Testing for Ancient Selection Using Cross-population Allele Frequency Differentiation

Fernando Racimo[1]
Department of Integrative Biology, University of California, Berkeley, California 94720

ABSTRACT A powerful way to detect selection in a population is by modeling local allele frequency changes in a particular region of the genome under scenarios of selection and neutrality and finding which model is most compatible with the data. A previous method based on a cross-population composite likelihood ratio (XP-CLR) uses an outgroup population to detect departures from neutrality that could be compatible with hard or soft sweeps, at linked sites near a beneficial allele. However, this method is most sensitive to recent selection and may miss selective events that happened a long time ago. To overcome this, we developed an extension of XP-CLR that jointly models the behavior of a selected allele in a three-population tree. Our method - called "3-population composite likelihood ratio" (3P-CLR) - outperforms XP-CLR when testing for selection that occurred before two populations split from each other and can distinguish between those events and events that occurred specifically in each of the populations after the split. We applied our new test to population genomic data from the 1000 Genomes Project, to search for selective sweeps that occurred before the split of Yoruba and Eurasians, but after their split from Neanderthals, and that could have led to the spread of modern-human-specific phenotypes. We also searched for sweep events that occurred in East Asians, Europeans, and the ancestors of both populations, after their split from Yoruba. In both cases, we are able to confirm a number of regions identified by previous methods and find several new candidates for selection in recent and ancient times. For some of these, we also find suggestive functional mutations that may have driven the selective events.

KEYWORDS composite likelihood; Denisova; Neanderthal; population differentiation; positive selection

GENETIC hitchhiking will distort allele frequency patterns at regions of the genome linked to a beneficial allele that is rising in frequency (Smith and Haigh 1974). This is known as a selective sweep. If the sweep is restricted to a particular population and does not affect other closely related populations, one can detect such an event by looking for extreme patterns of localized population differentiation, like high values of F_{st} at a specific locus (Lewontin and Krakauer 1973). This and other related statistics have been used to scan the genomes of present-day humans from different populations, to detect signals of recent positive selection (Akey *et al.* 2002; Weir *et al.* 2005; Oleksyk *et al.* 2008; Yi *et al.* 2010).

Once it became possible to sequence entire genomes of archaic humans (like Neanderthals) (Green *et al.* 2010; Meyer *et al.* 2012; Prüfer *et al.* 2014), researchers also began to search for selective sweeps that occurred in the ancestral population of all present-day humans. For example, Green *et al.* (2010) searched for genomic regions with a depletion of derived alleles in a low-coverage Neanderthal genome, relative to what would be expected given the derived allele frequency in present-day humans. This is a pattern that would be consistent with a sweep in present-day humans. Later on, Prüfer *et al.* (2014) developed a hidden Markov model (HMM) that could identify regions where Neanderthals fall outside of all present-day human variation (also called "external regions") and are therefore likely to have been affected by ancient sweeps in early modern humans. They applied their method to a high-coverage Neanderthal genome. Then, they ranked these regions by their genetic length, to find segments that were extremely long and therefore highly compatible with a selective sweep. Finally, Racimo *et al.* (2014) used summary statistics calculated in the neighborhood of sites that were ancestral in archaic humans but fixed derived in all or almost all present-day humans, to test whether any of these sites could be compatible with a selective sweep model. While these methods harnessed different summaries of the patterns of differentiation left by sweeps, they did not

doi: 10.1534/genetics.115.178095
Manuscript received May 11, 2015; accepted for publication November 18, 2015; published Early Online November 20, 2015.
Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178095/-/DC1.
[1]Address for correspondence: 1506 Oxford St., Berkeley, CA 94709.
E-mail: fernandoracimo@gmail.com

attempt to explicitly model the process by which these patterns are generated over time.

Chen *et al.* (2010) developed a test called "cross-population composite likelihood ratio" (XP-CLR), which is designed to test for selection in one population after its split from a second, outgroup, population t_{AB} generations ago. It does so by modeling the evolutionary trajectory of an allele under linked selection and under neutrality and then comparing the likelihood of the data for each of the two models. The method detects local allele frequency differences that are compatible with the linked selection model (Smith and Haigh 1974), along windows of the genome.

XP-CLR is a powerful test for detecting selective events restricted to one population. However, it provides little information about when these events happened, as it models all sweeps as if they had immediately occurred in the present generation. Additionally, if one is interested in selective sweeps that took place before two populations *a* and *b* split from each other, one would have to run XP-CLR separately on each population, with a third outgroup population *c* that split from the ancestor of *a* and *b* t_{ABC} generations ago (with $t_{ABC} > t_{AB}$). Then, one would need to check that the signal of selection appears in both tests. This may miss important information about correlated allele frequency changes shared by *a* and *b*, but not by *c*, limiting the power to detect ancient events.

To overcome this, we developed an extension of XP-CLR that jointly models the behavior of an allele in all three populations, to detect selective events that occurred before or after the closest two populations split from each other. Below we briefly review the modeling framework of XP-CLR and describe our new test, which we call the "3-population composite likelihood ratio" (3P-CLR). In *Results*, we show this method outperforms XP-CLR, when testing for selection that occurred before the split of two populations, and can distinguish between those events and events that occurred after the split, unlike XP-CLR. We then apply the method to population genomic data from the 1000 Genomes Project (Abecasis *et al.* 2012), to search for selective sweep events that occurred before the split of Yoruba and Eurasians, but after their split from Neanderthals. We also use it to search for selective sweeps that occurred in the Eurasian ancestral population and to distinguish those from events that occurred specifically in East Asians or specifically in Europeans.

Materials and Methods

XP-CLR

First, we review the procedure used by XP-CLR to model the evolution of allele frequency changes of two populations *a* and *b* that split from each other t_{AB} generations ago (Figure 1A). For neutral SNPs, Chen *et al.* (2010) use an approximation to the Wright–Fisher diffusion dynamics (Nicholson *et al.* 2002). Namely, the frequency of a SNP in a population *a* (p_A) in the present is treated as a random variable governed by a normal distribution with mean equal to the frequency in the ancestral population (β) and variance proportional to the drift time ω from the ancestral to the present population,

$$p_A|\beta \sim N(\beta, \omega\beta(1-\beta)), \tag{1}$$

where $\omega = t_{AB}/(2N_e)$ and N_e is the effective size of population *A*.

This is a Brownian motion approximation to the Wright–Fisher model, as the drift increment to variance is constant across generations. If a SNP is segregating in both populations—*i.e.*, has not hit the boundaries of fixation or extinction—this process is time reversible. Thus, one can model the frequency of the SNP in population *a* with a normal distribution having mean equal to the frequency in population *b* and variance proportional to the sum of the drift time (ω) between *a* and the ancestral population and the drift time between *b* and the ancestral population (ψ):

$$p_A|p_B \sim N(p_B, (\omega+\psi)p_B(1-p_B)). \tag{2}$$

For SNPs that are linked to a beneficial allele that has produced a sweep in population *a* only, Chen *et al.* (2010) model the allele as evolving neutrally until the present and then apply a transformation to the normal distribution that depends on the distance to the selected allele *r* and the strength of selection *s* (Fay and Wu 2000; Durrett and Schweinsberg 2004). Let $c = 1 - q_0^{r/s}$, where q_0 is the frequency of the beneficial allele in population *a* before the sweep begins. The frequency of a neutral allele is expected to increase from *p* to $1 - c + cp$ if the allele is linked to the beneficial allele, and this occurs with probability equal to the frequency of the neutral allele (*p*) before the sweep begins. Otherwise, the frequency of the neutral allele is expected to decrease from *p* to *cp*. This leads to the following transformation of the normal distribution,

$$\begin{aligned} f(p_A|p_B, r, s, \omega, \psi) &= \frac{1}{\sqrt{2\pi}\sigma}\frac{p_A + c - 1}{c^2} e^{-\left((p_A+c-1-cp_B)^2/2c^2\sigma^2\right)} I_{[1-c,1]}(p_A) \\ &+ \frac{1}{\sqrt{2\pi}\sigma}\frac{c - p_A}{c^2} e^{-\left((p_A-cp_B)^2/2c^2\sigma^2\right)} I_{[0,c]}(p_A), \end{aligned} \tag{3}$$

where $\sigma^2 = (\omega+\psi)p_B(1-p_B)$ and $I_{[x,y]}(z) = 1$ on the interval $[x, y]$ and 0 otherwise.

For $s \to 0$ or $r \gg s$, this distribution converges to the neutral case. Let $\mathbf{v}$ be the vector of all drift times that are relevant to the scenario we are studying. In this case, it will be equal to (ω, ψ) but in more complex cases below, it may include additional drift times. Let $\mathbf{r}$ be the vector of recombination fractions between the beneficial alleles and each of the SNPs within a window of arbitrary size. We can then calculate the product of likelihoods over all *k* SNPs in that window for either the neutral or the linked selection model, after binomial sampling of alleles from the population frequency and conditioning on the event that the allele is segregating in the population:

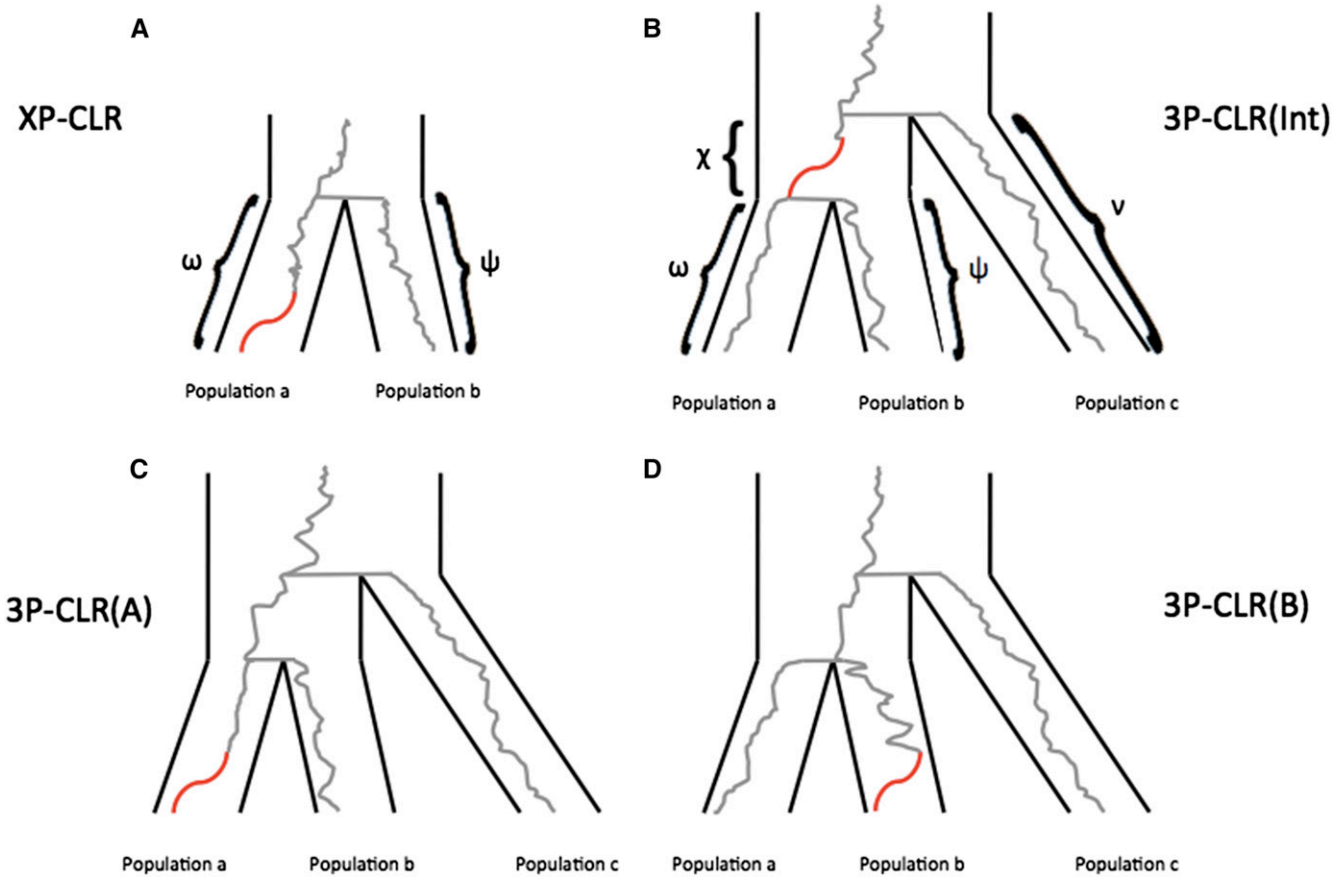

Figure 1 Schematic tree of selective sweeps detected by XP-CLR and 3P-CLR. While XP-CLR can use only two populations (an outgroup and a test) to detect selection (A), 3P-CLR can detect selection in the ancestral branch of two populations [3P-CLR(Int) (B)] or on the branches specific to each population [3P-CLR(A) (C) and 3P-CLR(B) (D)]. The Greek letters denote the known drift times for each branch of the population tree.

$$\mathrm{CL}_{\text{XP-CLR}}(\mathbf{r},\mathbf{v},s) = \prod_{j=1}^{k} \frac{\int_0^1 f\left(p_A^j | p_B^j, \mathbf{v}, s, r^j\right) \binom{n}{m_j} \left(p_A^j\right)^{m_j} \left(1-p_A^j\right)^{n-m_j} dp_A^j}{\int_0^1 f\left(p_A^j | p_B^j, \mathbf{v}, s, r^j\right) dp_A^j}. \quad (4)$$

This is a composite likelihood (Lindsay 1988; Varin *et al.*, 2011), because we are ignoring the correlation in frequencies produced by linkage among SNPs that is not strictly due to proximity to the beneficial SNP. We note that the denominator in the above equation is not explicitly stated in Chen *et al.* (2010) for ease of notation, but appears in the published online implementation of the method.

Finally, we obtain a composite-likelihood-ratio statistic $S_{\text{XP-CLR}}$ of the hypothesis of linked selection over the hypothesis of neutrality:

$$S_{\text{XP-CLR}} = 2\left[\sup_{\mathbf{r},\mathbf{v},s} \log(\mathrm{CL}_{\text{XP-CLR}}(\mathbf{r},\mathbf{v},s)) - \sup_{\mathbf{v}} \log(\mathrm{CL}_{\text{XP-CLR}}(\mathbf{r},\mathbf{v},s=0))\right]. \quad (5)$$

For ease of computation, Chen *et al.* (2010) assume that $\mathbf{r}$ is given (via a recombination map) instead of maximizing the likelihood with respect to it, and we do so too. Furthermore, they empirically estimate $\mathbf{v}$ using F_2 statistics (Patterson *et al.*, 2012) calculated over the whole genome and assume selection is not strong or frequent enough to affect their genome-wide values. Therefore, the likelihoods in the above equation are maximized only with respect to the selection coefficient, using a grid of coefficients on a logarithmic scale.

3P-CLR

We are interested in the case where a selective event occurred more anciently than the split of two populations (a and b) from each other, but more recently than their split from a third population c (Figure 1B). We begin by modeling p_A and p_B as evolving from an unknown common ancestral frequency β:

$$p_A|\beta,\omega \sim N\left(\beta, \omega\beta(1-\beta)\right) \quad (6)$$

$$p_B|\beta,\psi \sim N\left(\beta, \psi\beta(1-\beta)\right). \quad (7)$$

Let χ be the drift time separating the most recent common ancestor of a and b from the most recent common ancestor of a, b, and c. Additionally, let ν be the drift time separating

population c in the present from the most recent common ancestor of a, b, and c. Given these parameters, we can treat β as an additional random variable that either evolves neutrally or is linked to a selected allele that swept immediately more anciently than the split of a and b. In both cases, the distribution of β will depend on the frequency of the allele in population c (p_C) in the present. In the neutral case,

$$f_{\text{neut}}(\beta|p_C,\nu,\chi) = N\big(p_C,(\nu+\chi)p_C(1-p_C)\big). \tag{8}$$

In the linked selection case,

$$\begin{aligned} f_{\text{sel}}&(\beta|p_C,\nu,\chi,r,s) \\ &= \frac{1}{\sqrt{2\pi}\kappa}\frac{\beta+c-1}{c^2}e^{-\left((\beta+c-1-cp_C)^2/2c^2\kappa^2\right)}I_{[1-c,1]}(\beta) \\ &\quad + \frac{1}{\sqrt{2\pi}\kappa}\frac{c-\beta}{c^2}e^{-\left((\beta-cp_C)^2/2c^2\kappa^2\right)}I_{[0,c]}(\beta), \end{aligned} \tag{9}$$

where $\kappa^2 = (\nu+\chi)p_C(1-p_C)$.

The frequencies in a and b given the frequency in c can be obtained by integrating β out. This leads to a density function that models selection in the ancestral population of a and b:

$$\begin{aligned} f(&p_A,p_B|p_C,\mathbf{v},r,s) \\ &= \int_0^1 f_{\text{neut}}(p_A|\beta,\omega)\, f_{\text{neut}}(p_B|\beta,\psi)\, f_{\text{sel}}(\beta|p_C,\nu,\chi,r,s)\, d\beta. \end{aligned} \tag{10}$$

Additionally, Equation 10 can be modified to test for selection that occurred specifically in one of the terminal branches that lead to a or b (Figure 1, C and D), rather than in the ancestral population of a and b. For example, the density of frequencies for a scenario of selection in the branch leading to a can be written as

$$\begin{aligned} f(&p_A,p_B|p_C,\mathbf{v},r,s) \\ &= \int_0^1 f_{\text{sel}}(p_A|\beta,\omega,r,s)\, f_{\text{neut}}(p_B|\beta,\psi) f_{\text{neut}}(\beta|p_C,\nu,\chi)\, d\beta. \end{aligned} \tag{11}$$

We henceforth refer to the version of 3P-CLR that is tailored to detect selection in the internal branch that is ancestral to a and b as 3P-CLR(Int). In turn, the versions of 3P-CLR that are designed to detect selection in each of the daughter populations a and b are designated as 3P-CLR(A) and 3P-CLR(B), respectively.

We can now calculate the probability density of specific allele frequencies in populations a and b, given that we observe m_C derived alleles in a sample of size n_C from population c,

$$f(p_A,p_B|m_C,\mathbf{v},r,s) = \int_0^1 f(p_A,p_B|p_C,\mathbf{v},r,s)\, f(p_C|m_C)\, dp_C \tag{12}$$

and

$$f(p_C|m_C) = \frac{1}{B(m_C,n_C-m_C+1)}p_C^{m_C-1}(1-p_C)^{n_C-m_C}, \tag{13}$$

where $B(x,y)$ is the Beta function. We note that Equation 13 assumes that the unconditioned density function for the population derived allele frequency $f(p_C)$ comes from the neutral infinite-sites model at equilibrium and is therefore equal to the product of a constant and $1/p_C$ (Ewens 2012).

Conditioning on the event that the site is segregating in the population, we can then calculate the probability of observing m_A and m_B derived alleles in a sample of size n_A from population a and a sample of size n_B from population b, respectively, given that we observe m_C derived alleles in a sample of size n_C from population c, using binomial sampling,

$$\begin{aligned} P(&m_A,m_B|m_C,\mathbf{v},r,s) \\ &= \frac{\int_0^1\int_0^1 P(m_A|p_A)\,P(m_B|p_B) f(p_A,p_B|m_C,\mathbf{v},r,s)\,dp_A dp_B}{\int_0^1\int_0^1 f(p_A,p_B|m_C,\mathbf{v},r,s)\,dp_A\,dp_B}, \end{aligned} \tag{14}$$

where

$$P(m_A|p_A) = \binom{n_A}{m_A}p_A^{m_A}(1-p_A)^{n_A-m_A} \tag{15}$$

and

$$P(m_B|p_B) = \binom{n_B}{m_B}p_B^{m_B}(1-p_B)^{n_B-m_B}. \tag{16}$$

This allows us to calculate a composite likelihood of the derived allele counts in a and b given the derived allele counts in c:

$$\text{CL}_{\text{3P-CLR}}(\mathbf{r},\mathbf{v},s) = \prod_{j=1}^{k} P\Big(m_A^j,m_B^j|m_C^j,\mathbf{v},r^j,s\Big). \tag{17}$$

As before, we can use this composite likelihood to produce a composite-likelihood-ratio statistic that can be calculated over regions of the genome to test the hypothesis of linked selection centered on a particular locus against the hypothesis of neutrality. Due to computational costs in numerical integration, we skip the sampling step for population c (Equation 13) in our implementation of 3P-CLR. In other words, we assume $p_C = m_C/n_C$, but this is also assumed in XP-CLR when computing its corresponding outgroup frequency. To perform the numerical integrations, we used the package Cubature (v.1.0.2). We implemented our method in a freely available C++ program that can be downloaded from https://github.com/ferracimo. The program requires the neutral drift parameters α, β, and $(\nu+\chi)$ to be specified as input. These can be obtained using F_3 statistics (Felsenstein 1981; Patterson *et al.* 2012), which have previously been implemented in programs like MixMapper (Lipson *et al.* 2013). For example, α can be obtained via $F_3(A;B,C)$, while $(\nu+\chi)$ can be

Table 1 Description of models tested

Model	Population where selection occurred	t_{AB}	t_{ABC}	t_M	s	N_e
A	Ancestral population	500	2,000	1,800	0.1	10,000
B	Ancestral population	1,000	4,000	2,500	0.1	10,000
C	Ancestral population	2,000	4,000	3,500	0.1	10,000
D	Ancestral population	3,000	8,000	5,000	0.1	10,000
E	Ancestral population	2,000	16,000	8,000	0.1	10,000
F	Ancestral population	4,000	16,000	8,000	0.1	10,000
I	Daughter population *a*	2,000	4,000	1,000	0.1	10,000
J	Daughter population *a*	3,000	8,000	2,000	0.1	10,000

All times are in generations. Selection in the "ancestral population" refers to a selective sweep where the beneficial mutation and fixation occurred before the split time of the two most closely related populations. Selection in "daughter population *a*" refers to a selective sweep that occurred in one of the two most closely related populations (*a*), after their split from each other. t_{AB}, split time (in generations ago) of populations *a* and *b*; t_{ABC}, split time of population *c* and the ancestral population of *a* and *b*; t_M, time at which the selected mutation is introduced; s, selection coefficient; N_e, effective population size.

obtained via $F_3(C;A,B)$. When computing F_3 statistics, we use only sites where population C is polymorphic, and so we correct for this ascertainment in the calculation. Another way of calculating these drift times is via $\partial a \partial i$ (Gutenkunst *et al.* 2009). Focusing on two populations at a time, we can fix one population's size and allow the split time and the other population's size to be estimated by the program, in this case using all polymorphic sites, regardless of which population they are segregating in. We then obtain the two drift times by scaling the inferred split time by the two different population sizes. We provide scripts in our github page for the user to obtain these drift parameters, using both of the above ways.

Results

Simulations

We generated simulations in SLiM (Messer 2013) to test the performance of XP-CLR and 3P-CLR in a three-population scenario. We first focused on the performance of 3P-CLR (Int) in detecting ancient selective events that occurred in the ancestral branch of two sister populations. We assumed that the population history had been correctly estimated (*i.e.*, the drift parameters and population topology were known). First, we simulated scenarios in which a beneficial mutation arose in the ancestor of populations *a* and *b*, before their split from each other but after their split from *c* (Table 1). Although both XP-CLR and 3P-CLR are sensitive to partial or soft sweeps [as they do not rely on extended patterns of haplotype homozygosity (Chen *et al.*, 2010)], we required the beneficial allele to have fixed before the split (at time t_{ab}) to ensure that the allele had not been lost by then and also to ensure that the sweep was restricted to the internal branch of the tree. We fixed the effective size of all three populations at $N_e = 10,000$. Each simulation consisted of a 5-cM region and the beneficial mutation occurred in the center of this region. The mutation rate was set at 2.5×10^{-8} per generation and the recombination rate between adjacent nucleotides was set at 10^{-8} per generation.

To make a fair comparison to 3P-CLR(Int), and given that XP-CLR is a two-population test, we applied XP-CLR in two ways. First, we pretended population *b* was not sampled, and so the "test" panel consisted of individuals from *a* only, while the "outgroup" consisted of individuals from *c*. In the second implementation (which we call "XP-CLR-avg"), we used the same outgroup panel, but pooled the individuals from *a* and *b* into a single panel, and this pooled panel was the test. The window size was set at 0.5 cM and the number of SNPs sampled between each window's central SNP was set at 600 (this number is large because it includes SNPs that are not segregating in the outgroup, which are later discarded). To speed up computation, and because we are largely interested in comparing the relative performance of the three tests under different scenarios, we used only 20 randomly chosen SNPs per window in all tests. We note, however, that the performance of all of these tests can be improved by using more SNPs per window.

Figure 2 shows receiver operating characteristic (ROC) curves comparing the sensitivity and specificity of 3P-CLR (Int), 3P-CLR(A), XP-CLR, and XP-CLR-avg in the first six demographic scenarios described in Table 1. Each ROC curve was made from 100 simulations under selection (with $s = 0.1$ for the central mutation) and 100 simulations under neutrality (with $s = 0$ and no fixation required). In each simulation, 100 haploid individuals (or 50 diploids) were sampled from population *a*, 100 individuals from population *b*, and 100 individuals from the outgroup population *c*. For each simulation, we took the maximum value at a region in the neighborhood of the central mutation (± 0.5 cM) and used those values to compute ROC curves under the two models.

