

1 Testing for ancient selection using cross-population allele 2 frequency differentiation

3 Fernando Racimo^{1,*}

4 **1 Department of Integrative Biology, University of California, Berkeley, CA, USA**

5 * **E-mail: fernandoracimo@gmail.com**

6 **1 Abstract**

7 A powerful way to detect selection in a population is by modeling local allele frequency changes in a
8 particular region of the genome under scenarios of selection and neutrality, and finding which model is
9 most compatible with the data. Chen et al. [1] developed a composite likelihood method called XP-CLR
10 that uses an outgroup population to detect departures from neutrality which could be compatible with
11 hard or soft sweeps, at linked sites near a beneficial allele. However, this method is most sensitive to recent
12 selection and may miss selective events that happened a long time ago. To overcome this, we developed
13 an extension of XP-CLR that jointly models the behavior of a selected allele in a three-population tree.
14 Our method - called 3P-CLR - outperforms XP-CLR when testing for selection that occurred before two
15 populations split from each other, and can distinguish between those events and events that occurred
16 specifically in each of the populations after the split. We applied our new test to population genomic
17 data from the 1000 Genomes Project, to search for selective sweeps that occurred before the split of
18 Africans and Eurasians, but after their split from Neanderthals, and that could have presumably led to
19 the fixation of modern-human-specific phenotypes. We also searched for sweep events that occurred in
20 East Asians, Europeans and the ancestors of both populations, after their split from Africans.

21 **2 Introduction**

22 Genetic hitchhiking will distort allele frequency patterns at regions of the genome linked to a beneficial
23 allele that is rising in frequency [2]. This is known as a selective sweep. If the sweep is restricted to a
24 particular population and does not affect other closely related populations, one can detect such an event
25 by looking for extreme patterns of localized population differentiation, like high values of F_{st} at a specific
26 locus [3]. This and other related statistics have been used to scan the genomes of present-day humans

27 from different populations, so as to detect signals of recent positive selection [4–7].

28 Once it became possible to sequence entire genomes of archaic humans (like Neanderthals) [8–10],
29 researchers also began to search for selective sweeps that occurred in the ancestral population of all
30 present-day humans. For example, ref. [8] searched for genomic regions with a depletion of derived
31 alleles in a low-coverage Neanderthal genome, relative to what would be expected given the derived allele
32 frequency in present-day humans. This is a pattern that would be consistent with a sweep in present-
33 day humans. Later on, ref. [10] developed a hidden Markov model (HMM) that could identify regions
34 where Neanderthals fall outside of all present-day human variation (also called “external regions”), and
35 are therefore likely to have been affected by ancient sweeps in early modern humans. They applied
36 their method to a high-coverage Neanderthal genome. Then, they ranked these regions by their genetic
37 length, to find segments that were extremely long, and therefore highly compatible with a selective sweep.
38 Finally, ref. [11] used summary statistics calculated in the neighborhood of sites that were ancestral in
39 archaic humans but fixed derived in all or almost all present-day humans, to test if any of these sites
40 could be compatible with a selective sweep model. While these methods harnessed different summaries
41 of the patterns of differentiation left by sweeps, they did not attempt to explicitly model the process by
42 which these patterns are generated over time.

43 Chen et al. [1] developed a method called XP-CLR, which is designed to test for selection in one
44 population after its split from a second, outgroup, population t_{AB} generations ago. It does so by modeling
45 the evolutionary trajectory of an allele under linked selection and under neutrality, and then comparing
46 the likelihood of the data under each of the two models. The method detects local allele frequency
47 differences that are compatible with the linked selection model [2], along windows of the genome.

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

56 To overcome this, we developed an extension of XP-CLR that jointly models the behavior of an allele

57 in all 3 populations, to detect selective events that occurred before or after the closest two populations
58 split from each other. Below we briefly review the modeling framework of XP-CLR and describe our new
59 test, which we call 3P-CLR. In the Results, we show this method outperforms XP-CLR when testing for
60 selection that occurred before the split of two populations, and can distinguish between those events and
61 events that occurred after the split, unlike XP-CLR. We then apply the method to population genomic
62 data from the 1000 Genomes Project [12], to search for selective sweep events that occurred before the
63 split of Africans and Eurasians, but after their split from Neanderthals. We also use it to search for
64 selective sweeps that occurred in the Eurasian ancestral population, and to distinguish those from events
65 that occurred specifically in East Asians or specifically in Europeans.

66 3 Methods

67 3.1 XP-CLR

68 First, we review the procedure used by XP-CLR to model the evolution of allele frequency changes of
69 two populations a and b that split from each other t_{AB} generations ago (Figure 1.A). For neutral SNPs,
70 Chen et al. [1] use an approximation to the Wright-Fisher diffusion dynamics [13]. Namely, the frequency
71 of a SNP in a population a (p_A) in the present is treated as a random variable governed by a normal
72 distribution with mean equal to the frequency in the ancestral population (β) and variance proportional
73 to the drift time ω from the ancestral to the present population:

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

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

75 If a SNP is segregating in both populations - i.e. has not hit the boundaries of fixation or extinction
76 - this process is time-reversible. Thus, one can model the frequency of the SNP in population a with a
77 normal distribution having mean equal to the frequency in population b and variance proportional to the
78 sum of the drift time (ω) between a and the ancestral population, and the drift time between b and the
79 ancestral population (ψ):

$$p_A|p_B \sim N(p_B, (\omega + \psi)p_B(1 - p_B)) \quad (2)$$

80 For SNPs that are linked to a beneficial allele that has undergone a sweep in population a only, Chen
 81 et al. [1] model the allele as evolving neutrally until the present and then apply a transformation to the
 82 normal distribution that depends on the distance to the selected allele r and the strength of selection
 83 s [14, 15]. Let $c = 1 - q_0^{r/2}$ where q_0 is the frequency of the beneficial allele in population a before the
 84 sweep begins. The frequency of a neutral allele is expected to increase from p to $1 - c + cp$ if the allele
 85 is linked to the beneficial allele, and this occurs with probability equal to the frequency of the neutral
 86 allele (p) before the sweep begins. Otherwise, the frequency of the neutral allele is expected to decrease
 87 from p to cp . This leads to the following transformation of the normal distribution:

$$f(p_A|p_B, r, s, \omega, \psi) = \frac{1}{\sqrt{2\pi}\sigma} \frac{p_A + c - 1}{c^2} e^{-\frac{(p_A + c - 1 - cp_B)^2}{2c^2\sigma^2}} I_{[1-c, 1]}(p_A) + \frac{1}{\sqrt{2\pi}\sigma} \frac{c - p_A}{c^2} e^{-\frac{(p_A - cp_B)^2}{2c^2\sigma^2}} I_{[0, c]}(p_A) \quad (3)$$

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

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

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

96 We note that the denominator in the above equation is not explicitly stated in ref. [1] for ease of
 97 notation, but appears in the published online implementation of the method. Because we are ignoring
 98 the correlation in frequencies produced by linkage, this is a composite likelihood [16, 17]. Finally, we
 99 obtain a composite likelihood ratio statistic S_{XP-CLR} of the hypothesis of linked selection over the
 100 hypothesis of neutrality:

$$S_{XP-CLR} = 2[\sup_{\mathbf{r}, \mathbf{v}, s} \log(CL_{XP-CLR}(\mathbf{r}, \mathbf{v}, s)) - \sup_{\mathbf{v}} \log(CL_{XP-CLR}(\mathbf{r}, \mathbf{v}, s = 0))] \quad (5)$$

101 For ease of computation, Chen et al. [1] assume that \mathbf{r} is given (via a recombination map) and we
 102 will do so too. Furthermore, they empirically estimate \mathbf{v} using F_2 statistics [18] calculated over the
 103 whole genome, and assume selection is not strong or frequent enough to affect their genome-wide values.
 104 Because we are interested in selection over long time scales, the new methods we will present below
 105 are optimally run using drift times calculated from population split times and effective population sizes
 106 estimated using model-based demographic inference methods, like $\partial a \partial i$ [19] or fastsimcoal2 [20].

