0.1090, 0.8748)
0.01    0.000 (-0.0000, 0.0000)
0.1     0.497 (0.1101, 0.8838)
```

## YRI

```
ML estimates: (whole genome)
number jackknife samples: 140
standard error method: moving block jackknife
negative log-likelihood: 250673972070.93787
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00141668 (0.00142, 0.00142)
pi  = 0.00106447
mut. rate = 1.569e-08  (1.49e-08, 1.64e-08)
Ne = 35,416 (if mu=1e-8), Ne = 17,708 (if mu=2e-8)
R2 = 67.0084% (in-sample)  66.6718% (out-sample)
W =
        cadd8
------  ----------------------
1e-08   0.000 (-0.0445, 0.0445)
1e-07   0.000 (-0.0136, 0.0136)
1e-06   0.000 (-0.0002, 0.0002)
1e-05   0.000 (-0.0000, 0.0000)
0.0001  0.059 (0.0526, 0.0655)
0.001   0.298 (0.2808, 0.3142)
0.01    0.000 (-0.0000, 0.0000)
0.1     0.643 (0.6132, 0.6737)
```

## 6.9 Phastcons Priority (small DFE grid), Full Track

### CEU

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 197835157685.628
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00105552
pi  = 0.000801495
mut. rate = 1.913e-08
Ne = 26,387 (if mu=1e-8), Ne = 13,193 (if mu=2e-8)
R2 = 62.2548% (in-sample)  61.0392% (out-sample)
W =
```

|         | CDS | gene  | other | phastcons |
| ------- | --- | ----- | ----- | --------- |
| 1e-08   | 1   | 0.947 | 1     | 0.001     |
| 1e-07   | 0   | 0     | 0     | 0         |
| 1e-06   | 0   | 0     | 0     | 0         |
| 1e-05   | 0   | 0.03  | 0     | 0         |
| 0.0001  | 0   | 0     | 0     | 0         |
| 0.001   | 0   | 0     | 0     | 0.413     |
| 0.01    | 0   | 0.023 | 0     | 0.586     |

### CHB

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 61824235925.00332
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00100129
pi  = 0.000752601
mut. rate = 1.869e-08
Ne = 25,032 (if mu=1e-8), Ne = 12,516 (if mu=2e-8)
R2 = 60.4486% (in-sample)  58.7636% (out-sample)
W =
```

|         | CDS | gene  | other | phastcons |
| ------- | --- | ----- | ----- | --------- |
| 1e-08   | 1   | 0.926 | 1     | 0         |
| 1e-07   | 0   | 0.013 | 0     | 0         |
| 1e-06   | 0   | 0     | 0     | 0         |
| 1e-05   | 0   | 0.031 | 0     | 0         |
| 0.0001  | 0   | 0     | 0     | 0         |
| 0.001   | 0   | 0     | 0     | 0.457     |
| 0.01    | 0   | 0.03  | 0     | 0.543     |

**YRI**

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 250667315910.91177
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00139482
pi  = 0.00106447
mut. rate = 2.033e-08
Ne = 34,870 (if mu=1e-8), Ne = 17,435 (if mu=2e-8)
R2 = 67.3564% (in-sample)  67.274% (out-sample)
W =
         CDS    gene    other    phastcons
------  -----  ------  -------  -----------
1e-08      1   0.956      1        0.005
1e-07      0   0          0        0.007
1e-06      0   0          0        0.002
1e-05      0   0.028      0        0
0.0001     0   0          0        0.027
0.001      0   0          0        0.224
0.01       0   0.016      0        0.735
```