When the split times are recent or moderately ancient (models A–D), 3P-CLR(Int) outperforms the two versions of XP-CLR. Furthermore, 3P-CLR(A) is the test that is least sensitive to selection in the internal branch as it is meant to detect selection only in the terminal branch leading to population *a*. When the split times are very ancient (models E and F), none of the tests perform well. The root mean-squared error (RMSE) of the genetic distance between the true selected site and the highest scored window is comparable across tests in all six scenarios (Supporting Information, Figure S5). 3P-CLR(Int) is the best test at finding the true location of the selected site in almost all demographic scenarios. We observe that we lose almost all power if we simulate demographic scenarios where the population size is

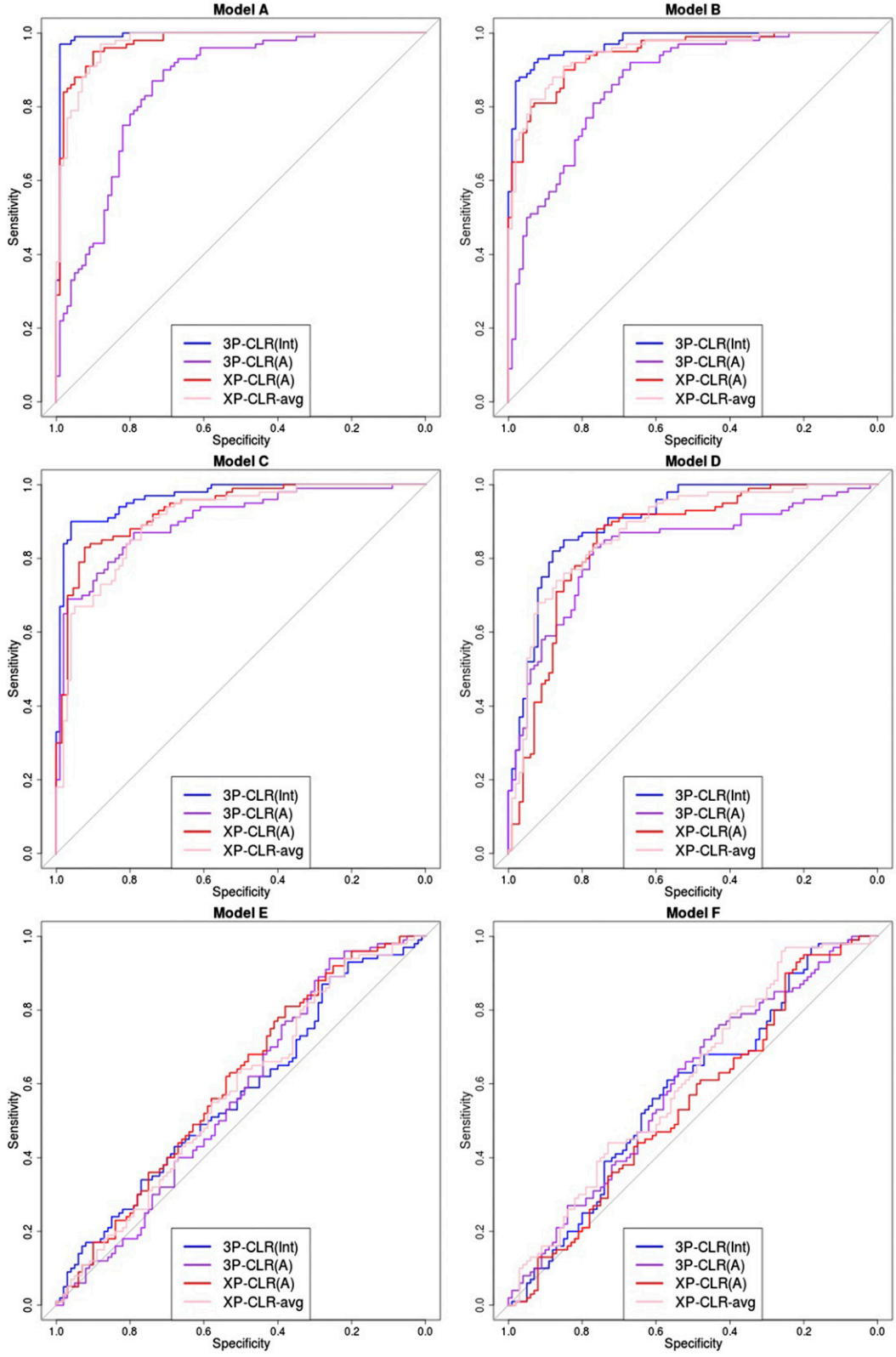

Figure 2 ROC curves for performance of 3P-CLR(Int), 3P-CLR(A), and two variants of XP-CLR in detecting selective sweeps that occurred before the split of two populations *a* and *b*, under different demographic models. In this case, the outgroup panel from population *c* contained 100 haploid genomes. The two sister population panels (from *a* and *b*) also have 100 haploid genomes each.

10 times smaller ($N_e = 1000$) (Figure S1). Additionally, we observe that the power and specificity of 3P-CLR decrease as the selection coefficient decreases (Figure S2).

We also simulated a situation in which only a few individuals are sequenced from the outgroup, while large numbers of sequences are available from the tests. Figure S3 and Figure S6 show the ROC curves and RMSE plots, respectively, for a scenario in which 100 individuals were sampled from the test populations but only 10 individuals (5 diploids) were sampled from the outgroup. Unsurprisingly, all tests have less power to detect selection when the split times and the selection events are recent to moderately ancient (models A–D). Interestingly though, when the split times and the selective events are very ancient (models E and F), both 3P-CLR and XP-CLR perform better when using a small outgroup panel (Figure S3) than when using a large outgroup panel (Figure 2). This is due to the Brownian motion approximation that these methods utilize. Under the Wright–Fisher model, the drift increment at generation t is proportional to $p(t)\times(1 - p(t))$, where $p(t)$ is the derived allele frequency. The derivative of this function gets smaller the closer $p(t)$ is to 0.5 (and is exactly 0 at that point). Small outgroup panels serve to filter out loci with allele frequencies far from 0.5, and so small changes in allele frequency will not affect the drift increment much, making Brownian motion a good approximation to the Wright–Fisher model. Indeed, when running 3P-CLR(Int) in a demographic scenario with very ancient split times (model E) and a large outgroup panel (100 sequences) but restricting only to sites that are at intermediate frequencies in the outgroup ($25\% \leq m_C/n_C \leq 75\%$), we find that performance is much improved relative to the case when we use all sites that are segregating in the outgroup (Figure S4).

Importantly, the usefulness of 3P-CLR(Int) resides not just in its performance at detecting selective sweeps in the ancestral population, but in its specific sensitivity to that particular type of events. Because the test relies on correlated allele frequency differences in both population a and population b (relative to the outgroup), selective sweeps that are specific to only one of the populations will not lead to high 3P-CLR (Int) scores, but will instead lead to high 3P-CLR(A) scores or 3P-CLR(B) scores, depending on where selection took place. Figure 3 shows ROC curves in two scenarios in which a selective sweep occurred only in population a (models I and J in Table 1), using 100 sampled individuals from each of the three populations. Here, XP-CLR performs well, but is outperformed by 3P-CLR(A). Furthermore, 3P-CLR(Int) shows almost no sensitivity to the recent sweep. For example, in model I, at a specificity of 90%, 3P-CLR(A) and XP-CLR(A) have 86% and 80% sensitivity, respectively, while at the same specificity, 3P-CLR(Int) has only 18% sensitivity. One can compare this to the same demographic scenario but with selection occurring in the ancestral population of a and b (model C, Figure 2), where at 90% specificity, 3P-CLR(A) and XP-CLR(A) have 72% and 84% sensitivity, respectively, while 3P-CLR(Int) has 90% sensitivity. We also observe that 3P-CLR(A) is the best test at finding the true location of the selected site when selection occurs in the terminal branch leading to population a (Figure S7).

Finally, we tested the behavior of 3P-CLR under selective scenarios that we did not explicitly model. First, we simulated a selective sweep in the outgroup population. We find that all three types of 3P-CLR statistics [3P-CLR(Int), 3P-CLR(A), and 3P-CLR(B)] are largely insensitive to this type of event, although 3P-CLR(Int) is relatively more sensitive than the other two. Second, we simulated two independent selective sweeps in populations a and b (convergent evolution). This results in elevated 3P-CLR(A) and 3P-CLR(B) statistics, but 3P-CLR(Int) remains largely insensitive (Figure S8). We note that 3P-CLR should not be used to detect selective events that occurred before the split of all three populations (*i.e.*, before the split of c and the ancestor of a and b), as it relies on allele frequency differences between the populations.

Selection in Eurasians

We first applied 3P-CLR to modern human data from phase 1 of the 1000 Genomes Project (Abecasis *et al.* 2012). We used the African–American recombination map (Hinch *et al.* 2011) to convert physical distances into genetic distances. We focused on Europeans - including Utah residents with European ancestry (CEU), Finnish (FIN), British (GBR), Spanish (IBS) and Toscani (TSI) - and East Asians - including Han Chinese (CHB), Southern Han Chinese (CHS) and Japanese (JPT) - as the two sister populations, using Yoruba (YRI) as the outgroup population (Figure S9A). We randomly sampled 100 individuals from each population and obtained sample derived allele frequencies every 10 SNPs in the genome. We then calculated likelihood-ratio statistics by a sliding-window approach, where we sampled a "central SNP" once every 10 SNPs. The central SNP in each window was the candidate beneficial SNP for that window. We set the window size to 0.25 cM and randomly sampled 100 SNPs from each window, centered around the candidate beneficial SNP. In each window, we calculated 3P-CLR to test for selection at three different branches of the population tree: the terminal branch leading to Europeans (3P-CLR Europe), the terminal branch leading to East Asians (3P-CLR East Asia), and the ancestral branch of Europeans and East Asians (3P-CLR Eurasia). Results are shown in Figure S10. For each scan, we selected the windows in the top 99.9% quantile of scores and merged them together if their corresponding central SNPs were contiguous, effectively resulting in overlapping windows being merged. Table S1, Table S2, and Table S3 show the top hits for Europeans, East Asians, and the ancestral Eurasian branch, respectively, while Table 2 shows the 10 strongest candidate regions for each population.

We observe several genes that were identified in previous selection scans. In the East Asian branch, one of the top hits is *EDAR*. Figure 4A shows that this gene appears to be under selection exclusively in this population branch. It codes for a protein involved in hair thickness and incisor tooth morphology (Fujimoto *et al.* 2008; Kimura *et al.* 2009) and has been

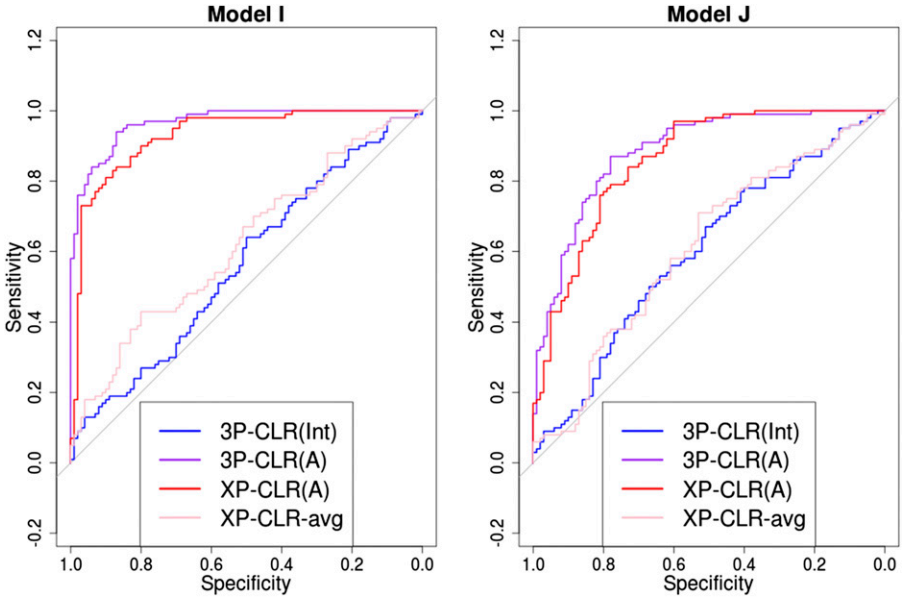

Figure 3 3P-CLR(Int) is tailored to detect selective events that happened before the split t_{ab}, so it is largely insensitive to sweeps that occurred after the split. ROC curves show performance of 3P-CLR(Int) and two variants of XP-CLR for models where selection occurred in population *a* after its split from *b*.

repeatedly identified as a candidate for a sweep in East Asians (Sabeti *et al.* 2007; Grossman *et al.* 2010).

Furthermore, 3P-CLR allows us to narrow down the specific time at which selection for previously found candidates occurred in the history of particular populations. For example, Chen *et al.* (2010) performed a scan of the genomes of East Asians, using XP-CLR with Yoruba as the outgroup, and identified a number of candidate genes. 3P-CLR confirms several of their loci when looking specifically at the East Asian branch: *OR56A1*, *OR56B4*, *OR52B2*, *SLC30A9*, *BBX*, *EPHB1*, *ACTN1*, and *XKR6*. However, when applied to the ancestral Eurasian branch, 3P-CLR finds some genes that were previously found in the XP-CLR analysis of East Asians, but that are not among the top hits in 3P-CLR applied to the East Asian branch: *COMMD3*, *BMI1*, *SPAG6*, *NGLY1*, *OXSM*, *CD226*, *ABCC12*, *ABCC11*, *LONP2*, *SIAH1*, *PPARA*, *PKDREJ*, *GTSE1*, *TRMU*, and *CELSR1*. This suggests selection in these regions occurred earlier, *i.e.*, before the European–East Asian split. Figure 4B shows a comparison between the 3P-CLR scores for the three branches in the region containing genes *BMI1* [a proto-oncogene (Siddique and Saleem 2012)] and *SPAG6* [involved in sperm motility (Sapiro *et al.* 2002)]. Here, the signal of Eurasia-specific selection is evidently stronger than the other two signals. Finally, we also find some candidates from Chen *et al.* (2010) that appear to be under selection in both the ancestral Eurasian branch and the East Asian daughter branch: *SFXN5*, *EMX1*, *SPR*, and *CYP26B1*. Interestingly, both CYP26B1 and CYP26A1 are very strong candidates for selection in the East Asian branch. These two genes lie in two different chromosomes, so they are not part of a gene cluster, but they both code for proteins that hydrolize retinoic acid, an important signaling molecule (White *et al.* 2000; Topletz *et al.* 2012).

Other selective events that 3P-CLR infers to have occurred in Eurasians include the region containing *HERC2* and *OCA2*, which are major determinants of eye color (Eiberg *et al.* 2008; Han *et al.* 2008; Branicki *et al.* 2009). There is also evidence that these genes underwent selection more recently in the history of Europeans (Mathieson *et al.* 2015), which could suggest an extended period of selection—perhaps influenced by migrations between Asia and Europe—or repeated selective events at the same locus.

When running 3P-CLR to look for selection specific to Europe, we find that *TYRP1*, which plays a role in human skin pigmentation (Halaban and Moellmann 1990), is among the top hits. This gene has been previously found to be under strong selection in Europe (Voight *et al.* 2006), using a statistic called *iHS*, which measures extended patterns of haplotype homozygosity that are characteristic of selective sweeps. Interestingly, a change in the gene *TYRP1* has also been found to cause a blonde hair phenotype in Melanesians (Kenny *et al.* 2012). Another of our top hits is the region containing *SH2B3*, which was identified previously as a candidate for selection in Europe based on both *iHS* and F_{st} (Pickrell *et al.* 2009). This gene contains a nonsynonymous SNP (rs3184504) segregating in Europeans. One of its alleles (the one in the selected haplotype) has been associated with celiac disease and type 1 diabetes (Todd *et al.* 2007; Hunt *et al.* 2008) but is also protective against bacterial infection (Zhernakova *et al.* 2010).

We used Gowinda (v1.12) (Kofler and Schlötterer 2012) to find enriched Gene Ontology (GO) categories among the regions in the 99.5% highest quantile for each branch score, relative to the rest of the genome [$P < 0.05$, false discovery rate (FDR) < 0.3]. The significantly enriched categories are listed in Table S4. In the East Asian branch, we find categories related to alcohol catabolism, retinol binding, vitamin metabolism, and epidermis development, among others. In the European branch, we find cuticle development and hydrogen peroxide metabolic process as enriched categories. We find

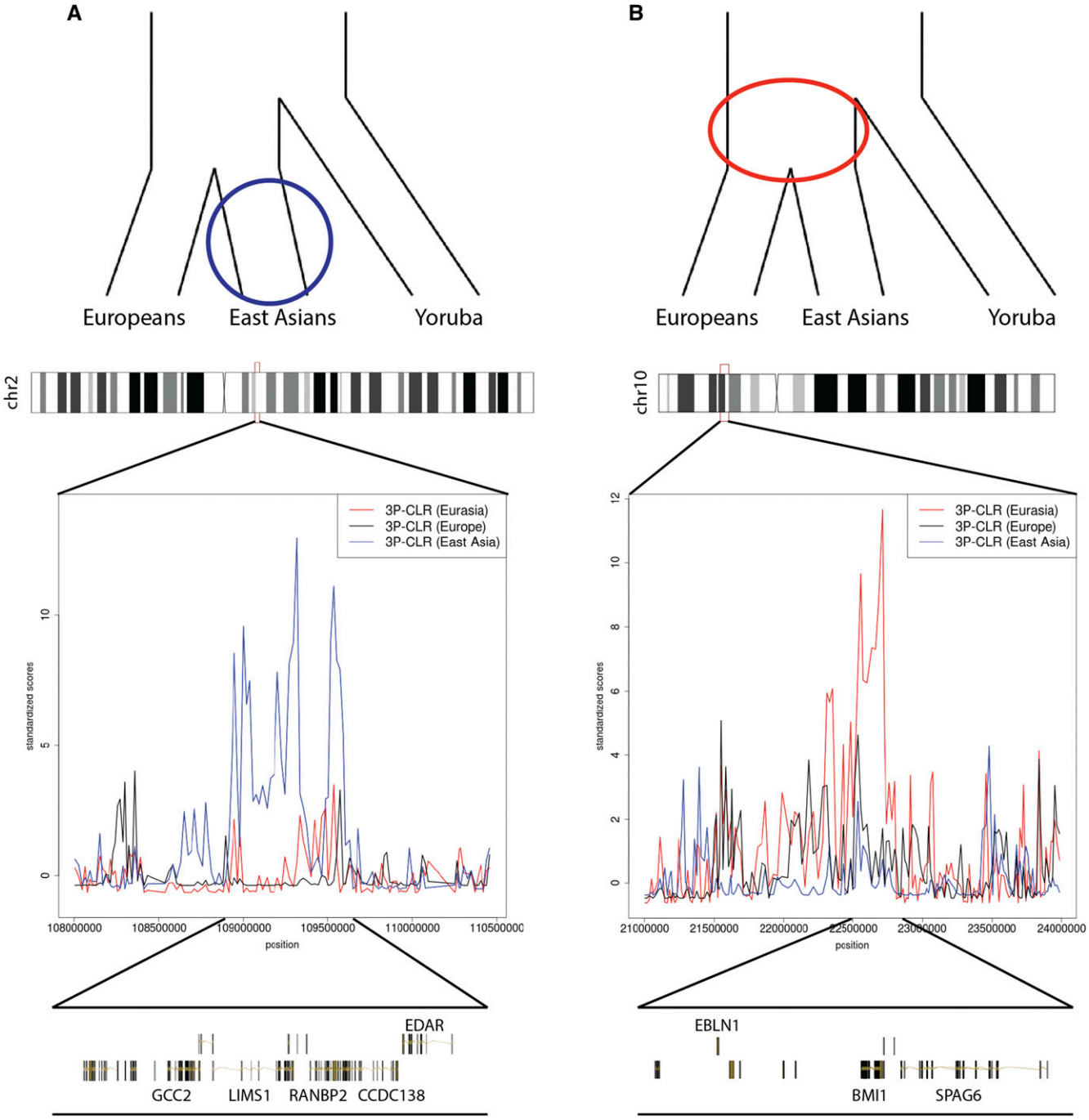

Figure 4 3P-CLR scan of Europeans (black), East Asians (blue), and the ancestral Eurasian population (red) reveals regions under selection in different branches of the population tree. To make a fair comparison, all 3P-CLR scores were standardized by substracting the chromosome-wide mean from each window and dividing the resulting score by the chromosome-wide standard deviation. (A) The region containing *EDAR* is a candidate for selection in the East Asian population. (B) The region containing genes *SPAG6* and *BMI1* is a candidate for selection in the ancestral population of Europeans and East Asians. The image was built using the GenomeGraphs package in Bioconductor.

no enriched categories in the Eurasian branch that pass the above cutoffs.

Selection in ancestral modern humans

We applied 3P-CLR to modern human data combined with recently sequenced archaic human data. We sought to find selective events that occurred in modern humans after their spit from archaic groups. We used the combined Neanderthal and Denisovan high-coverage genomes (Meyer *et al.* 2012; Prüfer *et al.* 2014) as the outgroup population, and, for our two test populations, we used Eurasians (CEU, FIN, GBR, IBS, TSI, CHB, CHS, and JPT) and YRI, again from phase 1 of the

Table 2 Genes from top 10 candidate regions for each of the branches on which 3P-CLR was run for the Eurasian population tree

Window size	Position (hg19)	Genes
European	Chr9:125424000–126089000	*ZBTB26*, *RABGAP1*, *GPR21*, *STRBP*, *OR1L1*, *OR1L3*, *OR1L4*, *OR1L6*, *OR5C1*, *PDCL*, *OR1K1*, *RC3H2*, *ZBTB6*
	Chr22:35,528,100–35,754,100	*HMGXB4*, *TOM1*
	Chr8:52,361,800–52,932,100	*PXDNL*, *PCMTD1*
	Chr2:74,450,100–74,972,700	*INO80B*, *WBP1*, *MOGS*, *MRPL53*, *CCDC142*, *TTC31*, *LBX2*, *PCGF1*, *TLX2*, *DQX1*, *AUP1*, *HTRA2*, *LOXL3*, *DOK1*, *M1AP*, *SEMA4F*, *SLC4A5*, *DCTN1*, *WDR54*, *RTKN*
	Chr1:35,382,000–36,592,200	*DLGAP3*, *ZMYM6NB*, *ZMYM6*, *ZMYM1*, *SFPQ*, *ZMYM4*, *KIAA0319L*, *NCDN*, *TFAP2E*, *PSMB2*, *C1orf216*, *CLSPN*, *AGO4*, *AGO1*, *AGO3*, *TEKT2*, *ADPRHL2*, *COL8A2*
	Chr15:29,248,000–29,338,300	*APBA2*
	Chr12:111,747,000–113,030,000	*BRAP*, *ACAD10*, *ALDH2*, *MAPKAPK5*, *TMEM116*, *ERP29*, *NAA25*, *TRAFD1*, *RPL6*, *PTPN11*, *RPH3A*, *CUX2*, *FAM109A*, *SH2B3*, *ATXN2*
	Chr9:90,909,300–91,210,000	*SPIN1*, *NXNL2*
	Chr19:33,504,200–33,705,700	*RHPN2*, *GPATCH1*, *WDR88*, *LRP3*, *SLC7A10*
	Chr9:30,085,400–31,031,600	—
East Asian	Chr15:63,693,900–64,188,300	*USP3*, *FBXL22*, *HERC1*
	Chr10:94,830,500–95,093,900	*CYP26A1*, *MYOF*
	Chr2:72,353,500–73,170,800	*CYP26B1*, *EXOC6B*, *SPR*, *EMX1*, *SFXN5*
	Chr2:72,353,500–73,170,800	*PCDH15*
	Chr1:234,209,000–234,396,000	*SLC35F3*
	Chr5:117,344,000–117,714,000	—
	Chr17:60,907,300–61,547,900	*TANC2*, *CYB561*
	Chr2:44,101,400–44,315,200	*ABCG8*, *LRPPRC*
	Chr11:6,028,090–6,191,240	*OR56A1*, *OR56B4*, *OR52B2*
	Chr2:108,905,000–109,629,000	*LIMS1*, *RANBP2*, *CCDC138*, *EDAR*, *SULT1C2*, *SULT1C4*, *GCC2*
Eurasian	Chr2:72,353,500–73,170,800	*CYP26B1*, *EXOC6B*, *SPR*, *EMX1*, *SFXN5*
	Chr20:53,876,700–54,056,200	—
	Chr10:22,309,300–22,799,200	*EBLN1*, *COMMD3*, *COMMD3-BMI1*, *BMI1*, *SPAG6*
	Chr3:25,726,300–26,012,000	*NGLY1*, *OXSM*
	Chr18:67,523,300–67,910,500	*CD226*, *RTTN*
	Chr10:65,794,400–66,339,100	—
	Chr11:39,587,400–39,934,300	—
	Chr7:138,806,000–139,141,000	*TTC26*, *UBN2*, *C7orf55*, *C7orf55-LUC7L2*, *LUC7L2*, *KLRG2*
	Chr9:90,909,300–91,202,200	*SPIN1*, *NXNL2*
	Chr4:41,454,200–42,195,300	*LIMCH1*, *PHOX2B*, *TMEM33*, *DCAF4L1*, *SLC30A9*, *BEND4*

All positions were rounded to the nearest 100 bp. Windows were merged together if the central SNPs that define them were contiguous.

1000 Genomes Project (Abecasis *et al.* 2012) (Figure S9B). As before, we randomly sampled 100 genomes for each of the two daughter populations at each site and tested for selective events that occurred more anciently than the split of Yoruba and Eurasians, but more recently than the split from Neanderthals. Figure S11 shows an ROC curve for a simulated scenario under these conditions, based on the history of population size changes inferred by the Pairwise Sequentially Markovian Coalescent (PSMC) model (Li and Durbin 2011; Prüfer *et al.* 2014), suggesting we should have power to detect strong ($s = 0.1$) selective events in the ancestral branch of present-day humans. We observe that 3P-CLR(Int) has similar power to XP-CLR and XP-CLR-avg at these timescales, but is less prone to also detect recent (postsplit) events, making it more specific to ancestral sweeps.