107 3.2 3P-CLR

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

$$111 \quad p_A | \beta, \omega \sim N(\beta, \omega \beta(1 - \beta)) \quad (6)$$

$$112 \quad p_B | \beta, \psi \sim N(\beta, \psi \beta(1 - \beta)) \quad (7)$$

112 Let χ be the drift time separating the most recent common ancestor of a and b from the most recent
 113 common ancestor of a , b and c . Additionally, let ν be the drift time separating population c in the
 114 present from the most recent common ancestor of a , b and c . Given these parameters, we can treat β
 115 as an additional random variable that either evolves neutrally or is linked to a selected allele that swept
 116 immediately more anciently than the split of a and b . In both cases, the distribution of β will depend on
 117 the frequency of the allele in population c (p_C) in the present. In the neutral case:

$$118 \quad f_{neut}(\beta | p_C, \nu, \chi) = N(p_C, (\nu + \chi)p_C(1 - p_C)) \quad (8)$$

118 In the linked selection case:

$$119 \quad f_{sel}(\beta | p_C, \nu, \chi, r, s) = \frac{1}{\sqrt{2\pi\kappa}} \frac{\beta + c - 1}{c^2} e^{-\frac{(\beta + c - 1 - cp_C)^2}{2c^2\kappa^2}} I_{[1-c,1]}(\beta) + \frac{1}{\sqrt{2\pi\kappa}} \frac{c - \beta}{c^2} e^{-\frac{(\beta - cp_C)^2}{2c^2\kappa^2}} I_{[0,c]}(\beta) \quad (9)$$

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

120 The frequencies in a and b given the frequency in c can be obtained by integrating β out. This leads

121 to a density function that models selection in the ancestral population of a and b .

$$f(p_A, p_B | p_C, \mathbf{v}, r, s) = \int_0^1 f_{neut}(p_A | \beta, \omega) f_{neut}(p_B | \beta, \psi) f_{sel}(\beta | p_C, \nu, \chi, r, s) d\beta \quad (10)$$

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

$$f(p_A, p_B | p_C, \mathbf{v}, r, s) = \int_0^1 f_{sel}(p_A | \beta, \omega, r, s) f_{neut}(p_B | \beta, \psi) f_{neut}(\beta | p_C, \nu, \chi) d\beta \quad (11)$$

126 We will henceforth refer to the version of 3P-CLR that is tailored to detect selection in the internal
 127 branch that is ancestral to a and b as 3P-CLR(Int). In turn, the versions of 3P-CLR that are designed
 128 to detect selection in each of the daughter populations a and b will be designated as 3P-CLR(A) and
 129 3P-CLR(B), respectively.

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

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

132 and

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

133 where $B(x, y)$ is the Beta function.

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

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

138 where

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

139 and

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

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

$$CL_{3P-CLR}(\mathbf{r}, \mathbf{v}, s) = \prod_{j=1}^k P(m_A^j, m_B^j | m_C^j, \mathbf{v}, r^j, s) \quad (17)$$

142 As before, we can use this composite likelihood to produce a composite likelihood ratio statistic
143 that can be calculated over regions of the genome to test the hypothesis of linked selection centered on a
144 particular locus against the hypothesis of neutrality. Due to computational costs in numerical integration,
145 we skip the sampling step for population c (formula 13) in our implementation of 3P-CLR. In other
146 words, we assume $p_C = m_C/n_C$, but this is also assumed in XP-CLR when computing its corresponding
147 outgroup frequency. To perform the numerical integrations, we used the package Cubature (v.1.0.2). We
148 implemented our method in a freely available C++ program that can be downloaded from here:

149 <https://github.com/ferracimo> [WILL POST IT AFTER PUBLICATION]

150 4 Results

151 4.1 Simulations

152 We generated simulations in SLiM [21] to test the performance of XP-CLR and 3P-CLR in a three-
153 population scenario. We first focused on the performance of 3P-CLR(Int) in detecting ancient selective
154 events that occurred in the ancestral branch of two sister populations. We assumed that the population
155 history had been correctly estimated by the researcher (i.e. the drift parameters and population topology
156 were known). First, we simulated scenarios in which a beneficial mutation arose in the ancestor of
157 populations a and b , before their split from each other but after their split from c (Table 1). Although

158 both XP-CLR and 3P-CLR are sensitive to partial or soft sweeps (as they do not rely on extended
159 patterns of haplotype homozygosity [1]), we required the allele to have fixed before the split (at time t_{ab})
160 to ensure that the allele had not been lost before it, and also to ensure that the sweep was restricted to
161 the internal branch of the tree. We fixed the effective size of all three populations at $N_e = 10,000$. Each
162 simulation consisted in a 5 cM region and the beneficial mutation occurred in the center of this region.
163 The mutation rate was set at $2.5 * 10^{-8}$ per generation and the recombination rate was set at 10^{-8} per
164 generation.

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

174 Figure 2 shows receiver operating characteristic (ROC) curves comparing the sensitivity and specificity
175 of 3P-CLR(Int), 3P-CLR(A), XP-CLR and XP-CLR-avg in the first six demographic scenarios described
176 in Table 1. Each ROC curve was made from 100 simulations under selection (with $s = 0.1$ for the central
177 mutation) and 100 simulations under neutrality (with $s = 0$ and no fixation required). In each simulation,
178 100 haploid individuals (or 50 diploids) were sampled from population a , 100 individuals from population
179 b and 100 individuals from the outgroup population c . For each simulation, we took the maximum value
180 at a region in the neighborhood of the central mutation (± 0.5 cM) and used those values to compute
181 ROC curves under the two models.

182 When the split times are recent or moderately ancient (models A to D), 3P-CLR(Int) outperforms
183 the two versions of XP-CLR. Furthermore, 3P-CLR(A) is the test that is least sensitive to selection in
184 the internal branch as it is only meant to detect selection in the terminal branch leading to population
185 a . When the split times are very ancient (models E and F), none of the tests perform well. The root
186 mean squared error (RMSE) of the genetic distance between the true selected site and the highest scored
187 window is comparable across tests in all six scenarios (Figure S2). 3P-CLR(Int) is the best test at finding

188 the true location of the selected site in almost all demographic scenarios.

189 We also simulated a situation in which only a few individuals (e.g. a small sample of archaic humans)
190 have been sequenced from the outgroup, while large numbers of sequences are available from the tests
191 (e.g. two populations of present-day humans). Figures S1 and S3 show the ROC curves and RMSE plots,
192 respectively, for a scenario in which 100 individuals were sampled from the test populations but only
193 10 individuals (5 diploids) were sampled from the outgroup. Unsurprisingly, all tests have less power to
194 detect selection when the split times and the selection events are recent to moderately ancient (models
195 A-D). Interestingly though, when the split times and the selective events are very ancient (models E-F),
196 both 3P-CLR and XP-CLR perform better when using a small outgroup panel (Figure S1) than when
197 using a large outgroup panel (Figure 2). This may be because both of these tests require the outgroup
198 sample at each site to be a segregating polymorphism, and sites that are polymorphic in a small panel
199 are, on average, more ancient than sites that are polymorphic in a large panel. Because the recent
200 polymorphisms in the outgroup carry little or no information about ancient selection in the tests, they
201 are less likely to contribute to differences in the likelihood functions for the selection and the neutrality
202 models, and so they make these tests less efficient at distinguishing these two models.

203 Importantly, the usefulness of 3P-CLR(Int) resides not just in its performance at detecting selective
204 sweeps in the ancestral population, but in its specific sensitivity to that particular type of events. Because
205 the test relies on correlated allele frequency differences in both population *a* and population *b* (relative
206 to the outgroup), selective sweeps that are specific to only one of the populations will not lead to high
207 3P-CLR(Int) scores, but will instead lead to high 3P-CLR(A) scores or 3P-CLR(B) scores, depending on
208 where selection took place. Figure 3 shows ROC curves in two scenarios in which a selective sweep event
209 occurred only in population *a* (models I and J in Table 1), using 100 sampled individuals from each of
210 the 3 populations. Here, XP-CLR performs well, but is outperformed by 3P-CLR(A). Furthermore, 3P-
211 CLR(Int) shows almost no sensitivity to the recent sweep. For example, in Model I, at a specificity of 90%,
212 3P-CLR(A) and XP-CLR(A) have 86% and 80% sensitivity, respectively, while at the same specificity,
213 3P-CLR(Int) only has 18% sensitivity. One can compare this to the same demographic scenario but with
214 selection occurring in the ancestral population of *a* and *b* (model C, Figure 2), where at 90% specificity,
215 3P-CLR(A) and XP-CLR(A) have 72% and 84% sensitivity, respectively, while 3P-CLR(Int) has 90%
216 sensitivity. We also observe that 3P-CLR(A) is the best test at finding the true location of the selected
217 site when selection occurs in the terminal branch leading to population *a* (Figure S4).