## 6.10   Feature Priority (small DFE grid), Full Track

**CEU**

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 197860660949.1676
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00109138
pi  = 0.000801495
mut. rate = 3.046e-08
Ne = 27,284 (if mu=1e-8), Ne = 13,642 (if mu=2e-8)
R2 = 59.1052% (in-sample)  57.2393% (out-sample)
W =
         CDS    gene    other    phastcons
------  -----  ------  -------  -----------
1e-08  0.412  0.887      1        0
1e-07  0      0          0        0
1e-06  0      0          0        0
1e-05  0      0.074      0        0
0.0001 0      0          0        0
0.001  0.066  0.008      0        0.136
0.01   0.521  0.031      0        0.863
```

## CHB

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 61830715800.047676
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00103388
pi  = 0.000752601
mut. rate = 3.172e-08
Ne = 25,847 (if mu=1e-8), Ne = 12,923 (if mu=2e-8)
R2 = 57.8792% (in-sample)  55.631% (out-sample)
W =
```

| | CDS | gene | other | phastcons |
|------|-----|------|-------|-----------|
| 1e-08 | 0.432 | 0.898 | 1 | 0 |
| 1e-07 | 0 | 0 | 0 | 0 |
| 1e-06 | 0 | 0 | 0 | 0 |
| 1e-05 | 0 | 0.064 | 0 | 0 |
| 0.0001 | 0 | 0 | 0 | 0 |
| 0.001 | 0.052 | 0.012 | 0 | 0.166 |
| 0.01 | 0.516 | 0.026 | 0 | 0.834 |

## YRI

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 250692005501.01343
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00144148
pi  = 0.00106447
mut. rate = 3.349e-08
Ne = 36,037 (if mu=1e-8), Ne = 18,018 (if mu=2e-8)
R2 = 64.2844% (in-sample)  63.6077% (out-sample)
W =
```

| | CDS | gene | other | phastcons |
|------|-----|------|-------|-----------|
| 1e-08 | 0.72 | 0.875 | 1 | 0 |
| 1e-07 | 0.033 | 0.003 | 0 | 0 |
| 1e-06 | 0 | 0 | 0 | 0 |
| 1e-05 | 0 | 0.085 | 0 | 0 |
| 0.0001 | 0 | 0 | 0 | 0 |
| 0.001 | 0.1 | 0.003 | 0 | 0.07 |
| 0.01 | 0.148 | 0.033 | 0 | 0.93 |

## 6.11 CADD 6% (small DFE grid), Full Track

### CEU

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 197839911915.8117
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00102951
pi  = 0.000801495
mut. rate = 1.635e-08
Ne = 25,737 (if mu=1e-8), Ne = 12,868 (if mu=2e-8)
R2 = 61.767% (in-sample)  60.2433% (out-sample)
W =
         cadd6    other
------  -------  -------
1e-08    0.002    0.997
1e-07    0.006    0
1e-06    0        0
1e-05    0        0
0.0001   0.049    0
0.001    0.325    0
0.01     0.617    0.003
```

### CHB

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 61824516495.02472
number of successful starts: 10000 (100.0% total)
pi_0 = 0.000977579
pi  = 0.000752601
mut. rate = 1.598e-08
Ne = 24,439 (if mu=1e-8), Ne = 12,219 (if mu=2e-8)
R2 = 60.4032% (in-sample)  58.4847% (out-sample)
W =
         cadd6    other
------  -------  -------
1e-08    0        0.992
1e-07    0.001    0.001
1e-06    0        0
1e-05    0        0
0.0001   0.03     0
0.001    0.44     0
0.01     0.529    0.008
```

## YRI

```
ML estimates: (whole genome)
standard error method: None
negative log-likelihood: 250675195472.4508
number of successful starts: 10000 (100.0% total)
pi_0 = 0.00136706
pi  = 0.00106447
mut. rate = 1.556e-08
Ne = 34,176 (if mu=1e-8), Ne = 17,088 (if mu=2e-8)
R2 = 66.6755% (in-sample)  66.5047% (out-sample)
W =
         cadd6   other
------  -------  -------
1e-08    0        0.994
1e-07    0.006    0
1e-06    0        0
1e-05    0        0
0.0001   0.099    0
0.001    0.115    0
0.01     0.78     0.006
```