We ran 3P-CLR using 0.25-cM windows as above (Figure S13). As before, we selected the top 99.9% windows and merged them together if their corresponding central SNPs were contiguous (Table S5). The top 20 regions are in Table 3. Figure S13 shows that the outliers in the genome-wide distribution of scores are not strong. We wanted to verify that the density of scores was robust to the choice of window size. By using a larger window size (1 cM), we obtained a distribution with slightly more extreme outliers (Figure S12 and Figure S13). For that reason, we also show the top hits from this large-window run (Table S6 and Table 3), using a smaller density of SNPs (200/1 cM rather than 100/0.25 cM), due to costs in speed. To find putative candidates for the beneficial variants in each region, we queried the catalogs of modern human-specific high-frequency or fixed derived changes that are ancestral in the Neanderthal and/or the Denisova genomes (Castellano *et al.* 2014; Prüfer *et al.* 2014) and overlapped them with our regions.

We found several genes that were identified in previous studies that looked for selection in modern humans after their split from archaic groups (Green *et al.* 2010; Prüfer *et al.*

2014), including *SIPA1L1*, *ANAPC10*, *ABCE1*, *RASA1*, *CCNH*, *KCNJ3*, *HBP1*, *COG5*, *CADPS2*, *FAM172A*, *POU5F2*, *FGF7*, *RABGAP1*, *SMURF1*, *GABRA2*, *ALMS1*, *PVRL3*, *EHBP1*, *VPS54*, *OTX1*, *UGP2*, *GTDC1*, *ZEB2*, and *OIT3*. One of our strongest candidate genes among these is *SIPA1L1* (Figure 5A), which is in the first and the fourth highest-ranking region, when using 1- and 0.25-cM windows, respectively. The protein encoded by this gene (E6TP1) is involved in actin cytoskeleton organization and controls neural morphology (UniProt by similarity). Interestingly, it is also a target of degradation of the oncoproteins of high-risk papillomaviruses (Gao *et al.* 1999).

Another candidate gene is *ANAPC10* (Figure 5B). This gene codes for a core subunit of the cyclosome, which is involved in progression through the cell cycle (Pravtcheva and Wise 2001) and may play a role in oocyte maturation and human T-lymphotropic virus infection [KEGG pathway (Kanehisa and Goto 2000)]. *ANAPC10* is noteworthy because it was found to be significantly differentially expressed in humans compared to great apes and macaques: it is upregulated in the testes (Brawand *et al.* 2011). The gene also contains two intronic changes that are fixed derived in modern humans and ancestral in both Neanderthals and Denisovans and that have evidence for being highly disruptive, based on a composite score that combines conservation and regulatory data [PHRED-scaled *C* scores >11 (Kircher *et al.* 2014; Prüfer *et al.* 2014)]. The changes, however, appear not to lie in any obvious regulatory region (Rosenbloom *et al.* 2011; Dunham *et al.* 2012).

We also find *ADSL* among the list of candidates. This gene is known to contain a nonsynonymous change that is fixed in all present-day humans but homozygous ancestral in the Neanderthal genome, the Denisova genome, and two Neanderthal exomes (Castellano *et al.* 2014) (Figure 6A). It was previously identified as lying in a region with strong support for positive selection in modern humans, using summary statistics implemented in an ABC method (Racimo *et al.* 2014). The gene is interesting because it is one of the members of the Human Phenotype ontology category "aggression/hyperactivity" that is enriched for nonsynonymous changes that occurred in the modern human lineage after the split from archaic humans (Robinson *et al.* 2008; Castellano *et al.* 2014). *ADSL* codes for adenylosuccinase, an enzyme involved in purine metabolism (Van Keuren *et al.* 1987). A deficiency of adenylosuccinase can lead to apraxia, speech deficits, delays in development, and abnormal behavioral features, like hyperactivity and excessive laughter (Gitiaux *et al.* 2009). The nonsynonymous mutation (A429V) is in the C-terminal domain of the protein (Figure 6B) and lies in a highly conserved position [primate PhastCons = 0.953; GERP score = 5.67 (Siepel *et al.* 2005; Cooper *et al.* 2010; Kircher *et al.* 2014)]. The ancestral amino acid is conserved across the tetrapod phylogeny, and the mutation is only three residues away from the most common causative SNP for severe adenylosuccinase deficiency (Maaswinkel-Mooij *et al.* 1997; Marie *et al.* 1999; Kmoch *et al.* 2000; Race *et al.* 2000; Edery *et al.* 2003). The change has the highest probability of being disruptive to protein function, of all the nonsynonymous modern-human-specific changes that lie in the top-scoring regions (*C* score = 17.69). While *ADSL* is an interesting candidate and lies in the center of the inferred selected region (Figure 6A), there are other genes in the region too, including *TNRC6B* and *MKL1*. *TNRC6B* may be involved in miRNA-guided gene silencing (Meister *et al.* 2005), while *MKL1* may play a role in smooth muscle differentiation (Du *et al.* 2004) and has been associated with acute megakaryocytic leukemia (Mercher *et al.* 2001).

RASA1 was also a top hit in a previous scan for selection (Green *et al.* 2010) and was additionally inferred to have evidence in favor of selection in Racimo *et al.* (2014). The gene codes for a protein involved in the control of cellular differentiation (Trahey *et al.* 1988) and has a modern human-specific fixed nonsynonymous change (G70E). Human diseases associated with *RASA1* include basal cell carcinoma (Friedman *et al.* 1993) and arteriovenous malformation (Eerola *et al.* 2003; Hershkovitz *et al.* 2008).

The $GABA_A$ gene cluster in chromosome 4p12 is also among the top regions. The gene within the putatively selected region codes for a subunit (*GABRA2*) of the $GABA_A$ receptor, which is a ligand-gated ion channel that plays a key role in synaptic inhibition in the central nervous system (see review by Whiting *et al.* 1999). *GABRA2* is significantly associated with risk of alcohol dependence in humans (Edenberg *et al.* 2004), perception of pain (Knabl *et al.* 2008), and asthma (Xiang *et al.* 2007).

Two other candidate genes that may be involved in brain development are *FOXG1* and *CADPS2*. *FOXG1* was not identified in any of the previous selection scans and codes for a protein called forkhead box G1, which plays an important role during brain development. Mutations in this gene are associated with a slowdown in brain growth during childhood, resulting in microcephaly, which in turn causes various intellectual disabilities (Ariani *et al.* 2008; Mencarelli *et al.* 2010). *CADPS2* was identified in Green *et al.* (2010) as a candidate for selection and has been associated with autism (Sadakata and Furuichi 2010). The gene has been suggested to be specifically important in the evolution of all modern humans, as it was not found to be selected earlier in great apes or later in particular modern human populations (Crisci *et al.* 2011).

Finally, we find a signal of selection in a region containing the genes *EHBP1* and *OTX1*. This region was identified in both of the two previous scans for modern human selection (Green *et al.* 2010; Prüfer *et al.* 2014). *EHBP1* codes for a protein involved in endocytic trafficking (Guilherme *et al.* 2004) and has been associated with prostate cancer (Gudmundsson *et al.* 2008). *OTX1* is a homeobox family gene that may play a role in brain development (Gong *et al.* 2003). Interestingly, *EHBP1* contains a single-nucleotide intronic change (chr2:63206488) that is almost fixed in all present-day humans and homozygous ancestral in Neanderthal and Denisova (Prüfer *et al.* 2014). This change is also predicted to be highly disruptive (*C* score = 13.1) and lies in a position that is extremely conserved across primates (PhastCons = 0.942), mammals (PhastCons = 1), and vertebrates

Table 3 Genes from top 20 candidate regions for the modern human ancestral branch

Window size	Position (hg19)	Genes
0.25 cM (100 SNPs)	Chr2:95,561,200–96,793,700	*ZNF514*, *ZNF2*, *PROM2*, *KCNIP3*, *FAHD2A*, *TRIM43*, *GPAT2*, *ADRA2B*, *ASTL*, *MAL*, *MRPS5*
	Chr5:86,463,700–87,101,400	*RASA1*, *CCNH*
	Chr17:60,910,700–61,557,700	*TANC2*, *CYB561*, *ACE*
	Chr14:71,649,200–72,283,600	*SIPA1L1*
	Chr18:15,012,100–19,548,600	*ROCK1*, *GREB1L*, *ESCO1*, *SNRPD1*, *ABHD3*, *MIB1*
	Chr3:110,513,000–110,932,000	*PVRL3*
	Chr2:37,917,900–38,024,200	*CDC42EP3*
	Chr3:36,836,900–37,517,500	*TRANK1*, *EPM2AIP1*, *MLH1*, *LRRFIP2*, *GOLGA4*, *C3orf35*, *ITGA9*
	Chr7:106,642,000–10,7310,000	*PRKAR2B*, *HBP1*, *COG5*, *GPR22*, *DUS4L*, *BCAP29*, *SLC26A4*
	Chr12:96,823,000–97,411,500	*NEDD1*
	Chr2:200,639,000–201,340,000	*C2orf69*, *TYW5*, *C2orf47*, *SPATS2L*
	Chr1:66,772,600–66,952,600	*PDE4B*
	Chr10:37,165,100–38,978,800	*ANKRD30A*, *MTRNR2L7*, *ZNF248*, *ZNF25*, *ZNF33A*, *ZNF37A*
	Chr2:155,639,000–156,767,000	*KCNJ3*
	Chr17:56,379,200–57,404,800	*BZRAP1*, *SUPT4H1*, *RNF43*, *HSF5*, *MTMR4*, *SEPT4*, *C17orf47*, *TEX14*, *RAD51C*, *PPM1E*, *TRIM37*, *SKA2*, *PRR11*, *SMG8*, *GDPD1*
	Chr5:18,493,900–18,793,500	—
	Chr2:61,050,900–61,891,900	*REL*, *PUS10*, *PEX13*, *KIAA1841*, *AHSA2*, *USP34*, *XPO1*
	Chr22:40,360,300–41,213,400	*GRAP2*, *FAM83F*, *TNRC6B*, *ADSL*, *SGSM3*, *MKL1*, *MCHR1*, *SLC25A17*
	Chr2:98,996,400–99,383,400	*CNGA3*, *INPP4A*, *COA5*, *UNC50*, *MGAT4A*
	Chr4:13,137,000–13,533,100	*RAB28*
1 cM (200 SNPs)	Chr14:71,349,200–72,490,300	*PCNX*, *SIPA1L1*, *RGS6*
	Chr4:145,023,000–146,522,000	*GYPB*, *GYPA*, *HHIP*, *ANAPC10*, *ABCE1*, *OTUD4*, *SMAD1*
	Chr2:155,391,000–156,992,000	*KCNJ3*
	Chr5:92,415,600–94,128,600	*NR2F1*, *FAM172A*, *POU5F2*, *KIAA0825*, *ANKRD32*, *MCTP1*
	Chr7:106,401,000–107,461,000	*PIK3CG*, *PRKAR2B*, *HBP1*, *COG5*, *GPR22*, *DUS4L*, *BCAP29*, *SLC26A4*, *CBLL1*, *SLC26A3*
	Chr7:151,651,000–152,286,000	*GALNTL5*, *GALNT11*, *KMT2C*
	Chr2:144,393,000–145,305,000	*ARHGAP15*, *GTDC1*, *ZEB2*
	Chr19:16,387,600–16,994,000	*KLF2*, *EPS15L1*, *CALR3*, *C19orf44*, *CHERP*, *SLC35E1*, *MED26*, *SMIM7*, *TMEM38A*, *NWD1*, *SIN3B*
	Chr2:37,730,400–38,054,600	*CDC42EP3*
	Chr2:62,639,800–64,698,300	*TMEM17*, *EHBP1*, *OTX1*, *WDPCP*, *MDH1*, *UGP2*, *VPS54*, *PELI1*, *LGALSL*
	Chr10:36,651,400–44,014,800	*ANKRD30A*, *MTRNR2L7*, *ZNF248*, *ZNF25*, *ZNF33A*, *ZNF37A*, *ZNF33B*, *BMS1*, *RET*, *CSGALNACT2*, *RASGEF1A*, *FXYD4*, *HNRNPF*
	Chr1:26,703,800–27,886,000	*LIN28A*, *DHDDS*, *HMGN2*, *RPS6KA1*, *ARID1A*, *PIGV*, *ZDHHC18*, *SFN*, *GPN2*, *GPATCH3*, *NUDC*, *NR0B2*, *C1orf172*, *TRNP1*, *FAM46B*, *SLC9A1*, *WDTC1*, *TMEM222*, *SYTL1*, *MAP3K6*, *FCN3*, *CD164L2*, *GPR3*, *WASF2*, *AHDC1*
	Chr12:102,308,000–103,125,000	*DRAM1*, *CCDC53*, *NUP37*, *PARPBP*, *PMCH*, *IGF1*
	Chr2:132,628,000–133,270,000	*GPR39*
	Chr15:42,284,300–45,101,400	*PLA2G4E*, *PLA2G4D*, *PLA2G4F*, *VPS39*, *TMEM87A*, *GANC*, *CAPN3*, *ZNF106*, *SNAP23*, *LRRC57*, *HAUS2*, *STARD9*, *CDAN1*, *TTBK2*, *UBR1*, *EPB42*, *TMEM62*, *CCNDBP1*, *TGM5*, *TGM7*, *LCMT2*, *ADAL*, *ZSCAN29*, *TUBGCP4*, *TP53BP1*, *MAP1A*, *PPIP5K1*, *CKMT1B*, *STRC*, *CATSPER2*, *CKMT1A*, *PDIA3*, *ELL3*, *SERF2*, *SERINC4HYPK*, *MFAP1*, *WDR76*, *FRMD5*, *CASC4*, *CTDSPL2*, *EIF3J*, *SPG11*, *PATL2*, *B2M*, *TRIM69*
	Chr2:73,178,500–74,194,400	*SFXN5*, *RAB11FIP5*, *NOTO*, *SMYD5*, *PRADC1*, *CCT7*, *FBXO41*, *EGR4*, *ALMS1*, *NAT8*, *TPRKB*, *DUSP11*, *C2orf78*, *STAMBP*, *ACTG2*, *DGUOK*
	Chr5:54,193,000–55,422,100	*ESM1*, *GZMK*, *GZMA*, *CDC20B*, *GPX8*, *MCIDAS*, *CCNO*, *DHX29*, *SKIV2L2*, *PPAP2A*, *SLC38A9*, *DDX4*, *IL31RA*, *IL6ST*, *ANKRD55*
	Chr3:50,184,000–53,602,300	*SEMA3F*, *GNAT1*, *GNAI2*, *LSMEM2*, *IFRD2*, *HYAL3*, *NAT6*, *HYAL1*, *HYAL2*, *TUSC2*, *RASSF1*, *ZMYND10*, *NPRL2*, *CYB561D2*, *TMEM115*, *CACNA2D2*, *C3orf18*, *HEMK1*, *CISH*, *MAPKAPK3*, *DOCK3*, *MANF*, *RBM15B*, *RAD54L2*, *TEX264*, *GRM2*, *IQCF6*, *IQCF3*, *IQCF2*, *IQCF5*, *IQCF1*, *RRP9*, *PARP3*, *GPR62*, *PCBP4*, *ABHD14B*, *ABHD14A*, *ACY1*, *RPL29*, *DUSP7*, *POC1A*, *ALAS1*, *TLR9*, *TWF2*, *PPM1M*, *WDR82*, *GLYCTK*, *DNAH1*, *BAP1*, *PHF7*, *SEMA3G*, *TNNC1*, *NISCH*, *STAB1*, *NT5DC2*, *SMIM4*, *PBRM1*, *GNL3*, *GLT8D1*, *SPCS1*, *NEK4*, *ITIH1*, *ITIH3*, *ITIH4*, *MUSTN1*, *TMEM110-MUSTN1*, *TMEM110*, *SFMBT1*, *RFT1*, *PRKCD*, *TKT*, *CACNA1D*
	Chr13:96,038,900–97,500,100	*CLDN10*, *DZIP1*, *DNAJC3*, *UGGT2*, *HS6ST3*
	Chr18:14,517,500–19,962,400	*POTEC*, *ANKRD30B*, *ROCK1*, *GREB1L*, *ESCO1*, *SNRPD1*, *ABHD3*, *MIB1*, *GATA6*

All positions were rounded to the nearest 100 bp. Windows were merged together if the central SNPs that define them were contiguous.

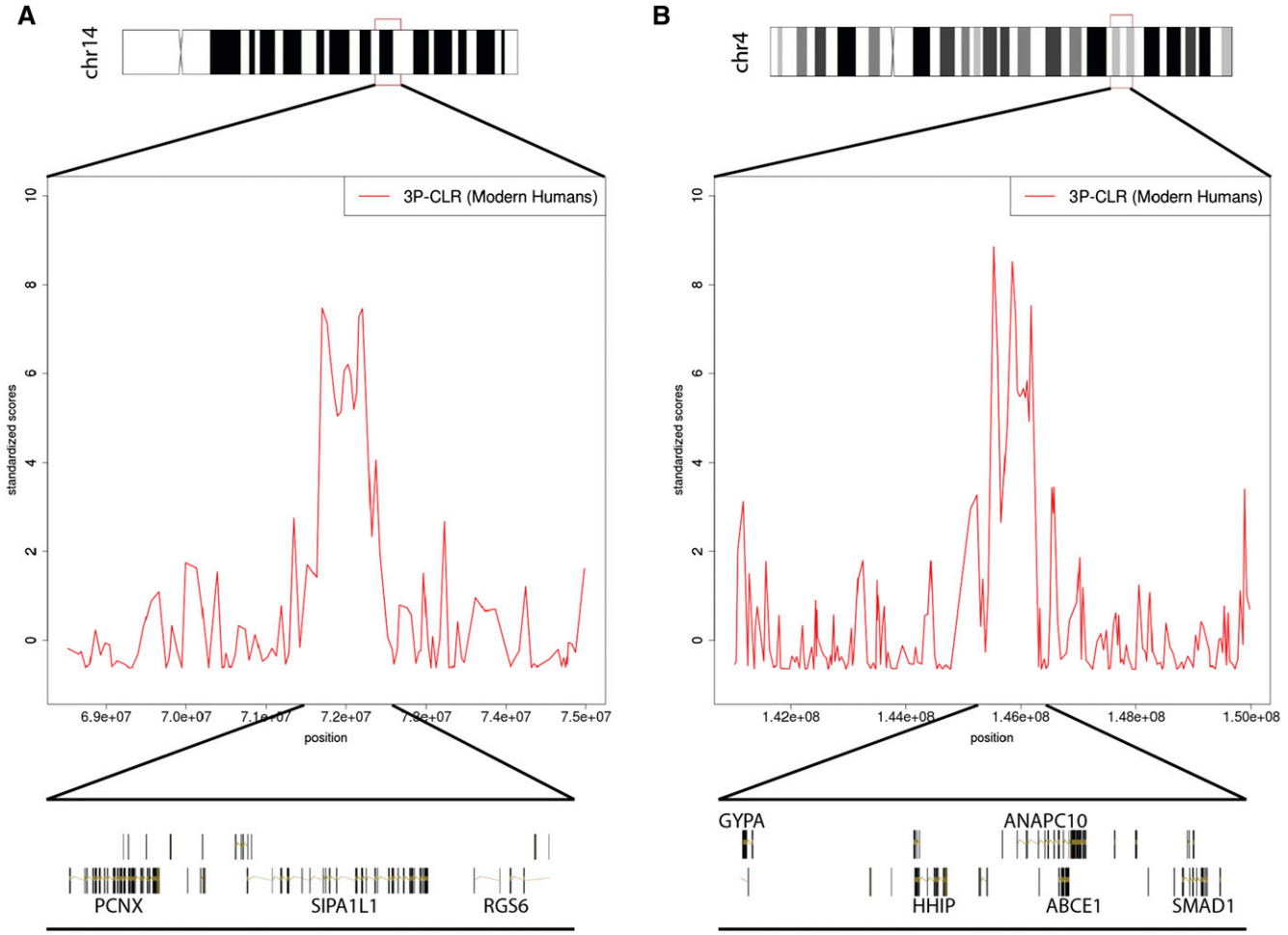

Figure 5 Two of the strongest candidates for selection in the modern human lineage, after the split from Neanderthal and Denisova. We show scores from the 1-cM scan, but the signals persist in the 0.25-cM scan. To make a fair comparison, all 3P-CLR scores were standardized by substracting the chromosome-wide mean from each window and dividing the resulting score by the chromosome-wide standard deviation. (A) The region containing *SIPA1L1*. (B) The region containing *ANAPC10*. The image was built using the GenomeGraphs package in Bioconductor.

(PhastCons = 1). The change is 18 bp away from the nearest splice site and overlaps a VISTA conserved enhancer region (element 1874) (Pennacchio *et al.* 2006), suggesting a putative regulatory role for the change.

We again used Gowinda (Kofler and Schlötterer 2012) to find enriched GO categories among the regions with high 3P-CLR scores in the modern human branch. The significantly enriched categories ($P < 0.05$, FDR < 0.3) are listed in Table S4. We find several GO terms related to the regulation of the cell cycle, T-cell migration, and intracellular transport.

We overlapped the genome-wide association studies (GWAS) database (Li *et al.* 2011; Welter *et al.* 2014) with the list of fixed or high-frequency modern human-specific changes that are ancestral in archaic humans (Prüfer *et al.* 2014) and that are located within our top putatively selected regions in modern humans (see Table S7 and Table S8 for the 0.25- and 1-cM scans, respectively). None of the resulting SNPs are completely fixed derived, because GWAS can yield associations only from sites that are segregating. We find several SNPs in the *RAB28* gene (Rosenbloom *et al.* 2011; Dunham *et al.* 2012), which are significantly associated with obesity (Paternoster *et al.* 2011). We also find a SNP with a high *C* score (rs10171434) associated with urinary metabolites (Suhre *et al.* 2011) and suicidal behavior in patients with mood disorders (Perlis *et al.* 2010). The SNP is located in an enhancer regulatory feature (Rosenbloom *et al.* 2011; Dunham *et al.* 2012) located between genes *PELI1* and *VPS54*, in the same putatively selected region as that of genes *EHBP1* and *OTX1* (see above). Finally, there is a highly *C*-scoring SNP (rs731108) that is associated with renal cell carcinoma (Henrion *et al.* 2013). This SNP is also located in an enhancer regulatory feature (Rosenbloom *et al.* 2011; Dunham *et al.* 2012), in an intron of *ZEB2*. In this last case, though, only the Neanderthal genome has the ancestral state, while the Denisova genome carries the modern human variant.

Discussion

We have developed a new method called 3P-CLR, which allows us to detect positive selection along the genome.

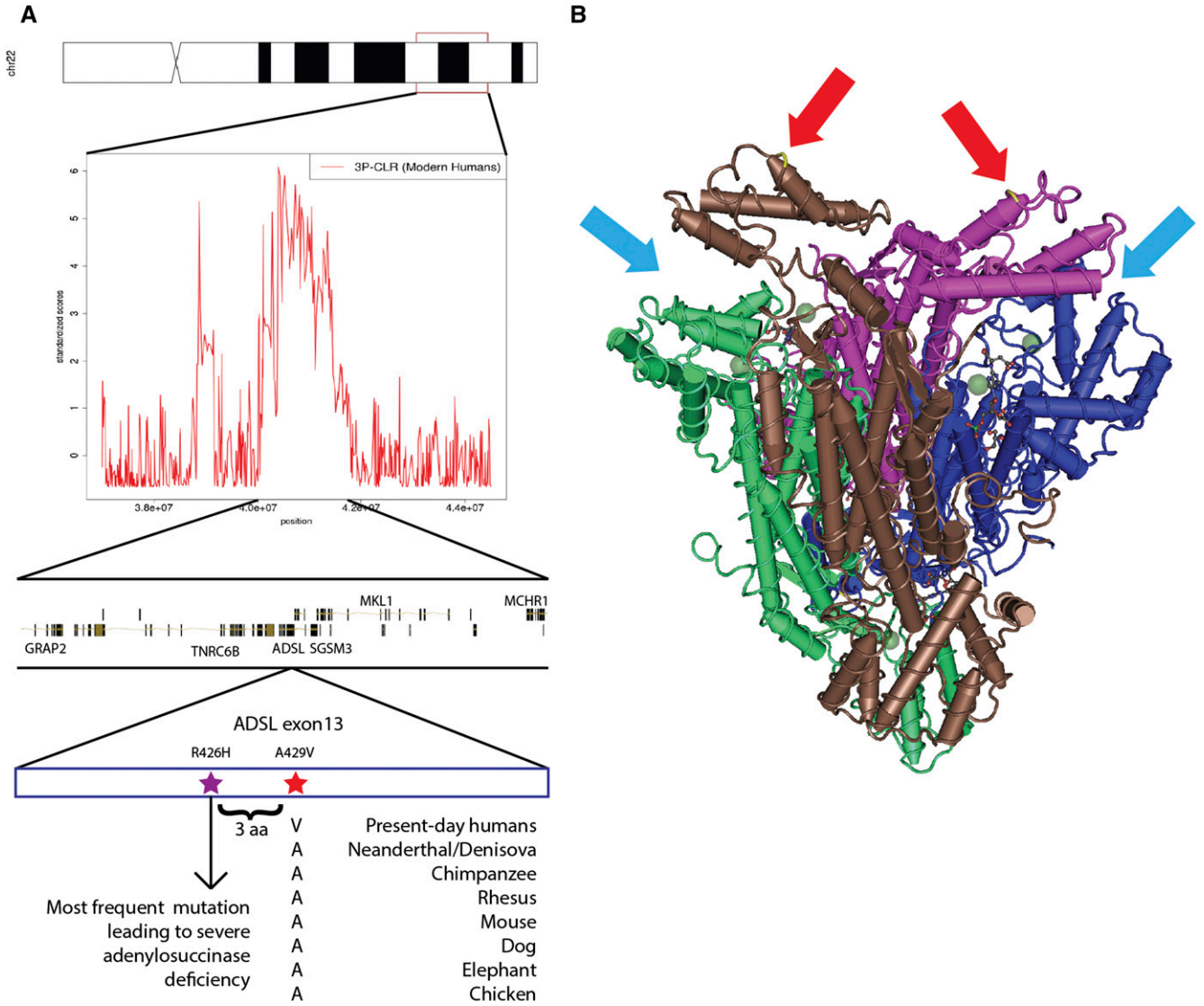

Figure 6 ADSL is a candidate for selection in the modern human lineage, after the split from Neanderthal and Denisova. (A) One of the top-scoring regions when running 3P-CLR (0.25-cM windows) on the modern human lineage contains genes *TNRC6B*, *ADSL*, *MKL1*, *MCHR1*, *SGSM3*, and *GRAP2*. The most disruptive nonsynonymous modern-human-specific change in the entire list of top regions is in an exon of *ADSL* and is fixed derived in all present-day humans but ancestral in archaic humans. It is highly conserved across tetrapods and lies only three residues away from the most common mutation leading to severe adenylosuccinase deficiency. (B) The *ADSL* gene codes for a tetrameric protein. The mutation is in the C-terminal domain of each of the tetrameric units (red arrows), which are near the active sites (light blue arrows). Scores in A were standardized using the chromosome-wide mean and standard deviation. Vertebrate alignments were obtained from the UCSC genome browser (Vertebrate Multiz Alignment and Conservation track) and the image was built using the GenomeGraphs package in Bioconductor and Cn3D.