218 4.2 Selection in Eurasians

219 We first applied 3P-CLR to modern human data from the 1000 Genomes Project [12]. We used the
220 African-American recombination map [22] to convert physical distances into genetic distances. We fo-
221 cused on Europeans and East Asians as the two sister populations, using Africans (excluding admixed
222 African-Americans) as the outgroup population (Figure S5.A). We randomly sampled 100 individuals
223 from each population and obtained sample derived allele frequencies every 10 SNPs in the genome. We
224 then calculated likelihood ratio statistics by a sliding window approach, where we sampled a "central
225 SNP" once every 20 SNPs. The central SNP in each window was the candidate beneficial SNP for that
226 window. We set the window size to 0.25 cM, and randomly sampled 100 SNPs from each window, cen-
227 tered around the candidate beneficial SNP. In each window, we calculated 3P-CLR to test for selection
228 at three different branches of the population tree: the terminal branch leading to Europeans (3P-CLR
229 Europe), the terminal branch leading to East Asians (3P-CLR East Asia) and the ancestral branch of
230 Europeans and East Asians (3P-CLR Eurasia). Results are shown in Figure 4. For each scan, we selected
231 the windows in the top 99.9% quantile of scores and merged them together if they were contiguous. Tables
232 2, 3 and 4 show the top hits for Europeans, East Asians and the ancestral Eurasian branch, respectively

233 We observe several genes that have been identified in previous selection scans. In the East Asian
234 branch, one of the top hits is *EDAR*. This gene codes for a protein involved in hair thickness and incisor
235 tooth morphology [23,24]. It has been repeatedly identified in earlier selections scans as having undergone
236 a sweep in East Asians [25,26].

237 Furthermore, 3P-CLR allows us to narrow down on the specific time at which selection occurred in
238 the history of particular populations. For example, ref. [1] performed a scan of the genomes of East
239 Asians using XP-CLR with Africans as the outgroup, and identified a number of genes as being under
240 selection [1]. 3P-CLR confirms this signal in several of these loci when looking specifically at the East
241 Asian branch: *CYP26B1*, *EMX1*, *SPR*, *SFXN5*, *PPARA*, *PKDREJ*, *GTSE1*, *TRMU*, *CELSR1*, *PINX1*,
242 *XKR6*, *CD226*, *ACD*, *PARD6A*, *GFOD2*, *RANBP10*, *TSNAXIP1*, *CENPT*, *THAP11*, *NUTF2*, *CDH16*,
243 *RRAD*, *FAM96B*, *CES2*, *CBFB*, *C16orf70*, *TRADD*, *FBXL8*, *HSF4*, *NOL3*, *EXOC3L1*, *E2F4*, *ELMO3*,
244 *LRRC29*, *FHOD1*, *SLC9A5*, *PLEKHG4*, *LRRC36*, *ZDHHC1*, *HSD11B2*, *ATP6V0D1*, *AGRP*, *FAM65A*,
245 *CTCF* and *RLTPR*. However, when applied to the ancestral Eurasian branch, 3P-CLR finds some genes
246 that were previously found in the XP-CLR analysis of East Asians, but that are not among the top hits
247 in 3P-CLR applied to the East Asian branch: *COMMD3*, *BMI1*, *SPAG6* and *ABCC12*. This suggests

248 selection in these regions occurred earlier, i.e. before the European-East Asian split. Figure 5.A shows
249 a comparison between the 3P-CLR scores for the three branches in the region containing genes *BMI1*
250 (a proto-oncogene [27]) and *SPAG6* (involved in sperm motility [28]). In that figure, the score within
251 each window was standardized using its chromosome-wide mean and standard deviation, to make a fair
252 comparison. One can observe that the signal of Eurasia-specific selection is evidently stronger than the
253 other two signals.

254 When running 3P-CLR to look for selection specific to Europe, we find that *TYRP1* (Figure 5.B) and
255 *MYO5A*, which play a role in human skin pigmentation [29–32], are among the top hits. Both of these
256 genes have been previously found to be under strong selection in Europe [33], using a statistic called
257 iHS, which measures extended patterns of haplotype homozygosity that are characteristic of selective
258 sweeps. Interestingly, a change in the gene *TYRP1* has also been found to cause a blonde hair phenotype
259 in Melanesians [34]. Another of our top hits is the region containing *SH2B3*, which was identified
260 previously as a candidate for selection in Europe based on both iHS and Fst [35]. This gene contains a
261 nonsynonymous SNP (rs3184504) segregating in Europeans. One of its alleles (the one in the selected
262 haplotype) has been associated with celiac disease and type 1 diabetes [36, 37] but is also protective
263 against bacterial infection [38].

264 We used Gowinda (v1.12) [39] to find enriched Gene Ontology (GO) categories among the regions
265 in the 99% highest quantile for each branch score, relative to the rest of the genome ($P < 0.05$, FDR
266 < 0.2). The significantly enriched categories are listed in Table 5. In the East Asian branch, we find
267 categories related to pyruvate metabolism, cholesterol absorption and peroxisome proliferation. In the
268 European branch, we find categories related to cuticle development, antioxidant activity and thyroid
269 hormone generation, among others. In the Eurasian branch, we only find two categories that are related
270 to the regulation of transcription by sequence-specific DNA-binding.

271 4.3 Selection in ancestral modern humans

272 We applied 3P-CLR to modern human data combined with recently sequenced archaic human data
273 [10]. We sought to find selective events that occurred in modern humans after their split from archaic
274 groups. We used the combined Neanderthal and Denisovan high-coverage genomes [9, 10] as the outgroup
275 population, and, for our two test populations, we randomly sampled 100 Eurasian genomes and 100
276 African genomes (excluding admixed African-Americans) from the 1000 Genomes data (Figure S5.B).

277 We used previously estimated drift times as fixed parameters [10], and tested for selective events that
278 occurred more anciently than the split of Africans and Eurasians, but more recently than the split from
279 Neanderthals. We ran 3P-CLR using 0.25 cM windows as above, but also verified that the density of
280 scores was robust to the choice of window size and spacing (Figure S6). As before, we selected the top
281 99.9% windows and merged them together if they were contiguous. Table 6 and Figure S7 show the
282 top hits. To find putative candidates for the beneficial variants in each region, we queried the catalogs
283 of modern human-specific high-frequency or fixed derived changes that are ancestral in the Neanderthal
284 and/or the Denisova genomes [10, 40].

285 We observe several genes that have been identified in previous scans that looked for selection in modern
286 humans after their split from archaic groups [8, 10]: *SIPA1L1*, *ANAPC10*, *ABCE1*, *RASA1*, *CCNH*,
287 *KCNJ3*, *HBP1*, *COG5*, *GPR22*, *DUS4L*, *BCAP29*, *CADPS2*, *RNF133*, *RNF148*, *FAM172A*, *POU5F2*,
288 *FGF7*, *RABGAP1*, *GPR21*, *STRBP*, *SMURF1*, *GABRA2*, *ALMS1*, *PVRL3*, *EHBP1*, *VPS54*, *OTX1*,
289 *UGP2*, *HCN1*, *GTDC1*, *ZEB2*, *OIT3*, *USP54* and *MYOZ1*. One of our strongest candidate genes among
290 these is *ANAPC10*. This gene is a core subunit of the cyclosome, is involved in progression through the
291 cell cycle [41], and may play a role in oocyte maturation and human T-lymphotropic virus infection
292 (KEGG pathway [42]). *ANAPC10* is noteworthy because it was found to be significantly differentially
293 expressed in humans compared to other great apes and macaques: it is up-regulated in the testes [43].
294 The gene also contains two intronic changes that are fixed derived in modern humans, ancestral in both
295 Neanderthals and Denisovans and that have evidence for being highly disruptive, based on a composite
296 score that combines conservation and regulatory data (PHRED-scaled C-scores > 11 [10, 44]). The
297 changes, however, appear not to lie in any obvious regulatory region [45, 46].

298 We also find *ADSL* among the list of candidates. This gene is known to contain a nonsynonymous
299 change that is fixed in all present-day humans but homozygous ancestral in the Neanderthal genome,
300 the Denisova genome and two Neanderthal exomes [40] (Figure 6.A). It was previously identified as ly-
301 ing in a region with strong support for positive selection in modern humans, using summary statistics
302 implemented in an ABC method [11]. The gene is interesting because it is one of the members of the
303 Human Phenotype ontology category "aggression / hyperactivity" which is enriched for nonsynonymous
304 changes that occurred in the modern human lineage after the split from archaic humans [40, 47]. *ADSL*
305 codes for adenylosuccinase, an enzyme involved in purine metabolism [48]. A deficiency of adenylosuc-
306 cinase can lead to apraxia, speech deficits, delays in development and abnormal behavioral features, like

307 hyperactivity and excessive laughter [49]. The nonsynonymous mutation (A429V) is in the C-terminal
308 domain of the protein (Figure 6.B) and lies in a highly conserved position (primate PhastCons = 0.953;
309 GERP score = 5.67 [44, 50, 51]). The ancestral amino acid is conserved across the tetrapod phylogeny,
310 and the mutation is only three residues away from the most common causative SNP for severe adeny-
311 losuccinase deficiency [52–56]. The change has the highest probability of being disruptive to protein
312 function, out of all the nonsynonymous modern-human-specific changes that lie in the top-scoring regions
313 (C-score = 17.69). While *ADSL* is an interesting candidate and lies in the center of the inferred selected
314 region (Figure 6.A), there are other genes in the region too, including *TNRC6B* and *MKL1*. *TNRC6B*
315 may be involved in miRNA-guided gene silencing [57], while *MKL1* may play a role in smooth muscle
316 differentiation [58], and has been associated with acute megakaryocytic leukemia [59].

317 *RASA1* was also a top hit in a previous scan for selection [8], and was additionally inferred to
318 have a high Bayes factor in favor of selection in ref. [11]. The gene codes for a protein involved in the
319 control of cellular differentiation [60], and has a modern human-specific fixed nonsynonymous change
320 (G70E). Human diseases associated with *RASA1* include basal cell carcinoma [61] and arteriovenous
321 malformation [62, 63].

322 The *GABA_A* gene cluster in chromosome 4p12 is also among the top regions. The genes within the
323 putatively selected region code for three of the subunits of the *GABA_A* receptor (*GABRA2*, *GABRA4*),
324 which codes for a ligand-gated ion channel that plays a key role in synaptic inhibition in the central
325 nervous system (see review by ref. [64]). *GABRA2* is significantly associated with the risk of alcohol
326 dependence in humans [65], perception of pain [66] and asthma [67]. In turn, *GABRA4* is associated with
327 autism risk [68, 69].

328 Two other candidate genes that may be involved in brain development are *FOXP1* and *CADPS2*.
329 *FOXP1* was not identified in any of the previous selection scans, and codes for a protein called forkhead
330 box G1, which plays an important role during brain development. Mutations in this gene have been
331 associated with a slow-down in brain growth during childhood resulting in microcephaly, which in turn
332 causes various intellectual disabilities [70, 71]. *CADPS2* was identified in [8] as a candidate for selection,
333 and has been associated with autism [72]. The gene has been suggested to be specifically important in
334 the evolution of all modern humans, as it was not found to be selected earlier in great apes or later in
335 particular modern human populations [73].

336 Finally, we find a signal of selection in a region containing the gene *EHBP1* and *OTX1*. This region

337 was identified in both of the two previous scans for modern human selection [8, 10]. *EHBPI* codes for a
338 protein involved in endocytic trafficking [74] and has been associated with prostate cancer [75]. *OTX1* is
339 a homeobox family gene that may play a role in brain development [76]. Interestingly, *EHBPI* contains
340 a single-nucleotide intronic change (chr2:63206488) that is almost fixed in all present-day humans and
341 homozygous ancestral in Neanderthal and Denisova [10]. This change is also predicted to be highly
342 disruptive (C-score = 13.1) and lies in a position that is extremely conserved across primates (PhastCons
343 = 0.942), mammals (PhastCons = 1) and vertebrates (PhastCons = 1). The change is 18 bp away
344 from the nearest splice site and overlaps a VISTA conserved enhancer region (element 1874) [77], which
345 suggests a putative regulatory role for the change.

346 We again used Gowinda [39] to find enriched GO categories among the regions with high 3P-CLR
347 scores in the Modern Human branch. This time, we used a stricter quantile cutoff (99.5%) to define
348 candidate regions than we did when running the program in the Eurasian tree (99%) because the less
349 strict cutoff yielded a very large number of enriched categories (114), though in both cases the enriched
350 terms were very similar. The significantly enriched categories ($P < 0.05$, $FDR < 0.2$) are listed in Table
351 7. We find several GO terms related to the regulation of the cell cycle, cell proliferation in the bone
352 marrow, lymphocyte chemotaxis and myeloid cell differentiation.

353 4.4 Modern human-specific high-frequency changes in GWAS catalog

354 We overlapped the genome-wide association studies (GWAS) database [78, 79] with the list of fixed or high-
355 frequency modern human-specific changes that are ancestral in archaic humans [10] and that are located
356 within our top putatively selected regions in modern humans (Table 8). None of the resulting SNPs
357 are completely fixed derived, because GWAS can only yield associations from sites that are segregating.
358 Among these SNPs, the one with the highest probability of being disruptive (rs10003958, C-score = 16.58,
359 Gerp score = 6.07) is located in a highly-conserved regulatory ("strong enhancer") region in the *RAB28*
360 gene [45, 46] (Primate PhastCons = 0.951), and is significantly associated with obesity [80] (Figure 7.A).
361 Interestingly, the region containing *RAB28* is inferred to have been under positive selection in both the
362 modern human and the Eurasian ancestral branches (Tables 4, 6). In line with this evidence, the derived
363 allele of rs10003958 is absent in archaic humans, at very high frequencies in Eurasians ($> 94\%$), and only
364 at moderately high frequencies in Africans (74%) (Figure 7.B).

365 We also find a highly disruptive SNP (rs10171434, C-score = 8.358) associated with urinary metabo-

366 lites [81] and suicidal behavior in patients with mood disorders [82]. The SNP is located in an enhancer
367 regulatory feature [45,46] located between genes *PEL11* and *VPS54*, in the same putatively selected
368 region as genes *EHBP1* and *OTX1* (see above). Finally, there is a highly disruptive SNP (rs731108, C-
369 score = 10.31) that is associated with renal cell carcinoma [83]. This SNP is also located in an enhancer
370 regulatory feature [45,46], in an intron of *ZEB2*. In this last case, though, only the Neanderthal genome
371 has the ancestral state, while the Denisova genome carries the modern human variant.

372 5 Discussion

373 We have developed a new method called 3P-CLR, which allows us to detect positive selection along
374 the genome. The method is based on an earlier test (XP-CLR [1]) that uses linked allele frequency
375 differences between two populations to detect population-specific selection. However, 3P-CLR can allow
376 us to distinguish between selective events that occurred before and after the split of two populations.
377 Our method also has some similarities to an earlier method developed by [84], which used an F_{st} -like
378 score to detect selection ancestral to two populations. In that case, though, the authors used summary
379 statistics and did not explicitly model the process leading to allele frequency differentiation.

380 We used our method to confirm previously found candidate genes in particular human populations, like
381 *EDAR*, *TYRP1*, *SH2B3* and *MYO5A*, and find some novel candidates too (Tables 2, 3, 4). Additionally,
382 we can infer that certain genes, which were previously known to have been under selection in East Asians
383 (like *SPAG6*), are more likely to have undergone a sweep in the population ancestral to both Europeans
384 and East Asians than in East Asians only. We find that genes involved in pyruvate and cholesterol
385 metabolism are particularly enriched among the East Asian candidate regions, which suggests these
386 biological functions may have been subject to positive selection in recent times.

387 We also used 3P-CLR to detect selective events that occurred in the ancestors of modern humans,
388 after their split from Neanderthals and Denisovans (Table 6). These events could perhaps have led to the
389 spread of phenotypes that set modern humans apart from other hominin groups. We find several interesting
390 candidates, like *SIPA1L1*, *ADSL*, *RASA1*, *OTX1*, *EHBP1*, *FOXP1*, *RAB28* and *ANAPC10*, some of
391 which were previously detected using other types of methods [8,10,11]. We also find an enrichment for
392 GO categories related to cell cycle regulation and lymphocyte chemotaxis among the candidate regions,
393 suggesting that these biological processes might have been affected by positive selection after the split

394 from archaic humans.