The method is based on an earlier test [XP-CLR (Chen *et al.* 2010)] that uses linked allele frequency differences between two populations to detect population-specific selection. However, unlike XP-CLR, 3P-CLR can allow us to distinguish between selective events that occurred before and after the split of two populations. Our method has some similarities to an earlier method developed by Schlebusch *et al.* (2012), which used an F_{st}-like score to detect selection ancestral to two populations. In that case, though, the authors used summary statistics and did not explicitly model the process leading to allele frequency differentiation. It is also similar to a more recent method (Fariello *et al.* 2013) that models differences in haplotype frequencies between populations, while accounting for population structure.

We used our method to confirm previously found candidate genes in particular human populations, like *EDAR*, *TYRP1*, and *CYP26B1*, and find some novel candidates too (Table S1, Table S2, and Table S3). Additionally, we can infer that certain genes, which were previously known to have been under selection in East Asians (like *SPAG6*), are more likely to have undergone a sweep in the population ancestral to both Europeans and East Asians than in East Asians only. We find that genes involved in epidermis development and alcohol catabolism are particularly enriched among the East

Asian candidate regions, while genes involved in peroxide catabolism and cuticle development are enriched in the European branch. This suggests these biological functions may have been subject to positive selection in recent times.

We also used 3P-CLR to detect selective events that occurred in the ancestors of modern humans, after their split from Neanderthals and Denisovans (Table S5). These events could perhaps have led to the spread of phenotypes that set modern humans apart from other hominin groups. We find several interesting candidates, like *SIPA1L1*, *ADSL*, *RASA1*, *OTX1*, *EHBP1*, *FOXG1*, *RAB28*, and *ANAPC10*, some of which were previously detected using other types of methods (Green *et al.* 2010; Prüfer *et al.* 2014; Racimo *et al.* 2014). We also find an enrichment for GO categories related to cell cycle regulation and T-cell migration among the candidate regions, suggesting that these biological processes might have been affected by positive selection after the split from archaic humans.

An advantage of differentiation-based tests like XP-CLR and 3P-CLR is that, unlike other patterns detected by tests of neutrality [like extended haplotype homozygostiy (Sabeti *et al.* 2002)] that are exclusive to hard sweeps, the patterns that both XP-CLR and 3P-CLR are tailored to find are based on regional allele frequency differences between populations. These patterns can also be produced by soft sweeps from standing variation or by partial sweeps (Chen *et al.* 2010), and there is some evidence that the latter phenomena may have been more important than classic sweeps during human evolutionary history (Hernandez *et al.* 2011).

Another advantage of both XP-CLR and 3P-CLR is that they do not rely on an arbitrary division of genomic space. Unlike other methods that require the partition of the genome into small windows of fixed size, our composite-likelihood ratios can theoretically be computed over windows that are as big as each chromosome, while switching only the central candidate site at each window. This is because the likelihood ratios use the genetic distance to the central SNP as input. SNPs that are very far away from the central SNP will not contribute much to the likelihood function of both the neutral and the selection models, while those that are close to it will. In the interest of speed, we heuristically limit the window size in our implementation and use fewer SNPs when calculating likelihoods over larger windows. Nevertheless, these parameters can be arbitrarily adjusted by the user as needed and if enough computing resources are available. The use of genetic distance in the likelihood function also allows us to take advantage of the spatial distribution of SNPs as an additional source of information, rather than only relying on patterns of population differentiation restricted to tightly linked SNPs.

3P-CLR also has an advantage over HMM-based selection methods, like the one implemented in Prüfer *et al.* (2014). The likelihood-ratio scores obtained from 3P-CLR can provide an idea of how credible a selection model is for a particular region, relative to the rest of the genome. The HMM-based method previously used to scan for selection in modern humans (Prüfer *et al.* 2014) can rank putatively selected regions only by genetic distance, but cannot output a statistical measure that may indicate how likely each region is to have been under selection in ancient times. In contrast, 3P-CLR provides a composite-likelihood-ratio score, which allows for a statistically rigorous way to compare the neutral model and a specific selection model (for example, recent or ancient selection).

The outliers from Figure S10 have much higher scores (relative to the rest of the genome) than the outliers from Figure S13. This may be due to both the difference in timescales in the two sets of tests and the uncertainty that comes from estimating outgroup allele frequencies using only two archaic genomes. This pattern can also be observed in Figure S14, where the densities of the scores looking for patterns of ancient selection (3P-CLR modern human and 3P-CLR Eurasia) have much shorter tails than the densities of scores looking for patterns of recent selection (3P-CLR Europe and 3P-CLR East Asia). Simulations show that 3P-CLR(Int) score distributions are naturally shorter than 3P-CLR(A) scores (Figure S15), which could explain the short tail of the 3P-CLR Eurasia distribution. Additionally, the even shorter tail in the distribution of 3P-CLR modern human scores may be a consequence of the fact that the split times of the demographic history in that case are older than the split times in the Eurasian tree, as simulations show that ancient split times tend to further shorten the tail of the 3P-CLR score distribution (Figure S15). We note, though, that using a larger window size produces a larger number of strong outliers (Figure S12).

A limitation of composite-likelihood-ratio tests is that the composite likelihood calculated for each model under comparison is obtained from a product of individual likelihoods at each site, and so it underestimates the correlation that exists between SNPs due to linkage effects (Lindsay 1988; Chen *et al.* 2010; Pace *et al.* 2011; Varin *et al.* 2011). One way to partially mitigate this problem is by using corrective weights based on linkage disequilibrium (LD) statistics calculated on the outgroup population (Chen *et al.* 2010). Our implementation of 3P-CLR allows the user to incorporate such weights, if appropriate LD statistics are available from the outgroup. However, in cases where these are unreliable, it may not be possible to correct for this (for example, when only a few unphased genomes are available, as in the case of the Neanderthal and Denisova genomes).

While 3P-CLR relies on integrating over the possible allele frequencies in the ancestors of populations a and b (Equation 10), one could envision using ancient DNA to avoid this step. Thus, if enough genomes could be sampled from that ancestral population that existed in the past, one could use the sample frequency in the ancient set of genomes as a proxy for the ancestral population frequency. This may soon be possible, as several early modern human genomes have already been sequenced in recent years (Fu *et al.* 2014; Lazaridis *et al.* 2014; Seguin-Orlando *et al.* 2014).

Although we have focused on a three-population model in this article, it should be straightforward to expand our method

to a larger number of populations, albeit with additional costs in terms of speed and memory. 3P-CLR relies on a similar framework to that of the demographic inference method implemented in TreeMix (Pickrell and Pritchard 2012), which can estimate population trees that include migration events, using genome-wide data. With a more complex modeling framework, it may be possible to estimate the time and strength of selective events with better resolution and using more populations and also to incorporate additional demographic forces, like continuous migration between populations or pulses of admixture.

Acknowledgments

We thank Montgomery Slatkin, Rasmus Nielsen, Joshua Schraiber, Nicolas Duforet-Frebourg, Emilia Huerta-Sánchez, Hua Chen, Benjamin Peter, Nick Patterson, David Reich, Joachim Hermisson, Graham Coop, and members of the Slatkin and Nielsen laboratories for helpful advice and discussions. We also thank two anonymous reviewers for their helpful comments. This work was supported by National Institutes of Health grant R01-GM40282 to Montgomery Slatkin.

Literature Cited

Abecasis, G. R., A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin *et al.*, 2012 An integrated map of genetic variation from 1,092 human genomes. Nature 491(7422): 56–65.

Akey, J. M., G. Zhang, K. Zhang, L. Jin, and M. D. Shriver, 2002 Interrogating a high-density SNP map for signatures of natural selection. Genome Res. 12(12): 1805–1814.

Ariani, F., G. Hayek, D. Rondinella, R. Artuso, M. A. Mencarelli *et al.*, 2008 Foxg1 is responsible for the congenital variant of rett syndrome. Am. J. Hum. Genet. 83(1): 89–93.

Branicki, W., U. Brudnik, and A. Wojas-Pelc, 2009 Interactions between herc2, oca2 and mc1r may influence human pigmentation phenotype. Ann. Hum. Genet. 73(2): 160–170.

Brawand, D., M. Soumillon, A. Necsulea, P. Julien, G. Csárdi *et al.*, 2011 The evolution of gene expression levels in mammalian organs. Nature 478(7369): 343–348.

Castellano, S., G. Parra, F. A. Sánchez-Quinto, F. Racimo, M. Kuhlwilm *et al.*, 2014 Patterns of coding variation in the complete exomes of three Neandertals. Proc. Natl. Acad. Sci. USA 111(18): 6666–6671.

Chen, H., N. Patterson, and D. Reich, 2010 Population differentiation as a test for selective sweeps. Genome Res. 20(3): 393–402.

Cooper, G. M., D. L. Goode, S. B. Ng, A. Sidow, M. J. Bamshad *et al.*, 2010 Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. Nat. Methods 7(4): 250–251.

Crisci, J. L., A. Wong, J. M. Good, and J. D. Jensen, 2011 On characterizing adaptive events unique to modern humans. Genome Biol. Evol. 3: 791–798.

Du, K. L., M. Chen, J. Li, J. J. Lepore, P. Mericko *et al.*, 2004 Megakaryoblastic leukemia factor-1 transduces cytoskeletal signals and induces smooth muscle cell differentiation from undifferentiated embryonic stem cells. J. Biol. Chem. 279(17): 17578–17586.

Dunham, I., A. Kundaje, S. F. Aldred, P. J. Collins, C. Davis *et al.*, 2012 An integrated encyclopedia of DNA elements in the human genome. Nature 489(7414): 57–74.

Durrett, R., and J. Schweinsberg, 2004 Approximating selective sweeps. Theor. Popul. Biol. 66(2): 129–138.

Edenberg, H. J., D. M. Dick, X. Xuei, H. Tian, L. Almasy *et al.*, 2004 Variations in gabra2, encoding the α2 subunit of the gaba a receptor, are associated with alcohol dependence and with brain oscillations. Am. J. Hum. Genet. 74(4): 705–714.

Edery, P., S. Chabrier, I. Ceballos-Picot, S. Marie, M.-F. Vincent *et al.*, 2003 Intrafamilial variability in the phenotypic expression of adenylosuccinate lyase deficiency: a report on three patients. Am. J. Med. Genet. A. 120(2): 185–190.

Eerola, I., L. M. Boon, J. B. Mulliken, P. E. Burrows, A. Dompmartin *et al.*, 2003 Capillary malformation–arteriovenous malformation, a new clinical and genetic disorder caused by rasa1 mutations. Am. J. Hum. Genet. 73(6): 1240–1249.

Eiberg, H., J. Troelsen, M. Nielsen, A. Mikkelsen, J. Mengel-From *et al.*, 2008 Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the herc2 gene inhibiting oca2 expression. Hum. Genet. 123(2): 177–187.

Ewens, W. J., 2012 *Mathematical Population Genetics 1: Theoretical Introduction*, Vol. 27. Springer Science & Business Media, New York, NY.

Fariello, M. I., S. Boitard, H. Naya, M. SanCristobal, and B. Servin, 2013 Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics 193: 929–941.

Fay, J. C., and C.-I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics 155: 1405–1413.

Felsenstein, J., 1981 Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. Evolution 35: 1229–1242.

Friedman, E., P. V. Gejman, G. A. Martin, and F. McCormick, 1993 Nonsense mutations in the c–terminal sh2 region of the gtpase activating protein (gap) gene in human tumours. Nat. Genet. 5(3): 242–247.

Fu, Q., H. Li, P. Moorjani, F. Jay, S. M. Slepchenko *et al.*, 2014 Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 514(7523): 445–449.

Fujimoto, A., R. Kimura, J. Ohashi, K. Omi, R. Yuliwulandari *et al.*, 2008 A scan for genetic determinants of human hair morphology: Edar is associated with Asian hair thickness. Hum. Mol. Genet. 17(6): 835–843.

Gao, Q., S. Srinivasan, S. N. Boyer, D. E. Wazer, and V. Band, 1999 The e6 oncoproteins of high-risk papillomaviruses bind to a novel putative gap protein, e6tp1, and target it for degradation. Mol. Cell. Biol. 19(1): 733–744.

Gitiaux, C., I. Ceballos-Picot, S. Marie, V. Valayannopoulos, M. Rio *et al.*, 2009 Misleading behavioural phenotype with adenylosuccinate lyase deficiency. Eur. J. Hum. Genet. 17(1): 133–136.

Gong, S., C. Zheng, M. L. Doughty, K. Losos, N. Didkovsky *et al.*, 2003 A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature 425(6961): 917–925.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel *et al.*, 2010 A draft sequence of the Neandertal genome. Science 328 (5979): 710–722.

Grossman, S. R., I. Shylakhter, E. K. Karlsson, E. H. Byrne, S. Morales *et al.*, 2010 A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327(5967): 883–886.

Gudmundsson, J., P. Sulem, T. Rafnar, J. T. Bergthorsson, and A. Manolescu *et al.*, 2008 Common sequence variants on 2p15 and xp11. 22 confer susceptibility to prostate cancer. Nat. Genet. 40(3): 281–283.

Guilherme, A., N. A. Soriano, P. S. Furcinitti, and M. P. Czech, 2004 Role of ehd1 and ehbp1 in perinuclear sorting and insulin-regulated glut4 recycling in 3t3-l1 adipocytes. J. Biol. Chem. 279(38): 40062–40075.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante, 2009 Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5(10): e1000695.
Halaban, R., and G. Moellmann, 1990 Murine and human b locus pigmentation genes encode a glycoprotein (gp75) with catalase activity. Proc. Natl. Acad. Sci. USA 87(12): 4809–4813.
Han, J., P. Kraft, H. Nan, Q. Guo, and C. Chen *et al.*, 2008 A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet. 4(5): e1000074.
Henrion, M., M. Frampton, G. Scelo, M. Purdue, Y. Ye *et al.*, 2013 Common variation at 2q22. 3 (zeb2) influences the risk of renal cancer. Hum. Mol. Genet. 22(4): 825–831.
Hernandez, R. D., J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton *et al.*, 2011 Classic selective sweeps were rare in recent human evolution. Science 331(6019): 920–924.
Hershkovitz, D., D. Bercovich, E. Sprecher, and M. Lapidot, 2008 Rasa1 mutations may cause hereditary capillary malformations without arteriovenous malformations. Br. J. Dermatol. 158(5): 1035–1040.
Hinch, A. G., A. Tandon, N. Patterson, Y. Song, N. Rohland *et al.*, 2011 The landscape of recombination in African Americans. Nature 476(7359): 170–175.
Hunt, K. A., A. Zhernakova, G. Turner, G. A. R. Heap, L. Franke *et al.*, 2008 Newly identified genetic risk variants for celiac disease related to the immune response. Nat. Genet. 40(4): 395–402.
Kanehisa, M., and S. Goto, 2000 Kegg: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28(1): 27–30.
Kenny, E. E., N. J. Timpson, M. Sikora, M.-C. Yee, A. Moreno-Estrada *et al.*, 2012 Melanesian blond hair is caused by an amino acid change in tyrp1. Science 336(6081): 554.
Kimura, R., T. Yamaguchi, M. Takeda, O. Kondo, T. Toma *et al.*, 2009 A common variation in edar is a genetic determinant of shovel-shaped incisors. Am. J. Hum. Genet. 85(4): 528–535.
Kircher, M., D. M. Witten, P. Jain, B. J. O'Roak, G. M. Cooper *et al.*, 2014 A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. 46(3): 310–315.
Kmoch, S., H. Hartmannová, B. Stibùrková, J. Krijt, M. Zikánová *et al.*, 2000 Human adenylosuccinate lyase (adsl), cloning and characterization of full-length cDNA and its isoform, gene structure and molecular basis for adsl deficiency in six patients. Hum. Mol. Genet. 9(10): 1501–1513.
Knabl, J., R. Witschi, K. Hösl, H. Reinold, U. B. Zeilhofer *et al.*, 2008 Reversal of pathological pain through specific spinal gabaa receptor subtypes. Nature 451(7176): 330–334.
Kofler, R., and C. Schlötterer, 2012 Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. Bioinformatics 28(15): 2084–2085.
Lazaridis, I., N. Patterson, A. Mittnik, G. Renaud, S. Mallick *et al.*, 2014 Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513(7518): 409–413.
Lewontin, R. C., and J. Krakauer, 1973 Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74: 175–195.
Li, H., and R. Durbin, 2011 Inference of human population history from individual whole-genome sequences. Nature 475(7357): 493–496.
Li, M. J., P. Wang, X. Liu, E. L. Lim, Z. Wang *et al.*, 2011 GWASdb: a database for human genetic variants identified by genome-wide association studies. Nucleic Acids Res. 40: D1047–D1054.
Lindsay, B. G., 1988 Composite likelihood methods. Contemp. Math. 80(1): 221–239.
Lipson, M., P.-R. Loh, A. Levin, D. Reich, N. Patterson *et al.*, 2013 Efficient moment-based inference of admixture parameters and sources of gene flow. Mol. Biol. Evol. 30(8): 1788–1802.
Maaswinkel-Mooij, P. D., L. A. E. M. Laan, W. Onkenhout, O. F. Brouwer, J. Jaeken *et al.*, 1997 Adenylosuccinase deficiency presenting with epilepsy in early infancy. J. Inherit. Metab. Dis. 20(4): 606–607.
Marie, S., H. Cuppens, M. Heuterspreute, M. Jaspers, E. Z. Tola *et al.*, 1999 Mutation analysis in adenylosuccinate lyase deficiency: eight novel mutations in the re-evaluated full adsl coding sequence. Hum. Mutat. 13(3): 197–202.
Mathieson, I., I. Lazaridis, N. Rohland, S. Mallick, B. Llamas *et al.*, 2015 Eight thousand years of natural selection in Europe. bioRxiv: 016477.
Meister, G., M. Landthaler, L. Peters, P. Y. Chen, H. Urlaub *et al.*, 2005 Identification of novel argonaute-associated proteins. Curr. Biol. 15(23): 2149–2155.
Mencarelli, M. A., A. Spanhol-Rosseto, R. Artuso, D. Rondinella, R. De Filippis *et al.*, 2010 Novel foxg1 mutations associated with the congenital variant of rett syndrome. J. Med. Genet. 47(1): 49–53.
Mercher, T., M. Busson-Le Coniat, R. Monni, M. Mauchauffé, F. N. Khac *et al.*, 2001 Involvement of a human gene related to the Drosophila spen gene in the recurrent t (1; 22) translocation of acute megakaryocytic leukemia. Proc. Natl. Acad. Sci. USA 98 (10): 5776–5779.
Messer, P. W., 2013 Slim: simulating evolution with selection and linkage. Genetics 194: 1037–1039.
Meyer, M., M. Kircher, M.-T. Gansauge, H. Li, and F. Racimo *et al.*, 2012 A high-coverage genome sequence from an archaic Denisovan individual. Science 338(6104): 222–226.
Nicholson, G., A. V. Smith, F. Jónsson, Ó. Gústafsson, K. Stefánsson *et al.*, 2002 Assessing population differentiation and isolation from single-nucleotide polymorphism data. J. R. Stat. Soc. Ser. B Stat. Methodol. 64(4): 695–715.
Oleksyk, T. K., K. Zhao, M. Francisco, D. A. Gilbert, S. J. O'Brien *et al.*, 2008 Identifying selected regions from heterozygosity and divergence using a light-coverage genomic dataset from two human populations. PLoS ONE 3(3): e1712.
Pace, L., A. Salvan, and N. Sartori, 2011 Adjusting composite likelihood ratio statistics. Stat. Sin. 21(1): 129.
Paternoster, L., D. M. Evans, E. A. Nohr, C. Holst, V. Gaborieau *et al.*, 2011 Genome-wide population-based association study of extremely overweight young adults–the Goya study. PLoS ONE 6(9): e24303.
Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland *et al.*, 2012 Ancient admixture in human history. Genetics 192: 1065–1093.
Pennacchio, L. A., N. Ahituv, A. M. Moses, S. Prabhakar, M. A. Nobrega *et al.*, 2006 In vivo enhancer analysis of human conserved non-coding sequences. Nature 444(7118): 499–502.
Perlis, R. H., J. Huang, S. Purcell, M. Fava, A. J. Rush *et al.*, 2010 Genome-wide association study of suicide attempts in mood disorder patients. Genome 167(12): 1499–1507.
Pickrell, J. K., and J. K. Pritchard, 2012 Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8(11): e1002967.
Pickrell, J. K., G. Coop, J. Novembre, S. Kudaravalli, J. Z. Li *et al.*, 2009 Signals of recent positive selection in a worldwide sample of human populations. Genome Res. 19(5): 826–837.
Pravtcheva, D. D., and T. L. Wise, 2001 Disruption of apc10/doc1 in three alleles of oligosyndactylism. Genomics 72(1): 78–87.
Prüfer, K., F. Racimo, N. Patterson, F. Jay, S. Sankararaman *et al.*, 2014 The complete genome sequence of a Neanderthal from the Altai mountains. Nature 505(7481): 43–49.
Race, V., S. Marie, M.-F. Vincent, and G. Van den Berghe, 2000 Clinical, biochemical and molecular genetic correlations

in adenylosuccinate lyase deficiency. Hum. Mol. Genet. 9(14): 2159–2165.

Racimo, F., M. Kuhlwilm, and M. Slatkin, 2014 A test for ancient selective sweeps and an application to candidate sites in modern humans. Mol. Biol. Evol. 31(12): 3344–3358.

Robinson, P. N., S. Köhler, S. Bauer, D. Seelow, D. Horn *et al.*, 2008 The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. Am. J. Hum. Genet. 83(5): 610–615.

Rosenbloom, K. R., T. R. Dreszer, J. C. Long, V. S. Malladi, C. A. Sloan *et al.*, 2011 ENCODE whole-genome data in the UCSC genome browser: update 2012. Nucleic Acids Res. 40: D912–D917.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter *et al.*, 2002 Detecting recent positive selection in the human genome from haplotype structure. Nature 419(6909): 832–837.

Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, and E. Hostetter *et al.*, 2007 Genome-wide detection and characterization of positive selection in human populations. Nature 449(7164): 913–918.

Sadakata, T., and T. Furuichi, 2010 Ca 2+-dependent activator protein for secretion 2 and autistic-like phenotypes. Neurosci. Res. 67(3): 197–202.

Sapiro, R., I. Kostetskii, P. Olds-Clarke, G. L. Gerton, G. L. Radice *et al.*, 2002 Male infertility, impaired sperm motility, and hydrocephalus in mice deficient in sperm-associated antigen 6. Mol. Cell. Biol. 22(17): 6298–6305.

Schlebusch, C. M., P. Skoglund, P. Sjödin, L. M. Gattepaille, D. Hernandez *et al.*, 2012 Genomic variation in seven khoe-san groups reveals adaptation and complex African history. Science 338(6105): 374–379.

Seguin-Orlando, A., T. S. Korneliussen, M. Sikora, A.-S. Malaspinas, A. Manica *et al.*, 2014 Genomic structure in Europeans dating back at least 36,200 years. Science 346(6213): 1113–1118.

Siddique, H. R., and M. Saleem, 2012 Role of bmi1, a stem cell factor, in cancer recurrence and chemoresistance: preclinical and clinical evidences. Stem Cells 30(3): 372–378.

Siepel, A., G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou *et al.*, 2005 Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 15(8): 1034–1050.

Smith, J. M., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 23(01): 23–35.

Suhre, K., H. Wallaschofski, J. Raffler, N. Friedrich, R. Haring *et al.*, 2011 A genome-wide association study of metabolic traits in human urine. Nat. Genet. 43(6): 565–569.

Todd, J. A., N. M. Walker, J. D. Cooper, D. J. Smyth, K. Downes *et al.*, 2007 Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat. Genet. 39(7): 857–864.

Topletz, A. R., J. E. Thatcher, A. Zelter, J. D. Lutz, and W. L. Su Tay, 2012 Comparison of the function and expression of cyp26a1 and cyp26b1, the two retinoic acid hydroxylases. Biochem. Pharmacol. 83(1): 149–163.