395 An advantage of differentiation-based tests like XP-CLR and 3P-CLR is that, unlike other patterns
396 detected by tests of neutrality (like extended haplotype homozygosity, [85]) that are exclusive to hard
397 sweeps, the patterns that both XP-CLR and 3P-CLR are tailored to find are based on regional allele
398 frequency differences between populations. These patterns can also be produced by soft sweeps from
399 standing variation or by partial sweeps [1], and there is some evidence that the latter phenomena may
400 have been more important than classic sweeps during human evolutionary history [86].

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

412 3P-CLR also has an advantage over HMM-based selection methods, like the one implemented in
413 ref. [10]. The likelihood ratio scores obtained from 3P-CLR can provide an idea of how credible a
414 selection model is for a particular region, relative to the rest of the genome. The HMM-based method
415 previously used to scan for selection in modern humans [10] can only rank putatively selected regions by
416 genetic distance, but cannot output a statistical measure that may indicate how likely each region is to
417 have been selected in ancient times. In contrast, 3P-CLR provides a composite likelihood ratio score,
418 which allows for a statistically rigorous way to compare the neutral model and a specific selection model
419 (for example, recent or ancient selection).

420 The score also gives an idea of how much fainter the signal of ancient selection in modern humans
421 is, relative to recent selection specific to a particular present-day population. For example, the outliers
422 from Figure 4 have much higher scores (relative to the rest of the genome) than the outliers from Figure
423 S7. This may be due to both the difference in time scales in the two sets of tests and to the uncertainty

424 that comes from estimating outgroup allele frequencies using only two archaic genomes. This pattern can
425 also be observed in Figure S8, where the densities of the scores looking for patterns of ancient selection
426 (3P-CLR Modern Human and 3P-CLR Eurasia) have much shorter tails than the densities of scores
427 looking for patterns of recent selection (3P-CLR Europe and 3P-CLR East Asia). Simulations show that
428 3P-CLR(Int) score distributions are naturally shorter than 3P-CLR(A) scores (Figure S9), which could
429 explain the short tail of the 3P-CLR Eurasia distribution. Additionally, the even shorter tail in the
430 distribution of 3P-CLR Modern Human scores may be a consequence of the fact that the split times of
431 the demographic history in that case are older than the split times in the Eurasian tree, as simulations
432 show that ancient split times tend to further shorten the tail of the 3P-CLR score distribution (Figure
433 S9).

434 A limitation of composite likelihood ratio tests is that the composite likelihood calculated for each
435 model under comparison is obtained from a product of individual likelihoods at each site, and so it
436 underestimates the correlation that exists between SNPs due to linkage effects [1, 16, 17, 87]. One way
437 to mitigate this problem is by using corrective weights based on linkage disequilibrium (LD) statistics
438 calculated on the outgroup population [1]. Our implementation of 3P-CLR allows the user to incorporate
439 such weights, if appropriate LD statistics are available from the outgroup. However, in cases where these
440 are unreliable, it may not be possible to fully correct for this (for example, when only a few unphased
441 genomes are available, as in the case of the Neanderthal and Denisova genomes).

442 While 3P-CLR relies on integrating over the possible allele frequencies in the ancestors of populations
443 a and b (formula 10), one could envision using ancient DNA to avoid this step. Thus, if enough genomes
444 could be sampled from that ancestral population that existed in the past, one could use the sample
445 frequency in the ancient set of genomes as a proxy for the ancestral population frequency. This may soon
446 be possible, as several early modern human genomes have already been sequenced in recent years [88–90].

447 Though we have limited ourselves to a three-population model in this manuscript, it should be straight-
448 forward to expand our model to a larger number of populations, albeit with additional costs in terms
449 of speed and memory. Our method relies on a similar framework to the demographic inference method
450 implemented in TreeMix [91], which can estimate complex population trees that include migration events,
451 using genome-wide data. With a more complex modeling framework, it may be possible to estimate the
452 time and strength of selective events with better resolution, and to incorporate additional demographic
453 forces, like continuous migration between populations or pulses of admixture.

454 Acknowledgments

455 We thank Montgomery Slatkin, Rasmus Nielsen, Joshua Schraiber, Nicolas Duforet-Frebourg, Emilia
456 Huerta-Sánchez, Hua Chen, Nick Patterson, David Reich, Joachim Hermisson, Graham Coop and mem-
457 bers of the Slatkin and Nielsen labs for helpful advice and discussions. This work was supported by NIH
458 grant R01-GM40282 to Montgomery Slatkin.

459 References

- 460 1. Chen H, Patterson N, Reich D (2010) Population differentiation as a test for selective sweeps.
461 *Genome research* 20: 393–402.
- 462 2. Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical research* 23:
463 23–35.
- 464 3. Lewontin R, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective
465 neutrality of polymorphisms. *Genetics* 74: 175–195.
- 466 4. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density snp map for
467 signatures of natural selection. *Genome research* 12: 1805–1814.
- 468 5. Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG (2005) Measures of human population
469 structure show heterogeneity among genomic regions. *Genome research* 15: 1468–1476.
- 470 6. Oleksyk TK, Zhao K, Francisco M, Gilbert DA, O’Brien SJ, et al. (2008) Identifying selected
471 regions from heterozygosity and divergence using a light-coverage genomic dataset from two human
472 populations. *PLoS One* 3: e1712.
- 473 7. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, et al. (2010) Sequencing of 50 human exomes
474 reveals adaptation to high altitude. *Science* 329: 75–78.
- 475 8. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, et al. (2010) A draft sequence of the
476 neandertal genome. *science* 328: 710–722.
- 477 9. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, et al. (2012) A high-coverage genome sequence
478 from an archaic denisovan individual. *Science* 338: 222–226.

- 479 10. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, et al. (2014) The complete genome
480 sequence of a neanderthal from the altai mountains. *Nature* 505: 43–49.
- 481 11. Racimo F, Kuhlwilm M, Slatkin M (2014) A test for ancient selective sweeps and an application
482 to candidate sites in modern humans. *Molecular biology and evolution* 31: 3344–3358.
- 483 12. Consortium GP, et al. (2012) An integrated map of genetic variation from 1,092 human genomes.
484 *Nature* 491: 56–65.
- 485 13. Nicholson G, Smith AV, Jónsson F, Gústafsson Ó, Stefánsson K, et al. (2002) Assessing popula-
486 tion differentiation and isolation from single-nucleotide polymorphism data. *Journal of the Royal*
487 *Statistical Society: Series B (Statistical Methodology)* 64: 695–715.
- 488 14. Durrett R, Schweinsberg J (2004) Approximating selective sweeps. *Theoretical population biology*
489 66: 129–138.
- 490 15. Fay JC, Wu CI (2000) Hitchhiking under positive darwinian selection. *Genetics* 155: 1405–1413.
- 491 16. Lindsay BG (1988) Composite likelihood methods. *Contemporary Mathematics* 80: 221–39.
- 492 17. Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. *Statistica Sinica*
493 21: 5–42.
- 494 18. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, et al. (2012) Ancient admixture in human
495 history. *Genetics* 192: 1065–1093.
- 496 19. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demo-
497 graphic history of multiple populations from multidimensional snp frequency data. *PLoS genetics*
498 5: e1000695.
- 499 20. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic infer-
500 ence from genomic and snp data. *PLoS genetics* 9: e1003905.
- 501 21. Messer PW (2013) Slim: simulating evolution with selection and linkage. *Genetics* 194: 1037–1039.
- 502 22. Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, et al. (2011) The landscape of recombina-
503 tion in african americans. *Nature* 476: 170–175.

- 504 23. Fujimoto A, Kimura R, Ohashi J, Omi K, Yuliwulandari R, et al. (2008) A scan for genetic
505 determinants of human hair morphology: Edar is associated with asian hair thickness. *Human*
506 *Molecular Genetics* 17: 835–843.
- 507 24. Kimura R, Yamaguchi T, Takeda M, Kondo O, Toma T, et al. (2009) A common variation in edar
508 is a genetic determinant of shovel-shaped incisors. *The American Journal of Human Genetics* 85:
509 528–535.
- 510 25. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, et al. (2007) Genome-wide detection and
511 characterization of positive selection in human populations. *Nature* 449: 913–918.
- 512 26. Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S, et al. (2010) A composite of
513 multiple signals distinguishes causal variants in regions of positive selection. *Science* 327: 883–
514 886.
- 515 27. Siddique HR, Saleem M (2012) Role of bmi1, a stem cell factor, in cancer recurrence and chemore-
516 sistance: preclinical and clinical evidences. *Stem Cells* 30: 372–378.
- 517 28. Sapiro R, Kostetskii I, Olds-Clarke P, Gerton GL, Radice GL, et al. (2002) Male infertility, impaired
518 sperm motility, and hydrocephalus in mice deficient in sperm-associated antigen 6. *Molecular and*
519 *cellular biology* 22: 6298–6305.
- 520 29. Pastural E, Ersoy F, Yalman N, Wulfraat N, Grillo E, et al. (2000) Two genes are responsible for
521 griscelli syndrome at the same 15q21 locus. *Genomics* 63: 299–306.
- 522 30. Fukuda M, Kuroda T, Mikoshiba K (2002) Slac2-a/melanophilin, the missing link between rab27
523 and myosin va: implications of a tripartite protein complex for melanosome transport. *The Journal*
524 *of biological chemistry* 277: 12432.
- 525 31. Halaban R, Moellmann G (1990) Murine and human b locus pigmentation genes encode a gly-
526 coprotein (gp75) with catalase activity. *Proceedings of the National Academy of Sciences* 87:
527 4809–4813.
- 528 32. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, et al. (2008) Two newly identified
529 genetic determinants of pigmentation in europeans. *Nature genetics* 40: 835–837.

- 530 33. Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the
531 human genome. *PLoS biology* 4: e72.
- 532 34. Kenny EE, Timpson NJ, Sikora M, Yee MC, Moreno-Estrada A, et al. (2012) Melanesian blond
533 hair is caused by an amino acid change in *tyrp1*. *Science* 336: 554–554.
- 534 35. Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, et al. (2009) Signals of recent positive
535 selection in a worldwide sample of human populations. *Genome research* 19: 826–837.
- 536 36. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, et al. (2007) Robust associations of
537 four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nature genetics* 39:
538 857–864.
- 539 37. Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, et al. (2008) Newly identified genetic risk
540 variants for celiac disease related to the immune response. *Nature genetics* 40: 395–402.
- 541 38. Zhernakova A, Elbers CC, Ferwerda B, Romanos J, Trynka G, et al. (2010) Evolutionary and
542 functional analysis of celiac risk loci reveals *sh2b3* as a protective factor against bacterial infection.
543 *The American Journal of Human Genetics* 86: 970–977.
- 544 39. Kofler R, Schlötterer C (2012) Gowinda: unbiased analysis of gene set enrichment for genome-wide
545 association studies. *Bioinformatics* 28: 2084–2085.
- 546 40. Castellano S, Parra G, Sánchez-Quinto FA, Racimo F, Kuhlwilm M, et al. (2014) Patterns of coding
547 variation in the complete exomes of three neandertals. *Proceedings of the National Academy of*
548 *Sciences* 111: 6666–6671.
- 549 41. Pravtcheva DD, Wise TL (2001) Disruption of *apc10/doc1* in three alleles of oligosyndactylism.
550 *Genomics* 72: 78–87.
- 551 42. Kanehisa M, Goto S (2000) Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*
552 28: 27–30.
- 553 43. Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, et al. (2011) The evolution of gene
554 expression levels in mammalian organs. *Nature* 478: 343–348.

- 555 44. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, et al. (2014) A general framework for
556 estimating the relative pathogenicity of human genetic variants. *Nature genetics* 46: 310–315.
- 557 45. Consortium EP, et al. (2012) An integrated encyclopedia of dna elements in the human genome.
558 *Nature* 489: 57–74.
- 559 46. Rosenbloom KR, Dreszer TR, Long JC, Malladi VS, Sloan CA, et al. (2011) Encode whole-genome
560 data in the ucsc genome browser: update 2012. *Nucleic acids research* : gkr1012.
- 561 47. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, et al. (2008) The human phenotype ontology:
562 a tool for annotating and analyzing human hereditary disease. *The American Journal of Human*
563 *Genetics* 83: 610–615.
- 564 48. Van Keuren M, Hart I, Kao FT, Neve R, Bruns G, et al. (1987) A somatic cell hybrid with a single
565 human chromosome 22 corrects the defect in the cho mutant (ade-i) lacking adenylosuccinase
566 activity. *Cytogenetic and Genome Research* 44: 142–147.
- 567 49. Gitiaux C, Ceballos-Picot I, Marie S, Valayannopoulos V, Rio M, et al. (2009) Misleading be-
568 havioural phenotype with adenylosuccinate lyase deficiency. *European Journal of Human Genetics*
569 17: 133–136.
- 570 50. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, et al. (2005) Evolutionarily conserved
571 elements in vertebrate, insect, worm, and yeast genomes. *Genome research* 15: 1034–1050.
- 572 51. Cooper GM, Goode DL, Ng SB, Sidow A, Bamshad MJ, et al. (2010) Single-nucleotide evolutionary
573 constraint scores highlight disease-causing mutations. *Nature methods* 7: 250–251.
- 574 52. Kmoč S, Hartmannová H, Stibůrková B, Krijt J, Zikánová M, et al. (2000) Human adenylosucci-
575 nate lyase (adsl), cloning and characterization of full-length cdna and its isoform, gene structure
576 and molecular basis for adsl deficiency in six patients. *Human molecular genetics* 9: 1501–1513.
- 577 53. Maaswinkel-Mooij P, Laan L, Onkenhout W, Brouwer O, Jaeken J, et al. (1997) Adenylosuccinase
578 deficiency presenting with epilepsy in early infancy. *Journal of inherited metabolic disease* 20:
579 606–607.

- 580 54. Marie S, Cuppens H, Heuterspreute M, Jaspers M, Tola EZ, et al. (1999) Mutation analysis in
581 adenylosuccinate lyase deficiency: Eight novel mutations in the re-evaluated full adsl coding se-
582 quence. *Human mutation* 13: 197–202.
- 583 55. Race V, Marie S, Vincent MF, Van den Berghe G (2000) Clinical, biochemical and molecular
584 genetic correlations in adenylosuccinate lyase deficiency. *Human molecular genetics* 9: 2159–2165.
- 585 56. Edery P, Chabrier S, Ceballos-Picot I, Marie S, Vincent MF, et al. (2003) Intrafamilial variability
586 in the phenotypic expression of adenylosuccinate lyase deficiency: a report on three patients.
587 *American Journal of Medical Genetics Part A* 120: 185–190.
- 588 57. Meister G, Landthaler M, Peters L, Chen PY, Urlaub H, et al. (2005) Identification of novel
589 argonaute-associated proteins. *Current biology* 15: 2149–2155.
- 590 58. Du KL, Chen M, Li J, Lepore JJ, Mericko P, et al. (2004) Megakaryoblastic leukemia factor-1
591 transduces cytoskeletal signals and induces smooth muscle cell differentiation from undifferentiated
592 embryonic stem cells. *Journal of Biological Chemistry* 279: 17578–17586.
- 593 59. Mercher T, Busson-Le Coniat M, Monni R, Mauchauffé M, Khac FN, et al. (2001) Involvement of
594 a human gene related to the drosophila spen gene in the recurrent t (1; 22) translocation of acute
595 megakaryocytic leukemia. *Proceedings of the National Academy of Sciences* 98: 5776–5779.
- 596 60. Trahey M, Wong G, Halenbeck R, Rubinfeld B, Martin GA, et al. (1988) Molecular cloning of two
597 types of gap complementary dna from human placenta. *Science* 242: 1697–1700.
- 598 61. Friedman E, Gejman PV, Martin GA, McCormick F (1993) Nonsense mutations in the c-terminal
599 sh2 region of the gtpase activating protein (gap) gene in human tumours. *Nature genetics* 5:
600 242–247.
- 601 62. Eerola I, Boon LM, Mulliken JB, Burrows PE, Dompmartin A, et al. (2003) Capillary
602 malformation–arteriovenous malformation, a new clinical and genetic disorder caused by *rasa1*
603 mutations. *The American Journal of Human Genetics* 73: 1240–1249.
- 604 63. Hershkovitz D, Bercovich D, Sprecher E, Lapidot M (2008) *Rasa1* mutations may cause hereditary
605 capillary malformations without arteriovenous malformations. *British Journal of Dermatology* 158:
606 1035–1040.