Trahey, M., G. Wong, R. Halenbeck, B. Rubinfeld, G. A. Martin *et al.*, 1988 Molecular cloning of two types of gap complementary DNA from human placenta. Science 242(4886): 1697–1700.

Van Keuren, M. L., I. M. Hart, F.-T. Kao, R. L. Neve, G. A. P. Bruns *et al.*, 1987 A somatic cell hybrid with a single human chromosome 22 corrects the defect in the cho mutant (ade–i) lacking adenylosuccinase activity. Cytogenet. Genome Res. 44(2–3): 142–147.

Varin, C., N. Reid, and D. Firth, 2011 An overview of composite likelihood methods. Stat. Sin. 21(1): 5–42.

Voight, B. F., S. Kudaravalli, X. Wen, and J. K. Pritchard, 2006 A map of recent positive selection in the human genome. PLoS Biol. 4(3): e72.

Weir, B. S., L. R. Cardon, A. D. Anderson, M. D. Nielsen, and W. G. Hill, 2005 Measures of human population structure show heterogeneity among genomic regions. Genome Res. 15(11): 1468–1476.

Welter, D., J. MacArthur, J. Morales, T. Burdett, P. Hall *et al.*, 2014 The NHGRI GWAS catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 42(D1): D1001–D1006.

White, J. A., H. Ramshaw, M. Taimi, W. Stangle, A. Zhang *et al.*, 2000 Identification of the human cytochrome p450, p450rai-2, which is predominantly expressed in the adult cerebellum and is responsible for all-trans-retinoic acid metabolism. Proc. Natl. Acad. Sci. USA 97(12): 6403–6408.

Whiting, P. J., T. P. Bonnert, R. M. McKernan, S. Farrar, B. Le Bourdelles *et al.*, 1999 Molecular and functional diversity of the expanding gaba-a receptor gene family. Ann. N. Y. Acad. Sci. 868(1): 645–653.

Xiang, Y.-Y., S. Wang, M. Liu, J. A. Hirota, J. Li *et al.*, 2007 A gabaergic system in airway epithelium is essential for mucus overproduction in asthma. Nat. Med. 13(7): 862–867.

Yi, X., Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. Ping Cuo *et al.*, 2010 Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329(5987): 75–78.

Zhernakova, A., C. C. Elbers, B. Ferwerda, J. Romanos, G. Trynka *et al.*, 2010 Evolutionary and functional analysis of celiac risk loci reveals sh2b3 as a protective factor against bacterial infection. Am. J. Hum. Genet. 86(6): 970–977.

Communicating editor: N. A. Rosenberg

GENETICS

Supporting Information
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178095/-/DC1

Testing for Ancient Selection Using Cross-population Allele Frequency Differentiation

Fernando Racimo

DOI: 10.1534/genetics.115.178095

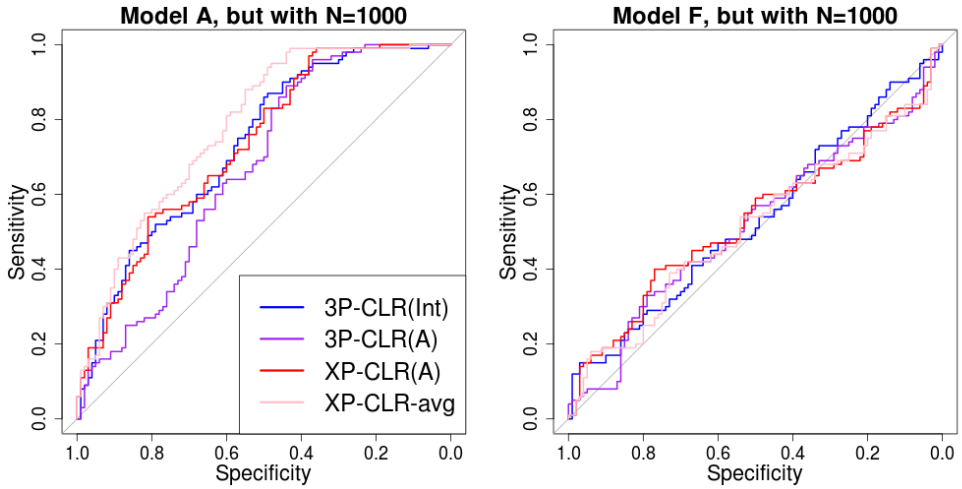

Figure S1. ROC curves for performance of 3P-CLR(Int), 3P-CLR(A) and two variants of XP-CLR in detecting selective sweeps that occurred before the split of two populations a and b, under two demographic models where the population size is extremely small (N_e = 1,000).

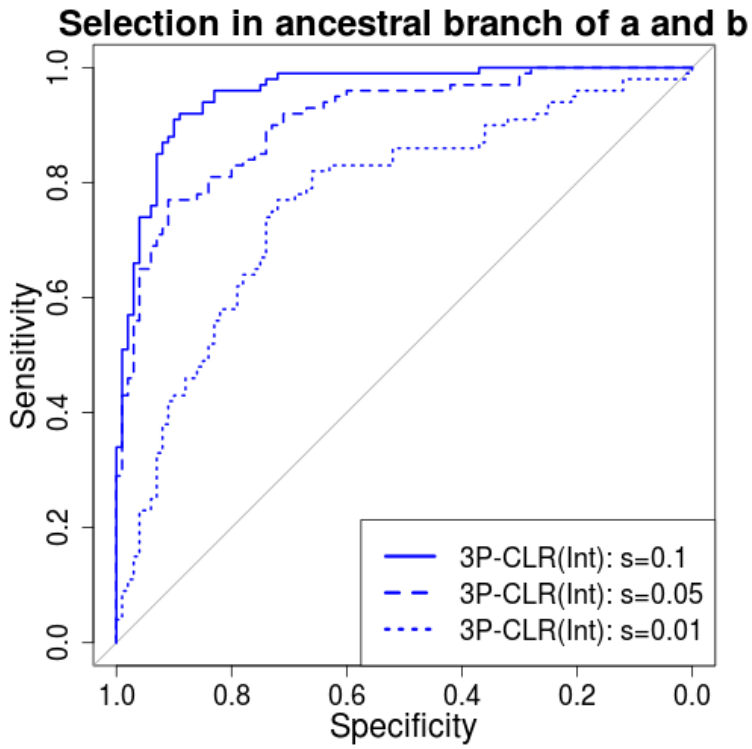

Figure S2. Performance of 3P-CLR(Int) for a range of selection coefficients. We used the demographic history from model B (Table 1) but extended the most ancient split time by 4,000 generations. The reason for this is that we wanted the internal branch to be long enough for it to be easy to sample simulations in which the beneficial allele fixed before the split of populations a and b, even for weak selection coefficients.

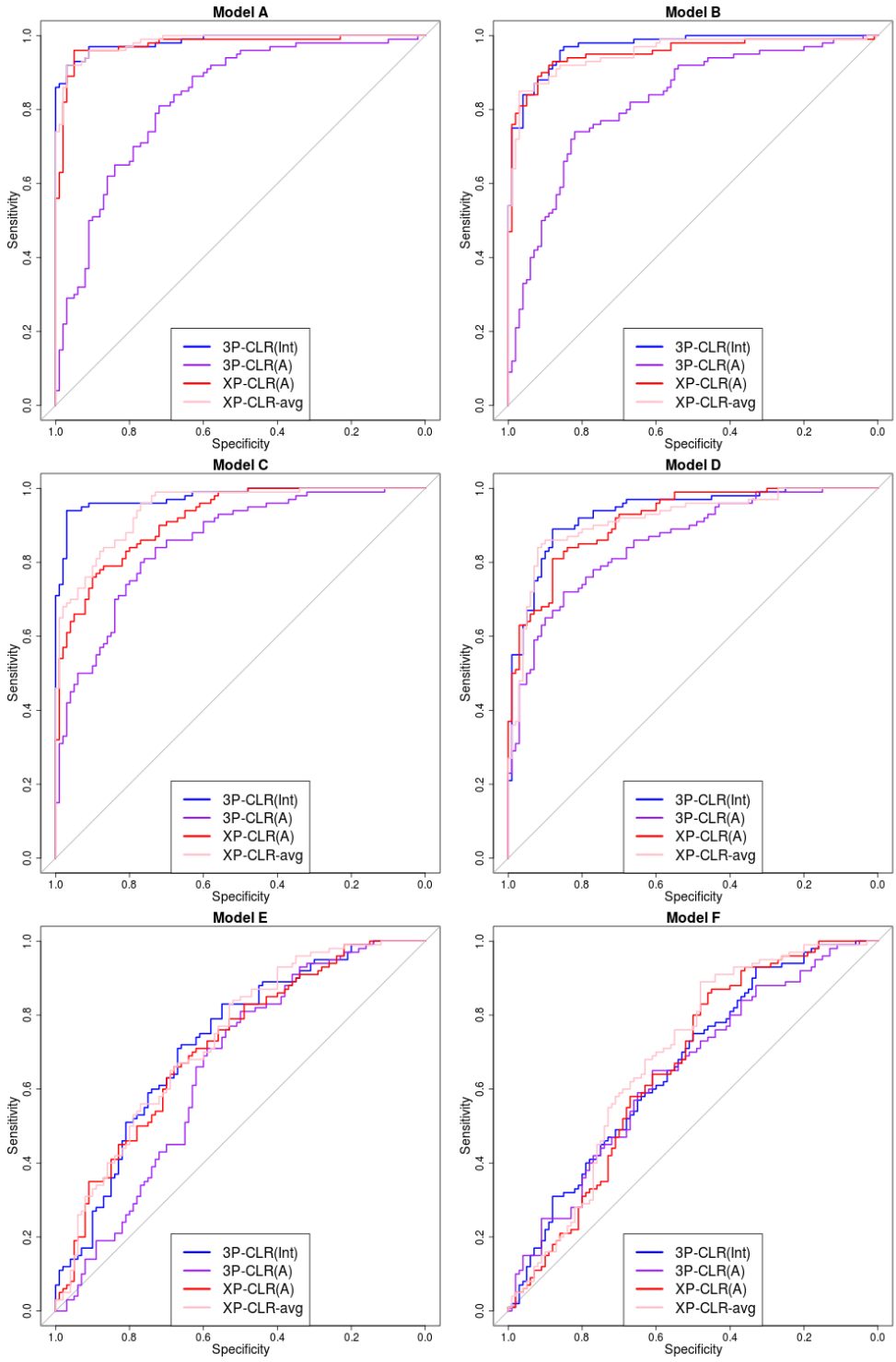

Figure S3. ROC curves for performance of 3P-CLR(Int), 3P-CLR(A) and two variants of XP-CLR in detecting selective sweeps that occurred before the split of two populations a **and** b**, under different demographic models**. In this case, the outgroup panel from population c contained 10 haploid genomes. The two sister population panels (from a and b) have 100 haploid genomes each.

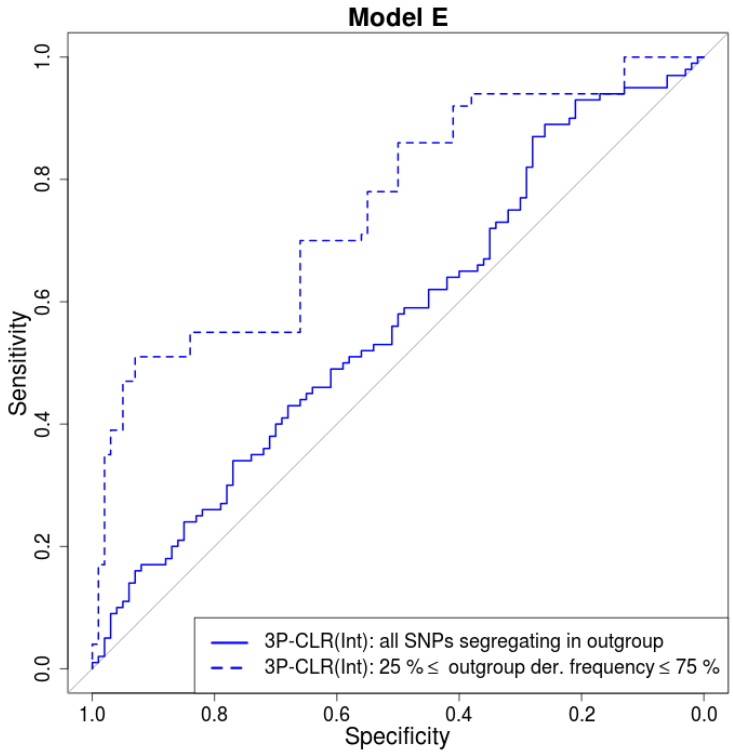

Figure S4. For demographic scenarios with very ancient split times, it is best to use sites segregating at intermediate frequencies in the outgroup. We compared the performance of 3P-CLR(Int) in a demographic scenario with very ancient split times (Model E) under two conditions: including all SNPs that are segregating in the outgroup, and only including SNPs segregating at intermediate frequencies in the outgroup. In both cases, the number of sampled sequences from the outgroup population was 100.

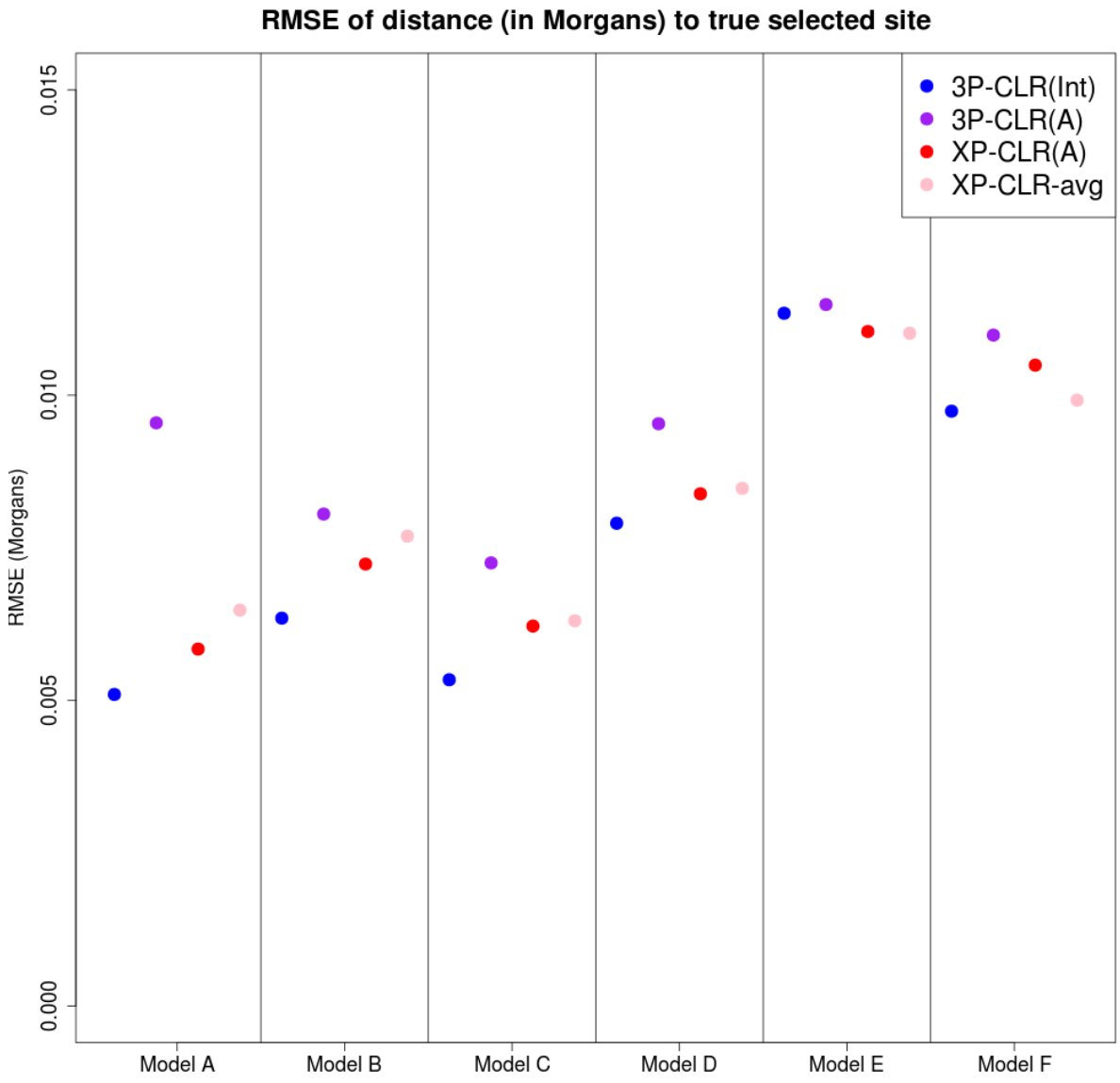

Figure S5. Root-mean squared error for the location of sweeps inferred by 3P-CLR(Int), 3P-CLR(A) and two variants of XP-CLR under different demographic scenarios, when the sweeps occurred before the split of populations a **and** b**.** In this case, the outgroup panel from population c contained 100 haploid genomes and the two sister population panels (from a and b) have 100 haploid genomes each.

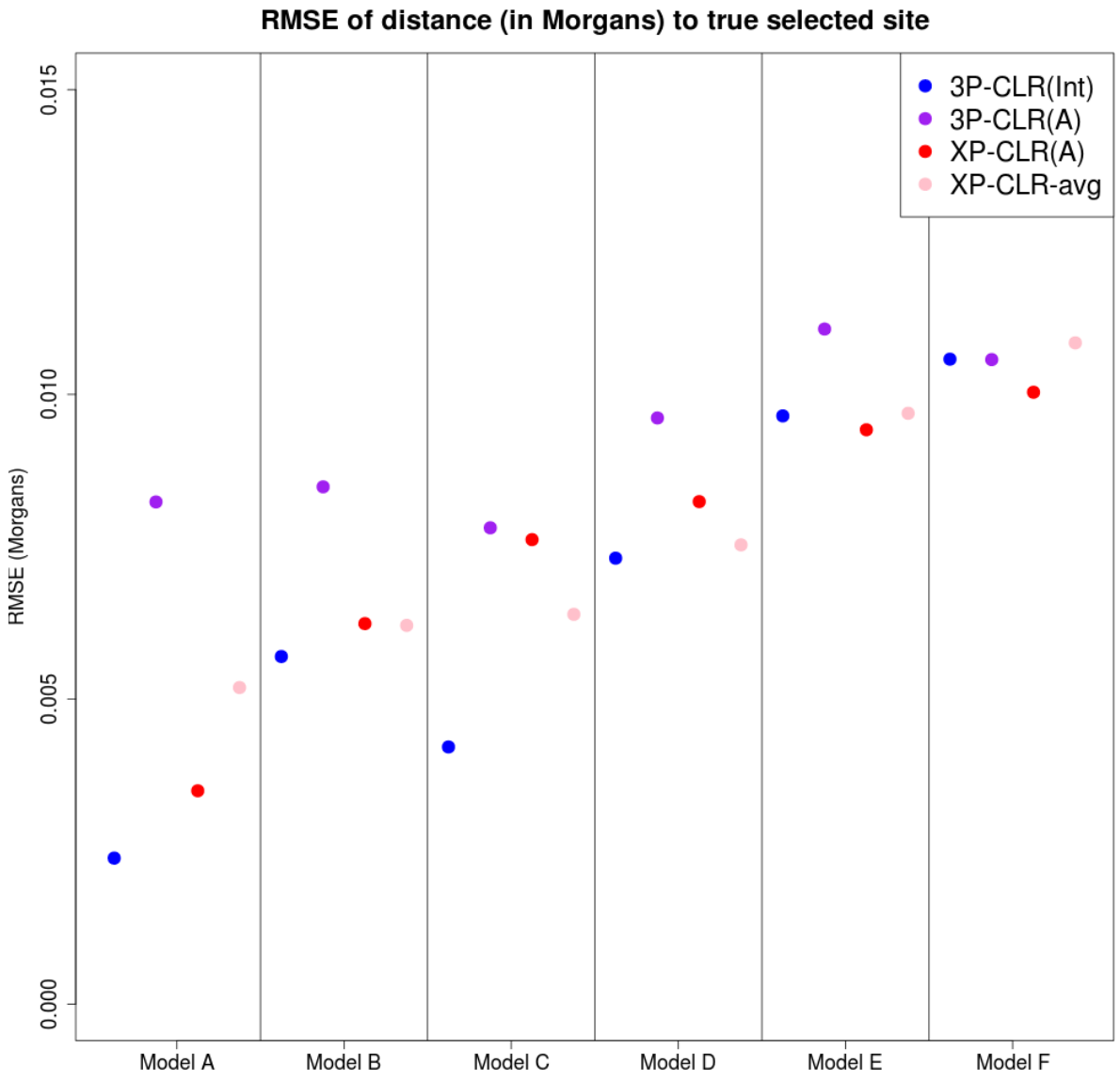

Figure S6. Root-mean squared error for the location of the sweep inferred by 3P-CLR(Int), 3P-CLR(A) and two variants of XP-CLR under different demographic scenarios, when the sweeps occurred before the split of populations a **and** b**.** the outgroup panel from population c contained 10 haploid genomes and the two sister population panels (from a and b) have 100 haploid genomes each.

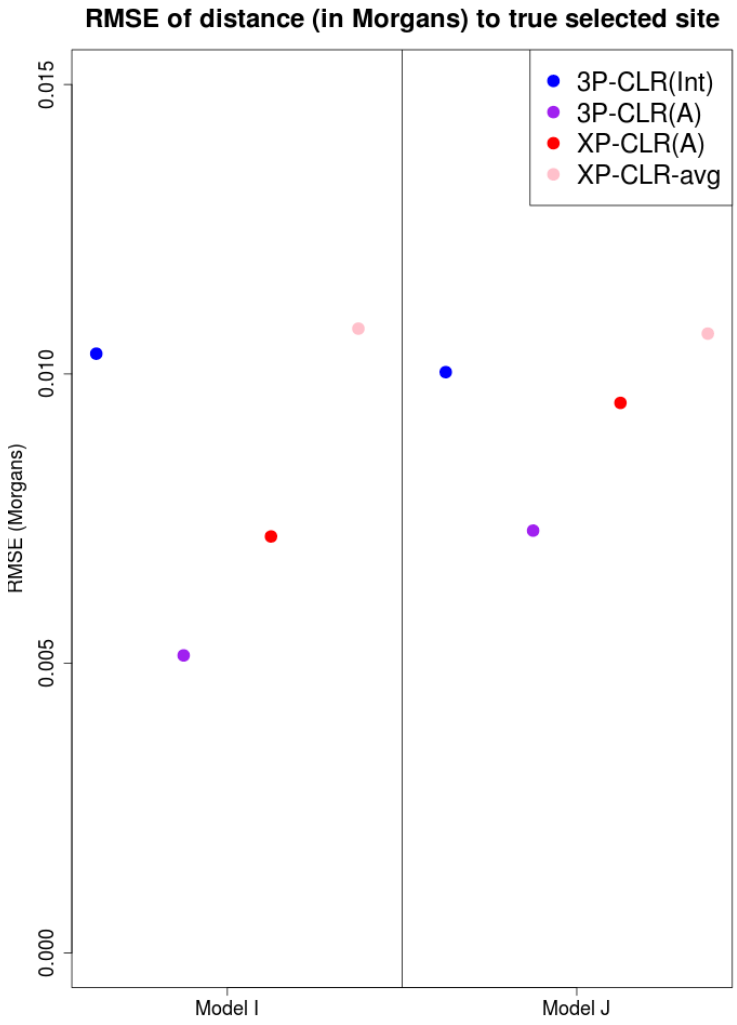

Figure S7. Root-mean squared error for the location of the sweep inferred by 3P-CLR(Int), 3P-CLR(A) and two variants of XP-CLR under different demographic scenarios, when the sweeps occurred in the terminal population branch leading to population a, after the split of populations a and b. In this case, the outgroup panel from population c contained 100 haploid genomes and the two sister population panels (from a and b) have 100 haploid genomes each.