- 607 64. Whiting PJ, Bonnert TP, McKernan RM, Farrar S, Bourdelles BL, et al. (1999) Molecular and
608 functional diversity of the expanding gaba-a receptor gene family. *Annals of the New York Academy
609 of Sciences* 868: 645–653.
- 610 65. Edenberg HJ, Dick DM, Xuei X, Tian H, Almasy L, et al. (2004) Variations in *gabra2*, encoding
611 the $\alpha 2$ subunit of the gaba a receptor, are associated with alcohol dependence and with brain
612 oscillations. *The American Journal of Human Genetics* 74: 705–714.
- 613 66. Knabl J, Witschi R, Hösl K, Reinold H, Zeilhofer UB, et al. (2008) Reversal of pathological pain
614 through specific spinal gaba_a receptor subtypes. *Nature* 451: 330–334.
- 615 67. Xiang YY, Wang S, Liu M, Hirota JA, Li J, et al. (2007) A gabaergic system in airway epithelium
616 is essential for mucus overproduction in asthma. *Nature medicine* 13: 862–867.
- 617 68. Ma D, Whitehead P, Menold M, Martin E, Ashley-Koch A, et al. (2005) Identification of significant
618 association and gene-gene interaction of gaba receptor subunit genes in autism. *The American
619 Journal of Human Genetics* 77: 377–388.
- 620 69. Collins AL, Ma D, Whitehead PL, Martin ER, Wright HH, et al. (2006) Investigation of autism
621 and gaba receptor subunit genes in multiple ethnic groups. *Neurogenetics* 7: 167–174.
- 622 70. Ariani F, Hayek G, Rondinella D, Artuso R, Mencarelli MA, et al. (2008) *Foxg1* is responsible for
623 the congenital variant of rett syndrome. *The American Journal of Human Genetics* 83: 89–93.
- 624 71. Mencarelli M, Spanhol-Rosseto A, Artuso R, Rondinella D, De Filippis R, et al. (2010) Novel *foxg1*
625 mutations associated with the congenital variant of rett syndrome. *Journal of medical genetics* 47:
626 49–53.
- 627 72. Sadakata T, Furuichi T (2010) Ca²⁺-dependent activator protein for secretion 2 and autistic-like
628 phenotypes. *Neuroscience research* 67: 197–202.
- 629 73. Crisci JL, Wong A, Good JM, Jensen JD (2011) On characterizing adaptive events unique to
630 modern humans. *Genome biology and evolution* 3: 791–798.
- 631 74. Guilherme A, Soriano NA, Furcinitti PS, Czech MP (2004) Role of *ehd1* and *ehbp1* in perinuclear
632 sorting and insulin-regulated glut4 recycling in 3t3-l1 adipocytes. *Journal of Biological Chemistry*
633 279: 40062–40075.

- 634 75. Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, et al. (2008) Common se-
635 quence variants on 2p15 and xp11. 22 confer susceptibility to prostate cancer. *Nature genetics* 40:
636 281–283.
- 637 76. Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, et al. (2003) A gene expression atlas of
638 the central nervous system based on bacterial artificial chromosomes. *Nature* 425: 917–925.
- 639 77. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, et al. (2006) In vivo enhancer
640 analysis of human conserved non-coding sequences. *Nature* 444: 499–502.
- 641 78. Li MJ, Wang P, Liu X, Lim EL, Wang Z, et al. (2011) Gwasdb: a database for human genetic
642 variants identified by genome-wide association studies. *Nucleic acids research* : gkr1182.
- 643 79. Welter D, MacArthur J, Morales J, Burdett T, Hall P, et al. (2014) The nhgri gwas catalog, a
644 curated resource of snp-trait associations. *Nucleic acids research* 42: D1001–D1006.
- 645 80. Paternoster L, Evans DM, Nohr EA, Holst C, Gaborieau V, et al. (2011) Genome-wide population-
646 based association study of extremely overweight young adults—the goya study. *PLoS One* 6: e24303.
- 647 81. Suhre K, Wallaschofski H, Raffler J, Friedrich N, Haring R, et al. (2011) A genome-wide association
648 study of metabolic traits in human urine. *Nature genetics* 43: 565–569.
- 649 82. Perlis RH, Huang J, Purcell S, Fava M, Rush AJ, et al. (2010) Genome-wide association study of
650 suicide attempts in mood disorder patients. *Genome* 167.
- 651 83. Henrion M, Frampton M, Scelo G, Purdue M, Ye Y, et al. (2013) Common variation at 2q22. 3
652 (*zeb2*) influences the risk of renal cancer. *Human molecular genetics* 22: 825–831.
- 653 84. Schlebusch CM, Skoglund P, Sjödin P, Gattepaille LM, Hernandez D, et al. (2012) Genomic varia-
654 tion in seven khoe-san groups reveals adaptation and complex african history. *Science* 338: 374–379.
- 655 85. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, et al. (2002) Detecting recent positive
656 selection in the human genome from haplotype structure. *Nature* 419: 832–837.
- 657 86. Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, et al. (2011) Classic selective sweeps
658 were rare in recent human evolution. *science* 331: 920–924.

- 659 87. Pace L, Salvan A, Sartori N (2011) Adjusting composite likelihood ratio statistics. *Statistica Sinica*
660 21: 129.
- 661 88. Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, et al. (2014) Genome sequence of a 45,000-year-old
662 modern human from western siberia. *Nature* 514: 445–449.
- 663 89. Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspinas AS, Manica A, et al. (2014) Genomic
664 structure in europeans dating back at least 36,200 years. *Science* 346: 1113–1118.
- 665 90. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, et al. (2014) Ancient human genomes
666 suggest three ancestral populations for present-day europeans. *Nature* 513: 409–413.
- 667 91. Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide
668 allele frequency data. *PLoS genetics* 8: e1002967.

669 Tables

Table 1. Description of models tested. 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.