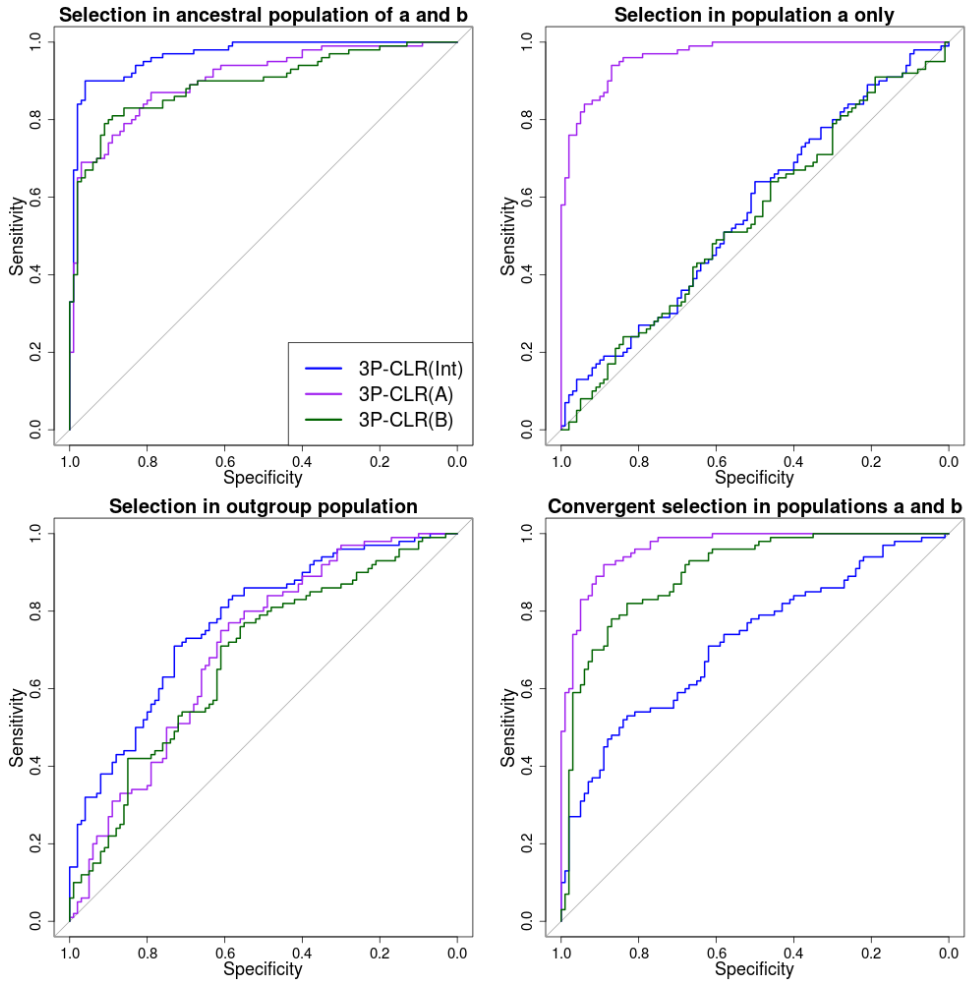

Figure S8. ROC curves for performance of 3P-CLR(Int), 3P-CLR(A) and 3P-CLR(B) when the selective events occur in different branches of the 3-population tree. Upper-left panel: Selection in the ancestral population of populations a and b. This is the type of events that 3P-CLR(Int) is designed to detect and, therefore, 3P-CLR(Int) is the most sensitive test in this case, though 3P-CLR(A) and 3P-CLR(B) show some sensitivity to these events too. Upper-right panel: Selection exclusive to population a. This is the type of events that 3P-CLR(A) is designed to detect, and it is therefore the best-performing statistic in that case, while 3P-CLR(B) and 3P-CLR(Int) are insensitive to selection. Lower-left panel: Selection in the outgroup population. In this case, none of the statistics seem very sensitive to the event, though 3P-CLR(Int) shows better relative sensitivity than the other two statistics. Lower-right panel: Independent selective events in populations a and b at the same locus. Here, both 3P-CLR(A) and 3P-CLR(B) perform best. In all cases, we used the split times and population sizes specified for Model C.

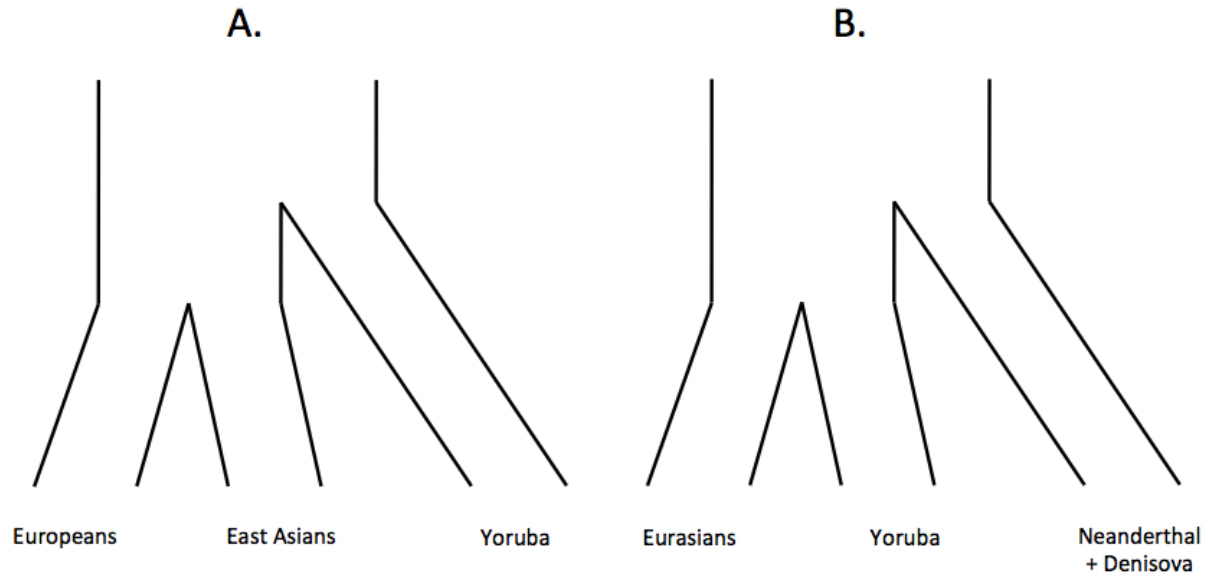

Figure S9. A. Three-population tree separating Europeans, East Asians and Yoruba. B. Three-population tree separating Eurasians, Yoruba and archaic humans (Neanderthal+Denisova).

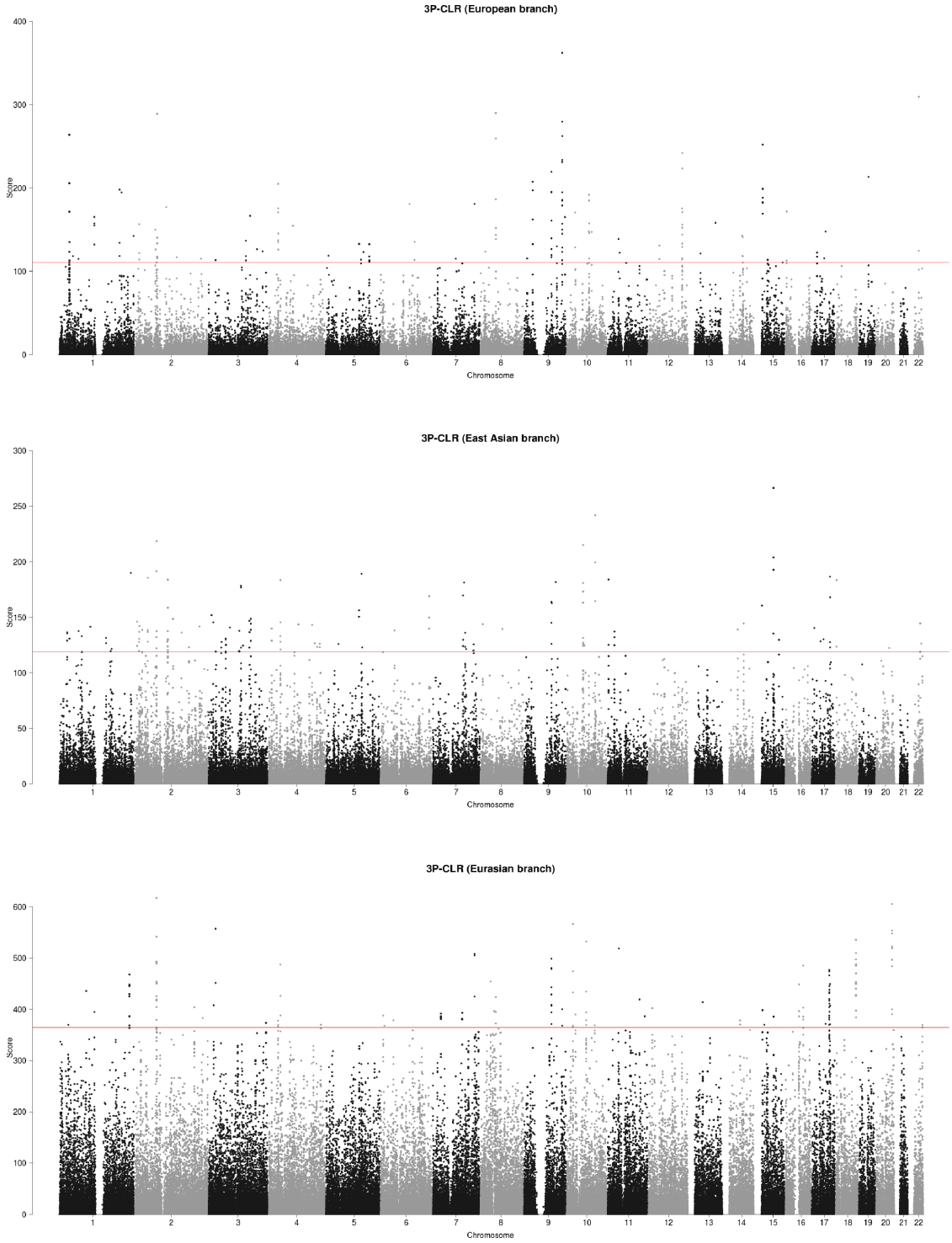

Figure S10. 3P-CLR scan of Europeans (upper panel), East Asians (middle panel) and the ancestral population to Europeans and East Asians (lower panel), using Yoruba as the outgroup in all 3 cases. The red line denotes the 99.9% quantile cutoff.

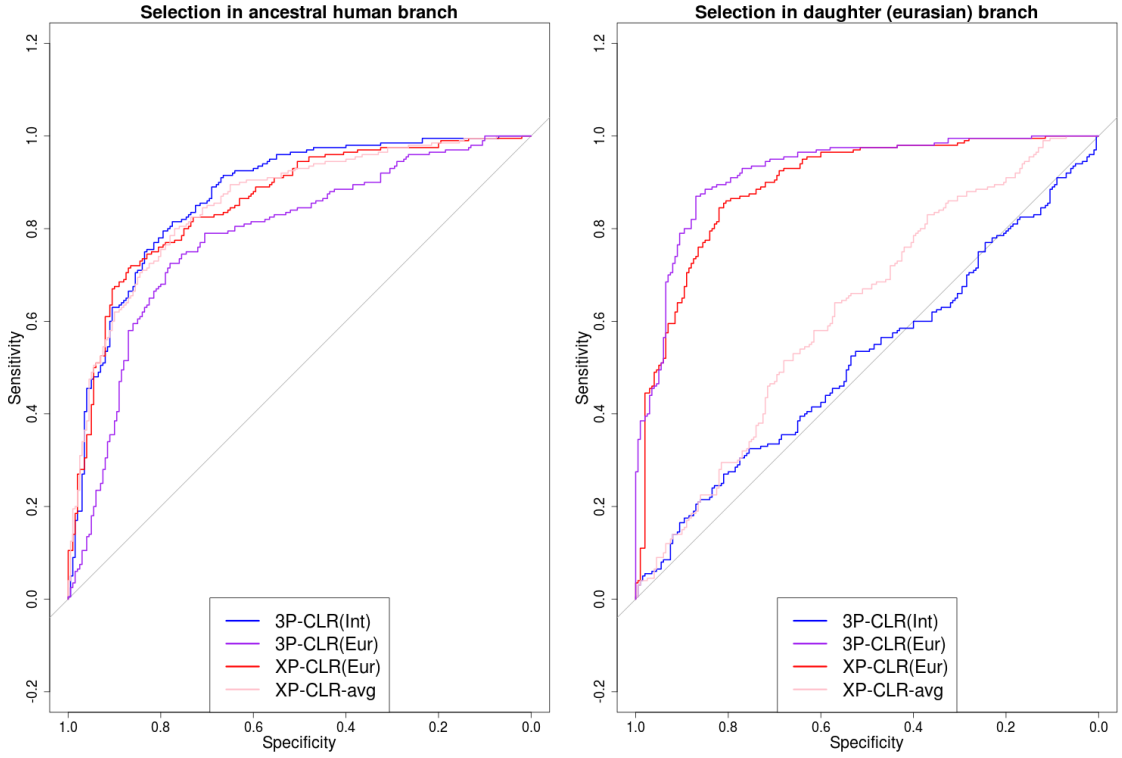

Figure S11. ROC curves for 3P-CLR run to detect selective events in the modern human ancestral branch, using simulations incorporating the history of population size changes and Neanderthal-to-Eurasian admixture inferred in Prüfer et al. (2014).

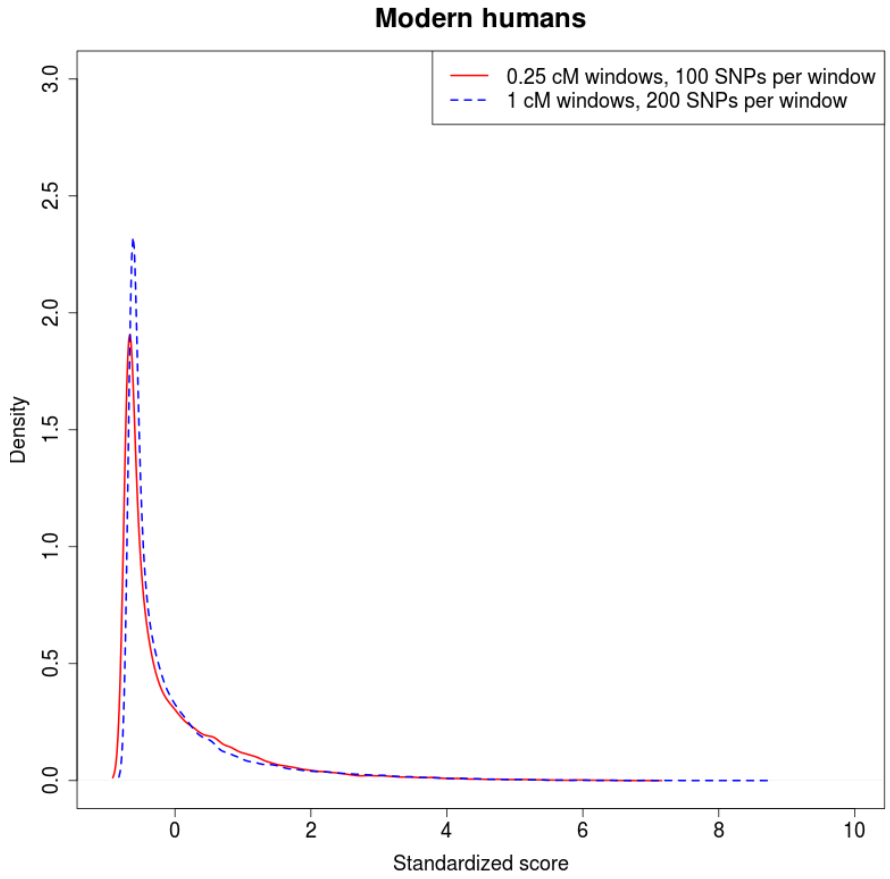

Figure S12. Comparison of 3P-CLR on the modern human ancestral branch under different window sizes and central SNP spacing. The red density is the density of standardized scores for 3P-CLR run using 0.25 cM windows, 100 SNPs per window and a spacing of 10 SNPs between each central SNP. The blue dashed density is the density of standardized scores for 3P-CLR run using 1 cM windows, 200 SNPs per window and a spacing of 40 SNPs between each central SNP.

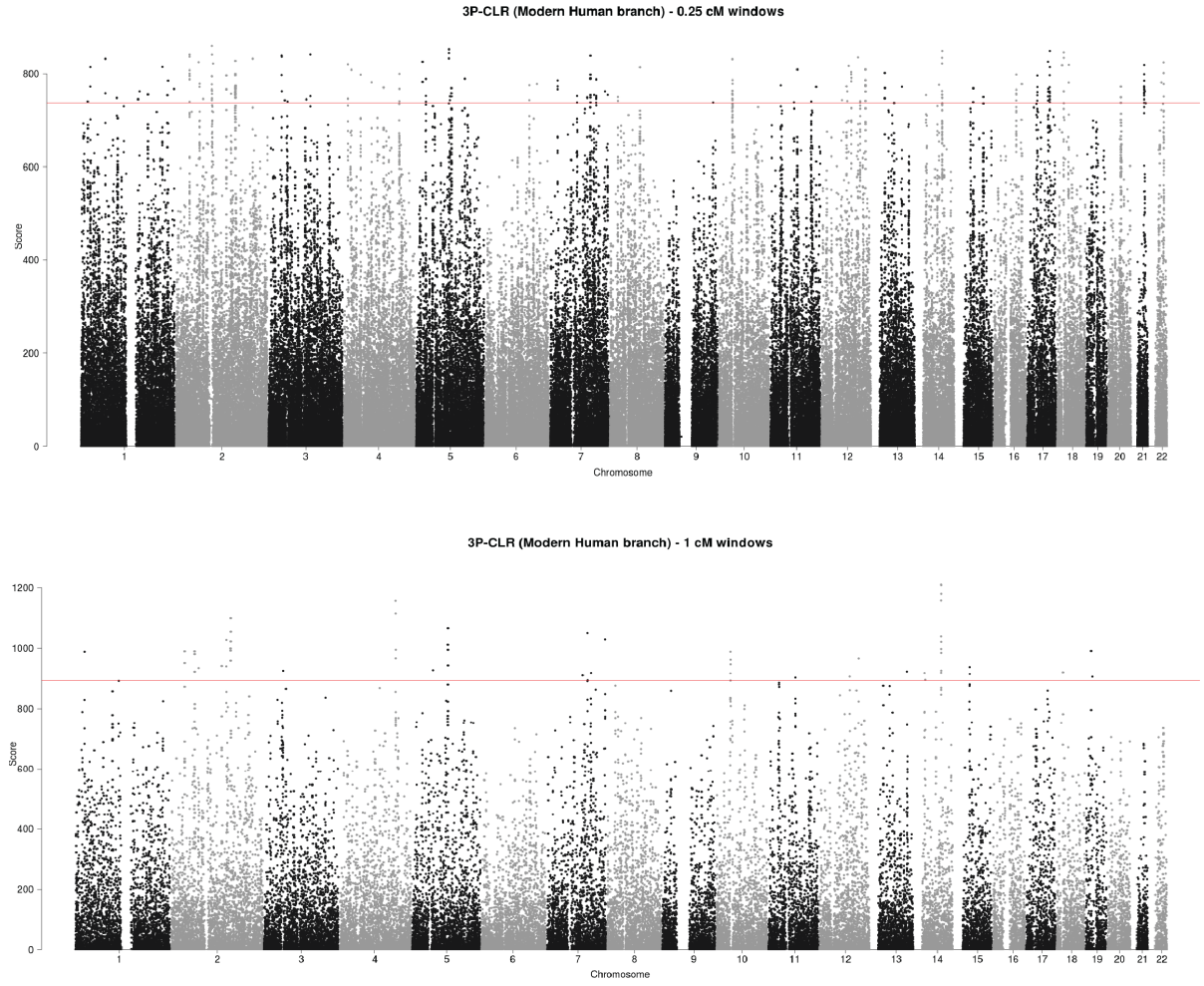

Figure S13. 3P-CLR scan of the ancestral branch to Yoruba and Eurasians, using the Denisovan and Neanderthal genomes as the outgroup. The red line denotes the 99.9% quantile cutoff. The top panel shows a run using 0.25 cM windows, each containing 100 SNPs, and sampling a candidate beneficial SNP every 10 SNPs. The bottom panels shows a run using 1 cM windows, each containing 200 SNPs, and sampling a candidate beneficial SNP every 40 SNPs.

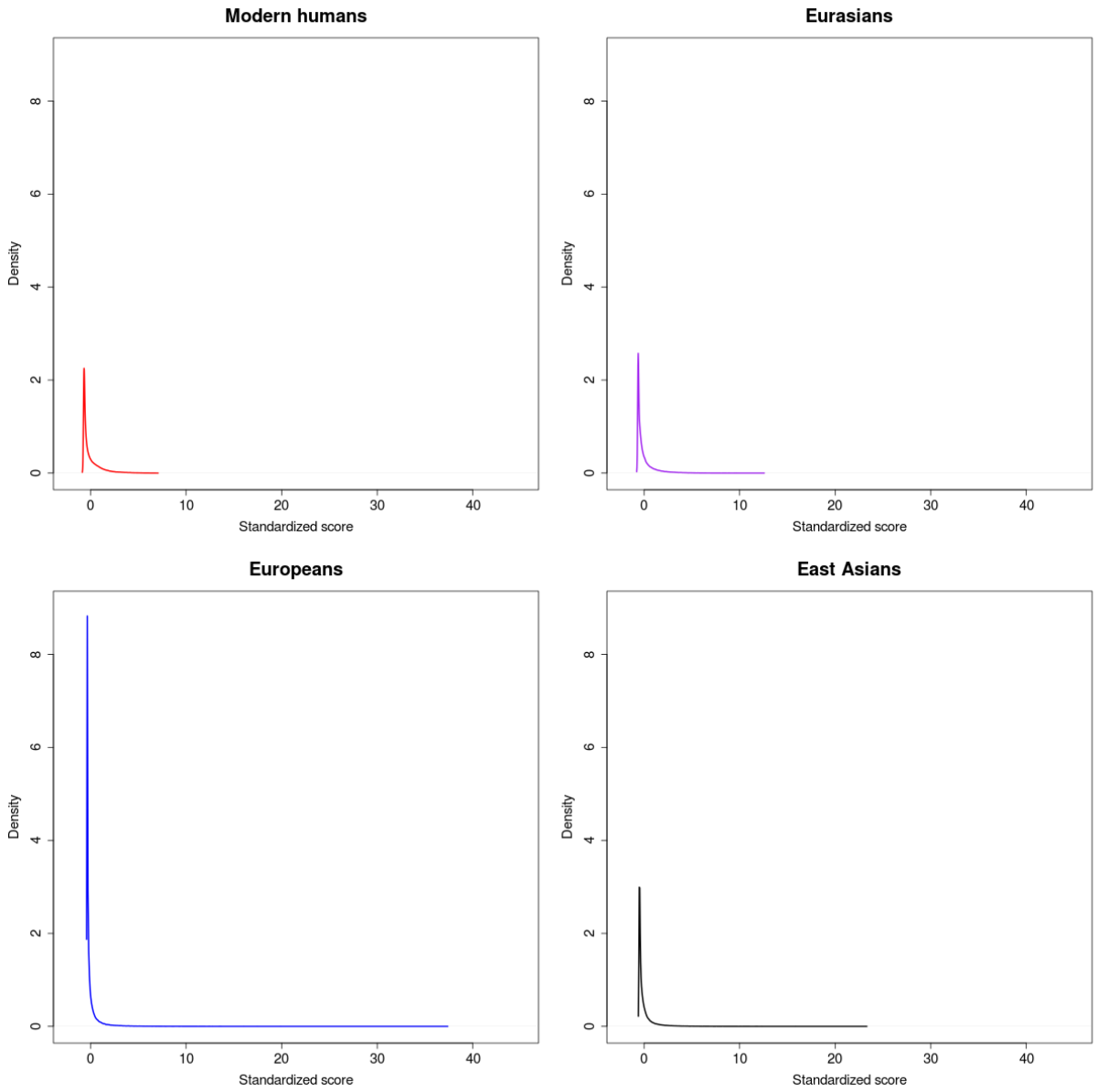

Figure S14. Genome-wide densities of each of the 3P-CLR scores described in this work. The distributions of scores testing for recent selection (Europeans and East Asians) have much longer tails than the distributions of scores testing for more ancient selection (Modern Humans and Eurasians). All scores were computed using 0.25 cM windows and were then standardized using their genome-wide means and standard deviations.

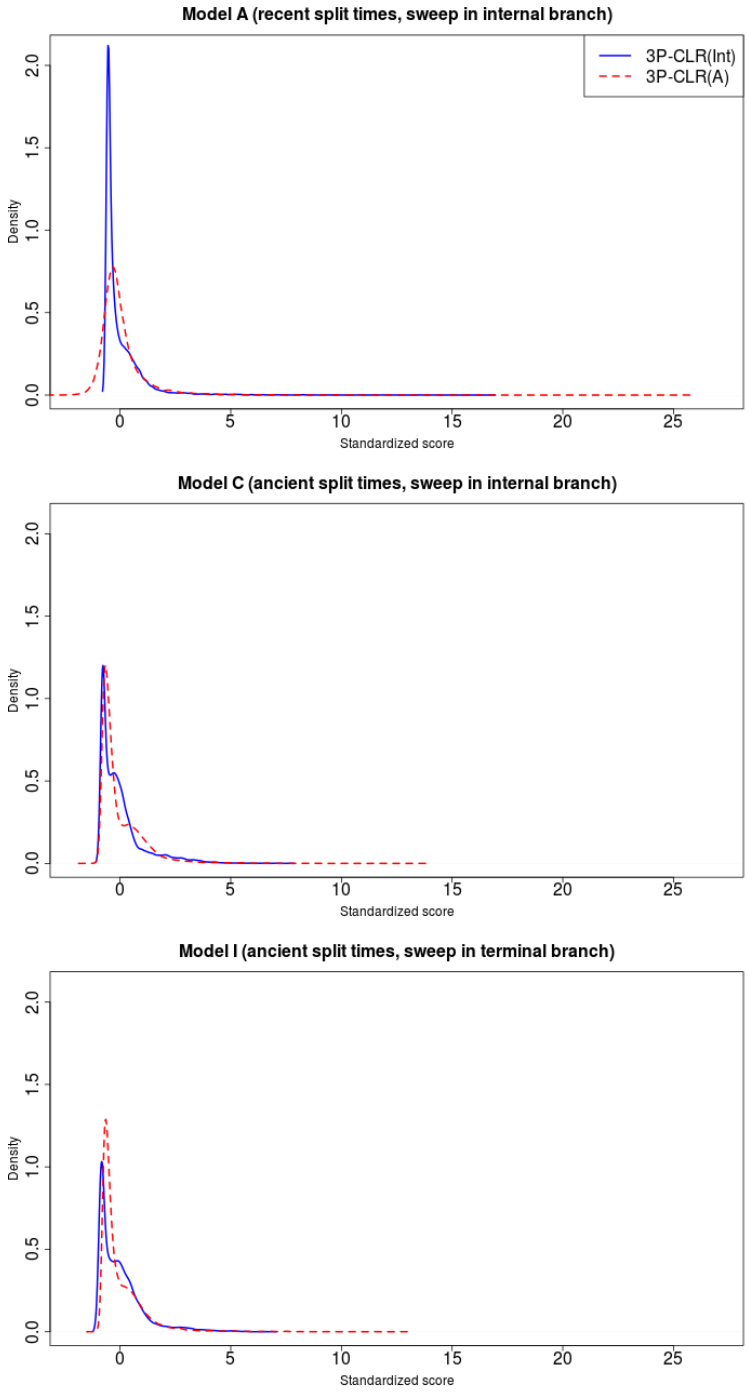

Figure S15. Distribution of 3P-CLR(Int) and 3P-CLR(A) scores under different demographic histories. We combined all scores obtained from 100 neutral simulations and 100 simulations with a selective sweep under different demographic and selection regimes. We then plotted the densities of the resulting scores. Top panel: Model A; Middle panel: Model C; Bottom panel: Model I. See Table 1 for details about each model.

Table S1. Top hits for 3P-CLR run on the European terminal branch, using Yoruba as the outgroup. We show the windows in the top 99.9% quantile of scores. Windows were merged together if the central SNPs that define them were contiguous. Win max = Location of window with maximum score. Win start = left-most end of left-most window for each region. Win end = right-most end of right-most window for each region. All positions were rounded to the nearest 100 bp. Score max = maximum score within region.

chr	Win max	Win start	Win end	Score max	Genes within region
9	125585000	125424000	126089000	362.273	ZBTB26,RABGAP1,GPR21,STRBP,OR1L1,OR1L3,OR1L4,OR1L6,OR5C1,PDCL,OR1K1, RC3H2,ZBTB6
22	35631900	35528100	35754100	309.488	HMGXB4,TOM1
8	52698800	52361800	52932100	289.921	PXDNL,PCMTD1
2	74967500	74450100	74972700	289.019	INO80B,WBP1,MOGS,MRPL53,CCDC142,TTC31,LBX2,PCGF1,TLX2,DQX1,AUP1, HTRA2,LOXL3,DOK1,M1AP,SEMA4F,SLC4A5,DCTN1,WDR54,RTKN
1	35634700	35382000	36592200	263.83	DL-GAP3,ZMYM6NB,ZMYM6,ZMYM1,SFPQ,ZMYM4,KIAA0319L,NCDN,TFAP2E,PSMB2, C1orf216,CLSPN,AGO4,AGO1,AGO3,TEKT2,ADPRHL2,COL8A2
15	29279800	29248000	29338300	251.944	APBA2
12	112950000	111747000	113030000	242.067	BRAP,ACAD10,ALDH2,MAPKAPK5,TMEM116,ERP29,NAA25,TRAFD1,RPL6,PTPN11, RPH3A,CUX2,FAM109A,SH2B3,ATXN2
9	90947700	90909300	91210000	219.285	SPIN1,NXNL2
19	33644300	33504200	33705700	213.189	RHPN2,GPATCH1,WDR88,LRP3,SLC7A10
9	30546800	30085400	31031600	207.378	-
4	33865300	33604700	34355600	204.96	-
1	198035000	197943000	198308000	197.96	NEK7
1	204868000	204681000	204873000	194.594	NFASC
10	74613800	73802300	75407100	191.864	SPOCK2,ASCC1,ANAPC16,DDIT4,DNAJB12,MICU1,MCU,OIT3,PLA2G12B,P4HA1, NUDT13,ECD,FAM149B1,DNAJC9,MRPS16,TTC18,ANXA7,MSS51,PPP3CB,USP54, MYOZ1,SYNPO2L
7	138809000	138798000	139136000	180.75	TTC26,UBN2,C7orf55,C7orf55-LUC7L2,LUC7L2
6	95678500	95351800	95831000	180.676	-
2	104752000	104592000	104951000	177.053	-
16	7602450	7528820	7612510	171.615	RBFOX1
10	30568100	30361300	30629500	170.714	KIAA1462,MTPAP
3	137183000	136873000	137250000	166.559	-
1	116731000	116709000	116919000	165.137	ATP1A1
9	135136000	135132000	135298000	165.004	SETX,TTF1,C9orf171
13	89882200	89262100	90103800	158.112	-
2	17094600	16977500	17173100	156.531	-
4	82050400	81981400	82125100	154.54	PRKG2
2	69245100	69147300	69342700	149.948	GKN2,GKN1,ANTXR1
17	46949100	46821000	47137900	147.537	ATP5G1,UBE2Z,SNF8,GIP,IGF2BP1,TTLL6,CALCOCO2
10	83993700	83977100	84328100	147.072	NRG3
14	63893800	63780300	64044700	142.831	PPP2R5E
1	244070000	243645000	244107000	142.335	SDCCAG8,AKT3
14	66636800	66417700	67889500	140.97	GPHN,FAM71D,MPP5,ATP6V1D,EIF2S1,PLEK2
11	38611200	38349600	39004500	138.731	-
3	123368000	123196000	123418000	136.651	PTPLB,MYLK
6	112298000	111392000	112346000	135.167	SLC16A10,KIAA1919,REV3L,TRAF3IP2,FYN
5	109496000	109419000	109608000	132.766	-
5	142160000	142070000	142522000	132.436	FGF1,ARHGAP26
12	39050200	33590600	39618900	130.832	SYT10,ALG10,ALG10B,CPNE8
9	108423000	108410000	108674000	129.893	TAL2,TMEM38B
3	159453000	159263000	159486000	126.462	IQCJ-SCHIP1
2	70182800	70020100	70563900	126.092	FAM136A,ANXA4,GMCL1,SNRNP27,MXD1,ASPRV1,PCBP1,C2orf42,TIA1,PCYOX1, SNRPG
3	177605000	177536000	177745000	123.927	-
8	18534300	18515900	18656800	123.593	PSD3
5	123555000	123371000	123603000	122.973	-
17	19287500	18887800	19443300	122.35	SLC5A10,FAM83G,GRAP,GRAPL,EPN2,B9D1,MAPK7,MFAP4,RNF112,SLC47A1
11	42236100	41807600	42311500	122.131	-
13	41623700	41119400	41801600	121.214	FOXO1,MRPS31,SLC25A15,ELF1,WBP4,KBTBD6,KBTBD7,MTRF1
5	10311500	10284000	10481500	118.766	CMBL,MARCH6,ROPN1L
14	65288500	65222500	65472700	118.576	SPTB,CHURC1,FNTB,GPX2,RAB15
1	47651700	47396900	47938300	118.241	CYP4A11,CYP4X1,CYP4Z1,CYP4A22,PDZK1IP1,TAL1,STIL,CMPK1,FOXE3,FOXD2
2	138527000	138428000	138694000	116.881	-
17	42294300	42056700	42351800	115.466	PYY,NAGS,TMEM101,LSM12,G6PC3,HDAC5,C17orf53,ASB16,TMUB2,ATXN7L3, UBTF,SLC4A1
9	12480000	12439900	12776500	115.209	TYRP1,LURAP1L
7	78743000	78688400	78897900	114.946	MAGI2
2	216626000	216556000	216751000	114.901	-
1	65511700	65377500	65611400	114.699	JAK1
5	115391000	115369000	115784000	113.862	ARL14EPL,COMMD10,SEMA6A
15	45402300	45096000	45490700	113.69	C15orf43,SORD,DUOX2,DUOXA2,DUOXA1,DUOX1,SHF
3	25840300	25705200	25934000	113.326	TOP2B,NGLY1,OXSM
2	73086900	72373800	73148200	110.523	CYP26B1,EXOC6B,SPR,EMX1

Table S2. Top hits for 3P-CLR run on the East Asian terminal branch, using Yoruba as the outgroup. We show the windows in the top 99.9% quantile of scores. Windows were merged together if the central SNPs that define them were contiguous. Win max = Location of window with maximum score. Win start = left-most end of left-most window for each region. Win end = right-most end of right-most window for each region. All positions were rounded to the nearest 100 bp. Score max = maximum score within region.

chr	Win max	Win start	Win end	Score max	Genes within region
15	64151100	63693900	64188300	266.459	USP3,FBXL22,HERC1
10	94962900	94830500	95093900	241.875	CYP26A1,MYOF
2	73086900	72353500	73170800	218.482	CYP26B1,EXOC6B,SPR,EMX1,SFXN5
10	55988000	55869200	56263600	215.051	PCDH15
1	234359000	234209000	234396000	189.946	SLC35F3
5	117350000	117344000	117714000	189.051	-
17	60964400	60907300	61547900	186.63	TANC2,CYB561
2	44268900	44101400	44315200	185.629	ABCG8,LRPPRC
11	6126830	6028090	6191240	184	OR56A1,OR56B4,OR52B2
2	109318000	108905000	109629000	183.859	LIMS1,RANBP2,CCDC138,EDAR,SULT1C2,SULT1C4,GCC2
4	41882900	41456100	42196500	183.481	LIMCH1,PHOX2B,TMEM33,DCAF4L1,SLC30A9,BEND4
18	5304160	5201440	5314680	183.476	ZBTB14
9	105040000	104779000	105042000	181.781	-
7	105097000	104526000	105128000	181.358	KMT2E,SRPK2,PUS7
3	107609000	107149000	107725000	178.27	BBX
7	101729000	101511000	101942000	169.558	CUX1
6	159274000	159087000	159319000	169.058	SYTL3,EZR,C6orf99
9	90947700	90909300	91202200	163.828	SPIN1,NXNL2
9	92311400	92294400	92495100	162.821	-
15	26885200	26723700	26911100	160.496	GABRB3
5	109197000	108988000	109240000	156.271	MAN2A1
3	12506200	12476600	12819300	151.978	TSEN2,C3orf83,MKRN2,RAF1,TMEM40
2	125998000	125740000	126335000	148.576	-
3	139052000	139033000	139351000	148.572	MRPS22,COPB2,RBP2,RBP1,NMNAT3
3	134739000	134629000	135618000	146.833	EPHB1
2	9766680	9354260	9774110	145.998	ASAP2,ITGB1BP1,CPSF3,IAH1,ADAM17,YWHAQ
3	17873800	17189600	18009400	145.345	TBC1D5
14	69592000	69423900	69791100	144.488	ACTN1,DCAF5,EXD2,GALNT16
22	39747800	39574300	39845300	144.477	PDGFB,RPL3,SYNGR1,TAB1
8	10875300	10731100	11094000	143.754	XKR6
4	99985900	99712200	100322000	143.554	EIF4E,METAP1,ADH5,ADH4,ADH6,ADH1A,ADH1B
4	144235000	143610000	144412000	143.124	INPP4B,USP38,GAB1
2	17596700	16574500	17994400	142.084	FAM49A,RAD51AP2,VSNL1,SMC6,GEN1
2	211707000	211652000	211873000	141.706	-
1	103763000	103353000	103785000	141.473	COL11A1
3	71482600	71372800	71685500	140.75	FOXP1
17	10519000	10280200	10564000	140.243	MYH8,MYH4,MYH1,MYH2,MYH3
4	13283100	13126100	13537100	139.729	RAB28
8	73836900	73815300	73953100	139.423	KCNB2,TERF1
14	50226700	49952500	50426100	139.052	RPS29,LRR1,RPL36AL,MGAT2,DNAAF2,POLE2,KLHDC1,KLHDC2,NEMF,ARF6
2	26167200	25895300	26238100	138.585	KIF3C,DTNB
6	47369600	47312800	47708400	138.112	CD2AP,GPR115
3	102005000	101899000	102361000	137.862	ZPLD1
1	65943500	65891700	66168800	137.68	LEPR,LEPROT
11	25169300	24892400	25274500	137.191	LUZP2
1	28846900	28430000	29177900	136.458	PTAFR,DNAJC8,ATPIF1,SESN2,MED18,PHACTR4,RCC1,TRNAU1AP,TAF12,RAB42, GMEB1,YTHDF2,OPRD1
2	154054000	154009000	154319000	136.247	-
7	108874000	108718000	109226000	135.996	-
1	75471000	75277400	75941000	133.055	LHX8,SLC44A5
1	154824000	154802000	155113000	131.45	KCNN3,PMVK,PBXIP1,PYGO2,SHC1,CKS1B,FLAD1,LENEP,ZBTB7B,DCST2,DCST1, ADAM15,EFNA4,EFNA3,EFNA1,SLC50A1,DPM3
3	58413700	58096400	58550500	130.828	FLNB,DNASE1L3,ABHD6,RPP14,PXK,PDHB,KCTD6,ACOX2,FAM107A
1	36170500	35690600	36592200	130.701	ZMYM4,KIAA0319L,NCDN,TFAP2E,PSMB2,C1orf216,CLSPN,AGO4,AGO1,AGO3, TEKT2,ADPRHL2,COL8A2
17	39768900	39673200	39865400	130.04	KRT15,KRT19,KRT9,KRT14,KRT16,KRT17,JUP,EIF1
15	82080400	81842500	82171400	129.682	-
17	30842700	30613600	30868000	128.36	RHBDL3,C17orf75,ZNF207,PSMD11,CDK5R1,MYO1D
2	107933000	107782000	108041000	128.04	-
3	44917100	44138200	45133100	127.824	TOPAZ1,TCAIM,ZNF445,ZKSCAN7,ZNF660,ZNF197,ZNF35,ZNF502,ZNF501,KIAA1143, KIF15,TMEM42,TGM4,ZDHHC3,EXOSC7,CLEC3B,CDCP1
4	153009000	152902000	153101000	126.503	-
22	43190000	43148300	43455100	126.326	ARFGAP3,PACSIN2,TTLL1
4	168849000	168619000	168995000	126.125	-
5	42286000	41478600	42623200	125.831	PLCXD3,OXCT1,C5orf51,FBXO4,GHR
7	136345000	135788000	136570000	125.551	CHRM2
3	60305100	60226500	60349500	125.16	FHIT
10	59763900	59572200	59825500	124.643	-
3	114438000	114363000	115146000	124.535	ZBTB20
4	160142000	159944000	160359000	123.391	C4orf45,RAPGEF2
2	177717000	177613000	177889000	123.094	-
5	119672000	119639000	119868000	122.93	PRR16
20	43771800	43592200	43969300	122.421	STK4,KCNS1,WFDC5,WFDC12,PI3,SEMG1,SEMG2,SLPI,MATN4,RBPJL,SDC4
1	172928000	172668000	172942000	121.532	-
7	112273000	112126000	112622000	121.336	LSMEM1,TMEM168,C7orf60
1	169523000	169103000	169525000	119.533	NME7,BLZF1,CCDC181,SLC19A2,F5
3	26265100	25931700	26512400	119.052	-

Table S3. Top hits for 3P-CLR run on the Eurasian ancestral branch, using Yoruba as the outgroup. We show the windows in the top 99.9% quantile of scores. Windows were merged together if the central SNPs that define them were contiguous. Win max = Location of window with maximum score. Win start = left-most end of left-most window for each region. Win end = right-most end of right-most window for each region. All positions were rounded to the nearest 100 bp. Score max = maximum score within region.

chr	Win max	Win start	Win end	Score max	Genes within region
2	72379700	72353500	73170800	617.695	CYP26B1,EXOC6B,SPR,EMX1,SFXN5
20	53879500	53876700	54056200	605.789	-
10	22712400	22309300	22799200	566.463	EBLN1,COMMD3,COMMD3-BMI1,BMI1,SPAG6
3	25856600	25726300	26012000	557.376	NGLY1,OXSM
18	67725100	67523300	67910500	535.743	CD226,RTTN
10	66262400	65794400	66339100	532.732	-
11	39695600	39587400	39934300	518.72	-
7	138927000	138806000	139141000	508.385	TTC26,UBN2,C7orf55,C7orf55-LUC7L2,LUC7L2,KLRG2
9	90934600	90909300	91202200	498.898	SPIN1,NXNL2
4	41554200	41454200	42195300	487.476	LIMCH1,PHOX2B,TMEM33,DCAF4L1,SLC30A9,BEND4
16	61271700	61121600	61458700	485.291	-
17	58509300	58113700	59307700	477.117	HEATR6,CA4,USP32,C17orf64,APPBP2,PPM1D,BCAS3
1	230132000	229910000	230208000	468.258	GALNT2
8	35540400	35533900	35913800	454.601	UNC5D
17	60964400	60907300	61547900	449.203	TANC2,CYB561
16	47972300	33707000	48480500	448.504	SHCBP1,VPS35,ORC6,MYLK3,C16orf87,GPT2,DNAJA2,NETO2,ITFG1,PHKB,ABCC12, ABCC11,LONP2,SIAH1
1	90393900	90329700	90521600	436.002	LRRC8D,ZNF326
8	52698800	52238900	52932100	423.865	PXDNL,PCMTD1
11	106237000	105877000	106256000	419.391	MSANTD4,KBTBD3,AASDHPPT
13	48798100	48722300	49288100	414.218	ITM2B,RB1,LPAR6,RCBTB2,CYSLTR2
3	19240300	19090800	19424900	408.064	KCNH8
2	194986000	194680000	195299000	404.394	-
12	15962600	15690100	16137200	402.558	PTPRO,EPS8,STRAP,DERA
9	125564000	125484000	126074000	400.096	ZBTB26,RABGAP1,GPR21,STRBP,OR1L4,OR1L6,OR5C1,PDCL,OR1K1,RC3H2,ZBTB6
15	28565300	28324600	28611900	398.519	OCA2,HERC2
8	47631700	42502000	49037700	396.687	CHRNB3,CHRNA6,THAP1,RNF170,HOOK3,FNTA,POMK,HGSNAT,SPIDR,CEBPD, MCM4,UBE2V2
1	116994000	116808000	117027000	395.221	ATP1A1
7	99338700	98717600	99376500	393.41	ZSCAN25,CYP3A5,CYP3A7,CYP3A4,SMURF1,KPNA7,ARPC1A,ARPC1B,PDAP1, BUD31,PTCD1,ATP5J2- PTCD1,CPSF4,ATP5J2,ZNF789,ZNF394,ZKSCAN5,FAM200A,ZNF655
7	30343200	30178800	30485700	391.828	MTURN,ZNRF2,NOD1
10	31583000	31430600	31907900	389.863	ZEB1
6	10647900	10583800	10778900	387.883	GCNT2,C6orf52,PAK1IP1,TMEM14C,TMEM14B,SYCP2L,MAK
11	123275000	123156000	123313000	386.485	-
15	64642400	64333700	65204100	385.748	DAPK2,FAM96A,SNX1,SNX22,PPIB,CSNK1G1,KIAA0101,TRIP4,ZNF609,OAZ2, RBPMS2,PIF1,PLEKHO2
2	222560000	222523000	222690000	383.336	-
6	43620800	43398400	43687800	378.463	ABCC10,DLK2,TJAP1,LRRC73,POLR1C,YIPF3,XPO5,POLH,GTPBP2,MAD2L1BP, RSPH9,MRPS18A
14	57643800	57603400	58047900	378.332	EXOC5,AP5M1,NAA30,C14orf105
4	33487100	33294500	34347100	377.815	-
3	188699000	188647000	188856000	373.617	TPRG1
17	46949100	46821000	47137900	371.886	ATP5G1,UBE2Z,SNF8,GIP,IGF2BP1,TTLL6,CALCOCO2
4	172656000	172565000	172739000	369.949	GALNTL6
15	34404500	34212600	34413500	369.949	AVEN,CHRM5,EMC7,PGBD4
1	32888000	32445400	33065900	369.725	KHDRBS1,TMEM39B,KPNA6,TXLNA,CCDC28B,IQCC,DCDC2B,TMEM234,EIF3I, FAM167B,LCK,HDAC1,MARCKSL1,TSSK3,FAM229A,BSDC1,ZBTB8B,ZBTB8A,ZBTB8OS
22	46820900	46593300	46834700	369.511	PPARA,CDPF1,PKDREJ,TTC38,GTSE1,TRMU,CELSR1
10	93143600	93060500	93324900	368.648	HECTD2
6	14845800	14753800	14948200	367.9	-

Table S4. Enriched GO categories in the European, East Asian and Modern Human branches. We tested for ontology enrichment among the regions in the 99.5% quantile of the 3P-CLR scores for each population branch ($P < 0.05$, FDR < 0.3). The Eurasian branch did not have any category that passed these cutoffs.

Population Branch	Raw p-value	FDR	GO category
European	0.00002	0.05977	cuticle development
European	0.00007	0.096085	hydrogen peroxide catabolic process
East Asian	0.00001	0.013385	regulation of cell adhesion mediated by integrin
East Asian	0.00001	0.013385	epidermis development
East Asian	0.00014	0.14102	cell-substrate adhesion
East Asian	0.00023	0.185135	nucleosomal DNA binding
East Asian	0.0003	0.185135	nuclear chromosome
East Asian	0.00033	0.185135	RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription
East Asian	0.00048	0.2023525	negative regulation of vitamin metabolic process
East Asian	0.00048	0.2023525	substrate adhesion-dependent cell spreading
East Asian	0.00058	0.219074444	regulation of ERK1 and ERK2 cascade
East Asian	0.00077	0.258110909	retinol binding
East Asian	0.00084	0.258110909	primary alcohol catabolic process
East Asian	0.00112	0.296474	D1 dopamine receptor binding
East Asian	0.00125	0.296474	RNA polymerase II transcription regulatory region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription
East Asian	0.00127	0.296474	positive regulation of protein kinase B signaling
East Asian	0.0013	0.296474	gap junction assembly
Modern Human	0.00002	0.031153333	nuclear division
Modern Human	0.00003	0.031153333	organelle fission
Modern Human	0.00003	0.031153333	mitosis
Modern Human	0.00006	0.0490675	intra-Golgi vesicle-mediated transport
Modern Human	0.00012	0.069241429	regulation of cell cycle
Modern Human	0.00014	0.069241429	retinoic acid-responsive element binding
Modern Human	0.00015	0.069241429	cell cycle process
Modern Human	0.00029	0.12784125	T cell migration
Modern Human	0.00041	0.162306667	chromosomal part
Modern Human	0.00055	0.198124	'de novo' IMP biosynthetic process
Modern Human	0.00072	0.237017273	intracellular organelle
Modern Human	0.00081	0.24451	SNAP receptor activity
Modern Human	0.00113	0.294514286	ATP-dependent protein binding
Modern Human	0.00114	0.294514286	RNA biosynthetic process

Table S5. Top hits for 3P-CLR run on the ancestral branch to Eurasians and Yoruba, using archaic humans as the outgroup and 0.25 cM windows.. We show the windows in the top 99.9% quantile of scores. Windows were merged together if the central SNPs that define them were contiguous. Win max = Location of window with maximum score. Win start = left-most end of left-most window for each region. Win end = right-most end of right-most window for each region. All positions were rounded to the nearest 100 bp. Score max = maximum score within region.

chr	Win max	Win start	Win end	Score max	Genes within region
2	95724900	95561200	96793700	859.783	ZNF514,ZNF2,PROM2,KCNIP3,FAHD2A,TRIM43,GPAT2,ADRA2B,ASTL,MAL,MRPS5
5	87054300	86463700	87101400	852.543	RASA1,CCNH
17	61538200	60910700	61557700	849.335	TANC2,CYB561,ACE
14	72207400	71649200	72283600	849.304	SIPA1L1
18	19089800	15012100	19548600	846.182	ROCK1,GREB1L,ESCO1,SNRPD1,ABHD3,MIB1
3	110675000	110513000	110932000	841.499	PVRL3
2	37990900	37917900	38024200	841.339	CDC42EP3
3	36938000	36836900	37517500	839.211	TRANK1,EPM2AIP1,MLH1,LRRFIP2,GOLGA4,C3orf35,ITGA9
7	107246000	106642000	107310000	838.948	PRKAR2B,HBP1,COG5,GPR22,DUS4L,BCAP29,SLC26A4
12	96986900	96823000	97411500	835	NEDD1
2	201056000	200639000	201340000	832.4	C2orf69,TYW5,C2orf47,SPATS2L
1	66851800	66772600	66952600	832.221	PDE4B
10	37795700	37165100	38978800	831.353	ANKRD30A,MTRNR2L7,ZNF248,ZNF25,ZNF33A,ZNF37A
2	156129000	155639000	156767000	827.839	KCNJ3
17	56516700	56379200	57404800	826.026	BZRAP1,SUPT4H1,RNF43,HSF5,MTMR4,SEPT4,C17orf47,TEX14,RAD51C,PPM1E, TRIM37,SKA2,PRR11,SMG8,GDPD1
5	18755900	18493900	18793500	825.858	-
2	61190300	61050900	61891900	824.962	REL,PUS10,PEX13,KIAA1841,AHSA2,USP34,XPO1
22	40392200	40360300	41213400	824.52	GRAP2,FAM83F,TNRC6B,ADSL,SGSM3,MKL1,MCHR1,SLC25A17
2	99013400	98996400	99383400	821.891	CNGA3,INPP4A,COA5,UNC50,MGAT4A
4	13294400	13137000	13533100	820.222	RAB28
18	32975600	32604100	33002800	819.128	MAPRE2,ZNF397,ZSCAN30,ZNF24,ZNF396
21	35204700	34737300	35222100	818.754	IFNGR2,TMEM50B,DNAJC28,GART,SON,DONSON,CRYZL1,ITSN1
12	73048100	72740100	73160400	816.903	TRHDE
1	213511000	213150000	213563000	814.632	VASH2,ANGEL2,RPS6KC1
1	27500300	26913700	27703900	814.332	ARID1A,PIGV,ZDHHC18,SFN,GPN2,GPATCH3,NUDC,NR0B2,C1orf172,TRNP1, FAM46B,SLC9A1,WDTC1,TMEM222,SYTL1,MAP3K6,FCN3
8	79219300	78698200	79558000	813.796	PKIA
12	116455000	116380000	116760000	809.406	MED13L
11	72857900	72416300	72912800	809.274	ARAP1,STARD10,ATG16L2,FCHSD2
4	22941400	22827300	23208900	808.696	-
12	79783400	79748800	80435300	804.117	SYT1,PAWR,PPP1R12A
13	35534800	35429700	36097500	801.815	NBEA,MAB21L1
4	146141000	145514000	146214000	799.686	HHIP,ANAPC10,ABCE1,OTUD4
16	61429300	61124400	61458700	798.318	-
4	46530000	46360000	46881700	797.876	GABRA2,COX7B2
2	133038000	132930000	133117000	796.277	-
17	28980100	28549700	29407200	796.136	SLC6A4,BLMH,TMIGD1,CPD,GOSR1,TBC1D29,CRLF3,ATAD5,TEFM,ADAP2,RNF135
5	127332000	127156000	127607000	789.339	SLC12A2,FBN2
5	27208300	27072700	27352900	788.924	CDH9
7	122294000	121973000	122559000	787.777	CADPS2,RNF133,RNF148
10	38218900	37175000	43224100	786.651	ANKRD30A,MTRNR2L7,ZNF248,ZNF25,ZNF33A,ZNF37A,ZNF33B
7	23100200	22888500	23114300	785.919	FAM126A
1	228050000	227587000	228112000	785.53	SNAP47,JMJD4,PRSS38,WNT9A
4	74891400	74846600	75086500	781.895	PF4,PPBP,CXCL5,CXCL3,CXCL2,MTHFD2L
22	34588400	34516300	34811800	781.522	-
2	63899700	62767900	64395700	778.951	EHBP1,OTX1,WDPCP,MDH1,UGP2,VPS54,PELI1
6	136666000	136527000	136967000	778.233	MTFR2,BCLAF1,MAP7,MAP3K5
16	75738400	75522400	75968000	778.171	CHST6,CHST5,TMEM231,GABARAPL2,ADAT1,KARS,TERF2IP
14	63446800	63288600	63597500	776.567	KCNH5
6	117528000	117080000	117579000	775.402	FAM162B,GPRC6A,RFX6
11	30206400	29986200	30443900	775.051	KCNA4,FSHB,ARL14EP,MPPED2
12	67533400	67436200	67639400	772.731	-
20	35460500	35049400	35710900	772.319	DLGAP4,MYL9,TGIF2,TGIF2- C20orf24,C20orf24,SLA2,NDRG3,DSN1,SOGA1,TLDC2,SAMHD1,RBL1
13	80131900	79801800	80268900	771.976	RBM26,NDFIP2
11	121408000	121310000	121493000	771.669	SORL1
4	105305000	104931000	105454000	770.437	CXXC4
5	93218900	92677500	93647600	769.192	NR2F1,FAM172A,POU5F2,KIAA0825
15	49975000	49247500	50040200	768.997	SECISBP2L,COPS2,GALK2,FAM227B,FGF7,DTWD1,SHC4
1	243669000	243505000	244087000	767.303	SDCCAG8,AKT3
21	36822500	36691000	36883300	762.715	RUNX1
1	154133000	153745000	154280000	762.43	INTS3,SLC27A3,GATAD2B,DENND4B,CRTC2,SLC39A1,CREB3L4,JTB,RAB13, RPS27,NUP210L,TPM3,C1orf189,C1orf43,UBAP2L,HAX1
7	144655000	144465000	144700000	762.429	TPK1
12	69177500	68890300	69290800	762.399	RAP1B,NUP107,SLC35E3,MDM2,CPM
2	145116000	144689000	145219000	757.235	GTDC1,ZEB2
1	176195000	175890000	176437000	755.81	RFWD2,PAPPA2
7	152155000	151699000	152199000	754.754	GALNTL5,GALNT11,KMT2C
7	116575000	116324000	116788000	754.606	MET,CAPZA2,ST7
14	29571400	29264600	29691100	754.435	-
1	226323000	226140000	226575000	754.04	SDE2,H3F3A,ACBD3,MIXL1,LIN9,PARP1
7	73051800	72317200	73134700	752.285	POM121,TRIM74,NSUN5,TRIM50,FKBP6,FZD9,BAZ1B,BCL7B,TBL2,MLXIPL,VPS37D, DNAJC30,WBSCR22,STX1A
5	89578700	89408400	89654700	751.498	-
8	22999100	22926500	23113900	749.992	TNFRSF10B,TNFRSF10C,TNFRSF10D,TNFRSF10A,CHMP7
15	75883900	75462000	76038100	749.953	C15orf39,GOLGA6C,GOLGA6D,COMMD4,NEIL1,MAN2C1,SIN3A,PTPN9,SNUPN,IMP3, SNX33,CSPG4,ODF3L1
7	98978400	98719400	99376100	749.35	ZSCAN25,CYP3A5,CYP3A7,CYP3A4,SMURF1,KPNA7,ARPC1A,ARPC1B,PDAP1, BUD31,PTCD1,ATP5J2- PTCD1,CPSF4,ATP5J2,ZNF789,ZNF394,ZKSCAN5,FAM200A,ZNF655

1	96340100	96155200	96608300	748.253	-
2	73508400	73482800	74054300	745.963	FBXO41,EGR4,ALMS1,NAT8,TPRKB,DUSP11,C2orf78
1	150868000	150224000	151137000	745.222	CA14,APH1A,C1orf54,C1orf51,MRPS21,PRPF3,RPRD2,TARS2,ECM1,ADAMTSL4, MCL1,ENSA,GOLPH3L,HORMAD1,CTSS,CTSK,ARNT,SETDB1,CERS2,ANXA9, FAM63A,PRUNE,BNIPL,C1orf56,CDC42SE1,MLLT11,GABPB2,SEMA6C,TNFAIP8L2, SCNM1,LYSMD1
3	99877600	99374500	100207000	744.933	COL8A1,CMSS1,FILIP1L,TBC1D23,NIT2,TOMM70A,LNP1
12	56244900	56086600	56360700	743.698	PMEL,CDK2,ITGA7,BLOC1S1,RDH5,CD63,GDF11,SARNP,ORMDL2,DNAJC14,MMP19, WIBG,DGKA
3	44843200	44139200	45128900	743.157	TOPAZ1,TCAIM,ZNF445,ZKSCAN7,ZNF660,ZNF197,ZNF35,ZNF502,ZNF501,KIAA1143, KIF15,TMEM42,TGM4,ZDHHC3,EXOSC7,CLEC3B,CDCP1
12	102922000	102388000	102964000	741.338	DRAM1,CCDC53,NUP37,PARPBP,PMCH,IGF1
1	21114300	21012100	21636800	740.553	KIF17,SH2D5,HP1BP3,EIF4G3,ECE1
11	108770000	108492000	108830000	740.463	DDX10
3	51678700	50188500	51919700	740.272	SEMA3F,GNAT1,GNAI2,LSMEM2,IFRD2,HYAL3,NAT6,HYAL1,HYAL2,TUSC2,RASSF1, ZMYND10,NPRL2,CYB561D2,TMEM115,CACNA2D2,C3orf18,HEMK1,CISH,MAPKAPK3, DOCK3,MANF,RBM15B,RAD54L2,TEX264,GRM2,IQCF6,IQCF3,IQCF2,IQCF5
11	64581900	64293300	64589300	738.648	RASGRP2,PYGM,SF1,MAP4K2,MEN1,SLC22A11,SLC22A12,NRXN2
9	126023000	125542000	126076000	738.221	ZBTB26,RABGAP1,GPR21,STRBP,OR5C1,PDCL,OR1K1,RC3H2,ZBTB6

Table S6. Top hits for 3P-CLR run on the ancestral branch to Eurasians and Yoruba, using archaic humans as the outgroup and 1 cM windows. We show the windows in the top 99.9% quantile of scores. Windows were merged together if the central SNPs that define them were contiguous. Win max = Location of window with maximum score. Win start = left-most end of left-most window for each region. Win end = right-most end of right-most window for each region. All positions were rounded to the nearest 100 bp. Score max = maximum score within region.

chr	Win max	Win start	Win end	Score max	Genes within region
14	71698500	71349200	72490300	1210.24	PCNX,SIPA1L1,RGS6
4	145534000	145023000	146522000	1157.25	GYPB,GYPA,HHIP,ANAPC10,ABCE1,OTUD4,SMAD1
2	156103000	155391000	156992000	1100.35	KCNJ3
5	93425300	92415600	94128600	1065.66	NR2F1,FAM172A,POU5F2,KIAA0825,ANKRD32,MCTP1
7	106717000	106401000	107461000	1049.82	PIK3CG,PRKAR2B,HBP1,COG5,GPR22,DUS4L,BCAP29,SLC26A4,CBLL1,SLC26A3
7	151831000	151651000	152286000	1028.93	GALNTL5,GALNT11,KMT2C
2	145008000	144393000	145305000	1027.28	ARHGAP15,GTDC1,ZEB2
19	16578500	16387600	16994000	991.083	KLF2,EPS15L1,CALR3,C19orf44,CHERP,SLC35E1,MED26,SMIM7,TMEM38A,NWD1, SIN3B
2	37996300	37730400	38054600	989.901	CDC42EP3
2	63467700	62639800	64698300	989.891	TMEM17,EHBP1,OTX1,WDPCP,MDH1,UGP2,VPS54,PELI1,LGALSL
10	38074100	36651400	44014800	988.663	ANKRD30A,MTRNR2L7,ZNF248,ZNF25,ZNF33A,ZNF37A,ZNF33B,BMS1,RET, CSGALNACT2,RASGEF1A,FXYD4,HNRNPF
1	27203100	26703800	27886000	988.598	LIN28A,DHDDS,HMGN2,RPS6KA1,ARID1A,PIGV,ZDHHC18,SFN,GPN2,GPATCH3, NUDC,NR0B2,C1orf172,TRNP1,FAM46B,SLC9A1,WDTC1,TMEM222,SYTL1,MAP3K6, FCN3,CD164L2,GPR3,WASF2,AHDC1
12	102906000	102308000	103125000	966.591	DRAM1,CCDC53,NUP37,PARPBP,PMCH,IGF1
2	133034000	132628000	133270000	941.856	GPR39
15	43507200	42284300	45101400	938.129	PLA2G4E,PLA2G4D,PLA2G4F,VPS39,TMEM87A,GANC,CAPN3,ZNF106,SNAP23, LRRC57,HAUS2,STARD9,CDAN1,TTBK2,UBR1,EPB42,TMEM62,CCNDBP1,TGM5, TGM7,LCMT2,ADAL,ZSCAN29,TUBGCP4,TP53BP1,MAP1A,PPIP5K1,CKMT1B,STRC, CATSPER2,CKMT1A,PDIA3,ELL3,SERF2,SERINC4,HYPK,MFAP1,WDR76,FRMD5, CASC4,CTDSPL2,EIF3J,SPG11,PATL2,B2M,TRIM69
2	73848400	73178500	74194400	934.997	SFXN5,RAB11FIP5,NOTO,SMYD5,PRADC1,CCT7,FBXO41,EGR4,ALMS1,NAT8,TPRKB, DUSP11,C2orf78,STAMBP,ACTG2,DGUOK
5	54861800	54193000	55422100	927.745	ESM1,GZMK,GZMA,CDC20B,GPX8,MCIDAS,CCNO,DHX29,SKIV2L2,PPAP2A,SLC38A9, DDX4,IL31RA,IL6ST,ANKRD55
3	52356200	50184000	53602300	925.895	SEMA3F,GNAT1,GNAI2,LSMEM2,IFRD2,HYAL3,NAT6,HYAL1,HYAL2,TUSC2,RASSF1, ZMYND10,NPRL2,CYB561D2,TMEM115,CACNA2D2,C3orf18,HEMK1,CISH,MAPKAPK3, DOCK3,MANF,RBM15B,RAD54L2,TEX264,GRM2,IQCF6,IQCF3,IQCF2,IQCF5,IQCF1, RRP9,PARP3,GPR62,PCBP4,ABHD14B,ABHD14A,ACY1,RPL29,DUSP7,POC1A,ALAS1, TLR9,TWF2,PPM1M,WDR82,GLYCTK,DNAH1,BAP1,PHF7,SEMA3G,TNNC1,NISCH, STAB1,NT5DC2,SMIM4,PBRM1,GNL3,GLT8D1,SPCS1,NEK4,ITIH1,ITIH3,ITIH4, MUSTN1,TMEM110-MUSTN1,TMEM110,SFMBT1,RFT1,PRKCD,TKT,CACNA1D
13	96364900	96038900	97500100	923.257	CLDN10,DZIP1,DNAJC3,UGGT2,HS6ST3
18	19248800	14517500	19962400	920.641	POTEC,ANKRD30B,ROCK1,GREB1L,ESCO1,SNRPD1,ABHD3,MIB1,GATA6
7	116587000	116214000	117339000	918.567	MET,CAPZA2,ST7,WNT2,ASZ1,CFTR
14	29544300	29031800	29913200	918.292	FOXG1
7	94710700	93964000	95170200	910.235	COL1A2,CASD1,SGCE,PEG10,PPP1R9A,PON1,PON3,PON2,ASB4
12	79783400	79231600	80435300	906.28	SYT1,PAWR,PPP1R12A
19	19290700	18936200	19885600	905.94	UPF1,CERS1,GDF1,COPE,DDX49,HOMER3,SUGP2,ARMC6,SLC25A42,TMEM161A, MEF2BNB-MEF2B,MEF2B,MEF2BNB,RFXANK,NR2C2AP,NCAN,HAPLN4,TM6SF2,SUGP1,MAU2, GATAD2A,TSSK6,NDUFA13,YJEFN3,CILP2,PBX4,LPAR2,GMIP,ATP13A1,ZNF101, ZNF14
11	72551000	72182800	72952400	902.837	PDE2A,ARAP1,STARD10,ATG16L2,FCHSD2,P2RY2
14	31685700	31255700	32384600	895.417	COCH,STRN3,AP4S1,HECTD1,DTD2,NUBPL

Table S7. Overlap between GWAS catalog and catalog of modern human-specific high-frequency changes in the top modern human selected regions (0.25 cM scan). Chr = chromosome. Pos = position (hg19). ID = SNP rs ID. Hum = Present-day human major allele. Anc = Human-Chimpanzee ancestor allele. Arch = Archaic human allele states (Altai Neanderthal, Denisova) where H=human-like allele and A=ancestral allele. Freq = present-day human derived frequency. Cons = consequence. C = C-score. PubMed = PubMed article ID for GWAS study.

Chr	Pos	ID	Hum	Anc	Arch	Freq	Gene	Cons	C	GWAS trait	PubMed
1	27138393	rs12748152	C	T	A/A,A/A	0.95	Metazoa SRP	upstream	4.193	HDL cholesterol	24097068
1	27138393	rs12748152	C	T	A/A,A/A	0.95	Metazoa SRP	upstream	4.193	LDL cholesterol	24097068
1	27138393	rs12748152	C	T	A/A,A/A	0.95	Metazoa SRP	upstream	4.193	Triglycerides	24097068
1	151009719	rs1534059	A	G	A/A,A/A	0.92	BNIPL	intron	7.111	DNA methylation, in blood cell lines	21251332
1	244044810	rs7553354	A	C	A/A,A/A	0.94	NA	intergenic	2.376	Response to taxane treatment (placlitaxel)	23006423
2	64279606	rs10171434	C	T	A/A,A/A	0.92	NA	intergenic	8.324	Suicide attempts in bipolar disorder	21041247
2	64279606	rs10171434	C	T	A/A,A/A	0.92	NA	intergenic	8.324	Urinary metabolites	21572414
2	144783214	rs16823411	T	C	A/A,A/A	0.93	GTDC1	intron	4.096	Body mass index	21701565
2	144783214	rs16823411	T	C	A/A,A/A	0.93	GTDC1	intron	4.096	Body mass index	21701565
2	145213638	rs731108	G	C	A/A,H/H	0.92	ZEB2	intron,nc	12.16	Renal cell carcinoma	23184150
2	156506516	rs4407211	C	T	A/A,A/A	0.92	NA	intergenic	2.077	Alcohol consumption	23953852
3	51142359	rs4286453	T	C	A/A,A/A	0.91	DOCK3	intron	2.344	Multiple complex diseases	17554300
3	51824167	rs6796373	G	C	A/A,A/A	0.94	NA	intergenic	2.285	Response to taxane treatment (placlitaxel)	23006423
4	13325741	rs2867467	G	C	A/A,A/A	0.91	NA	intergenic	0.56	Obesity (extreme)	21935397
4	13328373	rs6842438	T	C	A/A,A/A	0.92	NA	intergenic	3.609	Obesity (extreme)	21935397
4	13330095	rs10019897	C	T	A/A,A/A	0.92	NA	intergenic	0.303	Multiple complex diseases	17554300
4	13330095	rs10019897	C	T	A/A,A/A	0.92	NA	intergenic	0.303	Obesity (extreme)	21935397
4	13333413	rs9996364	A	G	A/A,A/A	0.92	HSP90AB2P	upstream	4.041	Obesity (extreme)	21935397
4	13338465	rs11945340	C	T	A/A,A/A	0.92	HSP90AB2P	intron,nc	10.31	Obesity (extreme)	21935397
4	13340249	rs6839621	T	C	A/A,A/A	0.92	HSP90AB2P	non coding exon,nc	0.873	Obesity (extreme)	21935397
4	13346602	rs11930614	C	T	A/A,A/A	0.92	NA	intergenic	0.22	Obesity (extreme)	21935397
4	13350973	rs10021881	T	C	A/A,A/A	0.92	NA	regulatory	3.346	Obesity (extreme)	21935397
4	13356393	rs16888596	G	A	A/A,A/A	0.94	NA	intergenic	1.347	Obesity (extreme)	21935397
4	13357274	rs11732938	A	G	A/A,A/A	0.94	NA	intergenic	20	Obesity (extreme)	21935397
4	13360622	rs11947529	T	A	A/A,A/A	0.93	RAB28	downstream	4.509	Obesity (extreme)	21935397
4	13363958	rs12331157	A	G	A/A,A/A	0.97	RAB28	intron	1.536	Obesity (extreme)	21935397
4	13363974	rs12332023	C	T	A/A,A/A	0.97	RAB28	intron	0.363	Obesity (extreme)	21935397
4	13366481	rs7673680	C	T	A/A,A/A	0.93	RAB28	intron	3.083	Obesity (extreme)	21935397
4	13370308	rs10003958	T	C	A/A,A/A	0.93	RAB28	intron	14.23	Obesity (extreme)	21935397
4	13373583	rs9999851	C	T	A/A,A/A	0.97	RAB28	intron	0.402	Obesity (extreme)	21935397
4	13374462	rs9291610	G	A	A/A,A/A	0.93	RAB28	intron	0.826	Obesity (extreme)	21935397
4	13393897	rs9998914	A	T	A/A,A/A	0.96	RAB28	intron	2.579	Obesity (extreme)	21935397
4	13403855	rs11943295	G	A	A/A,A/A	0.94	RAB28	intron	0.842	Multiple complex diseases	17554300
4	13403855	rs11943295	G	A	A/A,A/A	0.94	RAB28	intron	0.842	Obesity (extreme)	21935397
4	13403998	rs11943330	G	A	A/A,A/A	0.93	RAB28	intron	1.179	Obesity (extreme)	21935397
4	13404130	rs7677336	G	T	A/A,A/A	0.94	RAB28	intron	0.385	Obesity (extreme)	21935397
4	13404717	rs7673732	A	C	A/A,A/A	0.93	RAB28	intron	1.116	Obesity (extreme)	21935397
4	13440031	rs11737264	C	G	A/A,A/A	0.93	RAB28	intron	0.138	Obesity (extreme)	21935397
4	13440271	rs11737360	C	T	A/A,A/A	0.94	RAB28	intron	0.54	Obesity (extreme)	21935397
4	13449532	rs16888654	A	C	A/A,A/A	0.94	RAB28	intron	0.905	Obesity (extreme)	21935397
4	13452022	rs16888661	C	A	A/A,A/A	0.91	RAB28	intron	3.789	Obesity (extreme)	21935397
4	13463991	rs11933841	T	C	A/A,A/A	0.93	RAB28	intron	3.377	Obesity (extreme)	21935397
4	13465710	rs11947665	T	A	A/A,A/A	0.93	RAB28	intron	1.709	Obesity (extreme)	21935397
4	23095293	rs6825402	C	T	A/A,A/A	0.96	NA	intergenic	0.797	Multiple complex diseases	17554300
5	89540468	rs2935504	C	T	A/A,A/A	0.97	RP11-61G23.1	intron,nc	3.627	Multiple complex diseases	17554300
6	136947540	rs17723608	A	G	A/A,A/A	0.93	MAP3K5	intron	0.586	Blood pressure, CVD RF and other traits	20877124
7	72746648	rs6943090	C	T	A/A,A/A	0.97	TRIM50	upstream	1.88	Immune reponse to smallpox (secreted IL-12p40)	22610502
7	106720932	rs12154324	G	A	A/A,A/A	0.93	NA	regulatory	3.447	Multiple complex diseases	17554300
7	116668662	rs4730767	C	T	A/A,A/A	0.93	ST7-OT4	intron,nc	8.279	Response to gemcitabine or arabinosylcytosin in blood cell lines	19898621
7	116668662	rs4730767	C	T	A/A,A/A	0.93	ST7-OT4	intron,nc	8.279	Response to gemcitabine or arabinosylcytosin in blood cell lines	19898621
10	37579117	rs7096155	A	C	A/A,A/A	0.94	ATP8A2P1	intron,nc	2.346	Multiple complex diseases	17554300
10	37579117	rs7096155	A	C	A/A,A/A	0.94	ATP8A2P1	intron,nc	2.346	Multiple complex diseases	17554300
12	56308562	rs772464	G	T	A/A,A/A	0.96	NA	regulatory	1.192	Multiple complex diseases	17554300
12	72889122	rs17111252	A	T	A/A,A/A	0.93	TRHDE	intron	4.133	Multiple complex diseases	17554300
13	35811439	rs10129134	C	T	A/A,A/A	0.93	NBEA	intron	3.514	Body mass index	22446040
16	61340362	rs9922966	G	C	A/A,A/A	0.93	NA	intergenic	4.37	Multiple complex diseases	17554300
22	34557399	rs5999230	T	G	A/A,A/A	0.93	LL22NC03-86D4.1	intron,nc	1.126	HIV-1 viral setpoint	22174851

Table S8. Overlap between GWAS catalog and catalog of modern human-specific high-frequency changes in the top modern human selected regions (1 cM scan). Chr = chromosome. Pos = position (hg19). ID = SNP rs ID. Hum = Present-day human major allele. Anc = Human-Chimpanzee ancestor allele. Arch = Archaic human allele states (Altai Neanderthal, Denisova) where H=human-like allele and A=ancestral allele. Freq = present-day human derived frequency. Cons = consequence. C = C-score. PubMed = PubMed article ID for GWAS study.

Chr	Pos	ID	Hum	Anc	Arch	Freq	Gene	Cons	C	GWAS trait	PubMed
1	27138393	rs12748152	C	T	A/A,A/A	0.95	Metazoa SRP	upstream	4.193	HDL cholesterol	24097068
1	27138393	rs12748152	C	T	A/A,A/A	0.95	Metazoa SRP	upstream	4.193	LDL cholesterol	24097068
1	27138393	rs12748152	C	T	A/A,A/A	0.95	Metazoa SRP	upstream	4.193	Triglycerides	24097068
2	64279606	rs10171434	C	T	A/A,A/A	0.92	NA	intergenic	8.324	Suicide attempts in bipolar disorder	21041247
2	64279606	rs10171434	C	T	A/A,A/A	0.92	NA	intergenic	8.324	Urinary metabolites	21572414
2	144783214	rs16823411	T	C	A/A,A/A	0.93	GTDC1	intron	4.096	Body mass index	21701565
2	144783214	rs16823411	T	C	A/A,A/A	0.93	GTDC1	intron	4.096	Body mass index	21701565
2	145213638	rs731108	G	C	A/A,H/H	0.92	ZEB2	intron,nc	12.16	Renal cell carcinoma	23184150
2	156506516	rs4407211	C	T	A/A,A/A	0.92	NA	intergenic	2.077	Alcohol consumption	23953852
3	51142359	rs4286453	T	C	A/A,A/A	0.91	DOCK3	intron	2.344	Multiple complex diseases	17554300
3	51824167	rs6796373	G	C	A/A,A/A	0.94	NA	intergenic	2.285	Response to taxane treatment (placlitaxel)	23006423
3	52506426	rs6784615	T	C	A/A,A/A	0.96	NA	regulatory	0.316	Waist-hip ratio	20935629
5	54558972	rs897669	G	A	A/A,A/A	0.92	DHX29	intron	5.673	Alcohol and nicotine co-dependence	20158304
7	106720932	rs12154324	G	A	A/A,A/A	0.93	NA	regulatory	3.447	Multiple complex diseases	17554300
7	116668662	rs4730767	C	T	A/A,A/A	0.93	ST7-OT4	intron,nc	8.279	Response to gemcitabine or arabinosylcytosin in blood cell lines	19898621
7	116668662	rs4730767	C	T	A/A,A/A	0.93	ST7-OT4	intron,nc	8.279	Response to gemcitabine or arabinosylcytosin in blood cell lines	19898621
10	37579117	rs7096155	A	C	A/A,A/A	0.94	ATP8A2P1	intron,nc	2.346	Multiple complex diseases	17554300
12	79387804	rs17046168	C	T	A/A,A/A	0.92	RP11-390N6.1	intron,nc	2.716	Response to taxane treatment (placlitaxel)	23006423
15	42527218	rs2620387	C	A	A/A,A/A	0.91	TMEM87A	intron	10.12	Multiple complex diseases	17554300