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

Table 2. Top hits for 3P-CLR run on the European terminal branch, using Africans as the outgroup. We show the windows in the top 99.9% quantile of scores. Windows were merged together if they 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
17	19175100	18858600	19445800	213.864	SLC5A10,FAM83G,GRAP,GRAPL,EPN2,B9D1,MAPK7,MFAP4,RNF112,SLC47A1
15	29241200	29210600	29338200	180.648	APBA2
14	66765100	66417200	67923400	176.032	GPHN,FAM71D,MPP5,ATP6V1D,EIF2S1,PLEK2,TMEM229B
10	74736500	74007800	75402200	161.652	DDIT4,DNAJB12,MICU1,MCU,OIT3,PLA2G12B,P4HA1,NUDT13,ECD,FAM149B1,DNAJC9,MRPS16,TTC18,ANXA7,MSS51,PPP3CB,USP54,MYOZ1,DLGAP3,ZMYM6NB,ZMYM6,ZMYM1,SFPQ,ZMYM4,KIAA0319L,NCDN,TFAP2E,PSMB2,C1orf216,CLSPN,AGO4,AGO1,AGO3,TEKT2,ADPRHL2,COL8A2,RASGRP2,PYGM,SF1,MAP4K2,MEN1,SLC22A11,SLC22A12,NRXN2,BRAP,ACAD10,ALDH2,MAPKAPK5,TMEM116,ERP29,NAA25,TRAFD1,RPL6,PTPN11,RPH3A,CUX2,FAM109A,SH2B3,ATXN2
1	35623900	35380800	36584500	150.189	INO80B,WBP1,MOGS,MRPL53,CCDC142,TTC31,LBX2,PCGF1,TLX2,DQX1,AUP1,NOTO,SMYD5,PRADC1,CCT7,HTRA2,LOXL3,DOK1,M1AP,SEMA4F,FBXO41,EGR4,ALMS1,NAT8,TPRKB,DUSP11,C2orf78,STAMBP,ACTG2,DGUOK,TET3,BOLA3,MOB1A,MTHFD2,SLC4A5,DCTN1,WDR54,RTKN
11	64581000	64217100	64588600	148.952	C15orf43,SORD,DUOX2,DUOXA2,DUOXA1,DUOX1
12	113010000	111691000	113030000	145.64	ZSCAN25,CYP3A5,CYP3A7,CYP3A4,SMURF1,KPNA7,ARPC1A,ARPC1B,PDAP1,BUD31,PTCD1,ATP5J2
2	74507400	73404200	74970500	137.92	PTCD1,CPSF4,ATP5J2,ZNF789,ZNF394,ZKSCAN5,FAM200A,ZNF655
15	45332100	45094600	45436600	137.885	THSD4,MYO9A,SENPA8,GRAMD2,PKM,PARP6,CELF6,HEXA,TMEM202,ARIH1,GOLGA6B,BBS4,ADPGK
7	98882800	98717700	99369400	135.106	SPIN1,NXN2
15	72654000	72057800	73138200	132.697	NRG3
9	91155000	90913100	91201600	125.597	FGF1,ARHGAP26
10	83601100	83597800	83761400	120.455	ZEB1
5	142116000	142074000	142194000	117.87	CCDC102B
10	31863100	31479100	31908500	113.751	-
18	66807000	66646700	66883000	108.186	IMPG1
2	104933000	104749000	105027000	107.684	-
6	76751700	76636000	77261200	106.577	-
7	81142700	81087600	81298600	104.727	-
4	167411000	167094000	167644000	104.554	-
21	21424100	21378700	21643900	104.405	-
2	216600000	216551000	216628000	103.798	-
17	58512800	58075800	59174400	103.791	HEATR6,CA4,USP32,C17orf64,APPBP2,PPM1D,BCAS3
5	123509000	123370000	123604000	103.386	-
6	150686000	150637000	150738000	103.115	IYD
15	35551700	35444900	35727400	102.676	DPH6
6	121627000	121082000	121788000	102.59	TBC1D32,GJA1
4	60872200	60814500	61356600	98.9547	-
9	108572000	108412000	108755000	98.9047	TAL2,TMEM38B
1	204823000	204680000	204872000	98.6962	NFASC
20	53878400	53876100	54051800	95.6355	-
10	93143500	93060300	93325000	95.5239	HECTD2
1	162116000	162002000	162228000	95.4304	NOS1AP
9	12777200	12488900	12787600	95.1877	TYRP1,LURAP1L
18	7330950	7259810	7374120	93.1786	-
3	188661000	188641000	188840000	91.5491	TPRG1
15	52859500	52581800	52992200	91.5176	MYO5C,MYO5A,ARPP19,FAM214A
15	48211200	48153900	48308500	90.9933	-
12	80298900	80117100	80435100	90.0939	PPP1R12A
11	38229500	37879000	38607000	86.7175	-
5	82679100	82488400	82790300	85.7742	XRCC4,VCAN
6	43624100	43419100	43688500	85.5502	DLK2,TJAP1,LRRRC73,POLR1C,YIPF3,XPO5,POLH,GTPBP2,MAD2L1BP,RSPH9,MRPS18A
13	48977600	48726500	49291500	85.2064	ITM2B,RB1,LPAR6,RCBTB2,CYSLTR2
1	53568300	53124700	53633500	84.8166	FAM159A,COA7,ZYG11B,ZYG11A,ECHDC2,SCP2,PODN,SLC1A7
10	54166700	54130800	54335700	84.1889	-
8	15994200	15838700	15997700	83.0166	MSR1
15	94107900	94022200	94185300	82.8563	-
11	129910000	129805000	130073000	82.6407	PRDM10,APLP2,ST14

Table 3. Top hits for 3P-CLR run on the East Asian terminal branch, using Africans as the outgroup. We show the windows in the top 99.9% quantile of scores. Windows were merged together if they 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
5	117510000	117345000	117716000	270.211	-
3	58238900	58104900	58557500	257.128	FLNB, DNASE1L3, ABHD6, RPP14, PXX, PDHB, KCTD6, ACOX2, FAM107A
10	94874300	94840100	95720400	235.225	MYOF, CEP55, FFAR4, RBP4, PDE6C, FRA10AC1, LGI1, SLC35G1
15	64166100	63723000	64339800	231.645	USP3, FBXL2, HERC1, DAPK2
4	42193900	41823000	42206800	231.205	TMEM33, DCAF4L1, SLC30A9, BEND4
2	72378700	72353700	73177300	221.359	CYP26B1, EXOC6B, SPR, EMX1, SFXN5
1	234347000	234207000	234380000	208.215	SLC35F3
11	25172400	25098500	25276200	206.418	LUZP2
4	158638000	158481000	158740000	190.705	-
17	61536300	60912100	61549600	185.021	TANC2, CYB561
20	24793800	24570100	25037800	177.801	SYNDIG1, CST7, APMAP, ACSS1
4	86504400	86438300	86602900	173.091	ARHGAP24
10	56026900	55868800	56209000	172.266	PCDH15
1	75622900	75277800	76729300	167.716	LHX8, SLC44A5, ACADM, RABGGTB, MSH4, ASB17, ST6GALNAC3
18	5299800	5203000	5314080	163.435	ZBTB14
7	112265000	112125000	112622000	157.219	LSMEM1, TMEM168, C7orf60
8	10836400	10559800	11126200	150.728	RP1L1, SOX7, PINX1, XKR6
1	172931000	172670000	172950000	149.67	-
4	135424000	134792000	135547000	149.303	-
10	53363100	53226200	53440300	147.8	PRKG1
7	1.09E+08	108741000	109226000	146.998	-
13	63542000	63261200	63971200	146.966	-
3	102005000	101902000	102242000	146.338	ZPLD1
6	69974500	69524500	70359500	144.236	BAI3
2	56096500	55929400	56198400	139.435	EFEMP1
2	109534000	108937000	109626000	139.089	LIMS1, RANBP2, CCDC138, EDAR, SULT1C4, GCC2
22	46760700	46594600	46831200	138.989	PPARA, CDPF1, PKDREJ, TTC38, GTSE1, TRMU, CELSR1
3	104826000	104604000	104910000	138.473	-
18	67572500	67533400	67877100	138.033	CD226, RTTN
2	26159900	25853900	26233500	135.742	KIF3C, DTNB
20	31604100	31304800	31614200	134.323	COMMD7, DNMT3B, MAPRE1, SUN5, BPIFB2
2	17456900	17247000	17564600	134.267	-
4	28858500	28537900	28879000	131.871	-
9	107052000	106657000	107058000	131.263	SMC2
12	93322200	92983200	93454700	129.603	C12orf74, PLEKHG7, EEA1
4	80074800	79878800	80250300	129.59	NAA11
5	153541000	153053000	153736000	129.577	GRIA1, FAM114A2, MFAP3, GALNT10
14	77987200	77720700	78086400	126.295	TMEM63C, NGB, POMT2, GSTZ1, TMED8, SAMD15, NOXRED1, VIPAS39, AHSA1, ISM2, SPTLC2
6	8002470	7975450	8112550	125.312	TXNDC5, BLOC1S5, EEF1E1
12	124021000	123925000	124275000	125.199	SNRNP35, RILPL1, TMED2, DDX55, EIF2B1, GTF2H3, TCTN2, ATP6V0A2, DNAH10
10	97039700	96682100	97059000	125.048	CYP2C9, CYP2C8, C10orf129, PDLIM1
16	67607200	66947800	68430200	124.191	ACD, PARD6A, ENKD1, C16orf86, GFOD2, RANBP10, TSNAXIP1, CENPT, THAP11, NUTF2, EDC4, NRNL, PSKH1, CTRL, PSME10, LCAT, SLC12A4, DPEP3, DPEP2, DUS2, DDX28, NFATC3, ESRP2, PLA2G15, SLC7A6, SLC7A6OS, PRMT7, SMPD3, CDH16, RRAD, FAM96B, CES2, CES3, CES4A, CBFB, C16orf70, B3GNT9, TRADD, FBXL8, HSF4, NOL3, KIAA0895L, EXOC3L1, E2F4, ELMO3, LRRC29, TMEM208, FHOD1, SLC9A5, PLEKHG4, KCTD19, LRRC36, TPPP3, ZDHHC1, HSD11B2, ATP6V0D1, AGRP, FAM65A, CTCF, RLTPR
12	103350000	103178000	103439000	123.87	PAH, ASCL1
5	111988000	111981000	112344000	123.442	APC, SRP19, REEP5, DCP2
14	69489100	69424000	69782800	122.252	ACTN1, DCAF5, EXD2, GALNT16
3	17197100	17188900	17897600	122.173	TBC1D5

Table 4. Top hits for 3P-CLR run on the Eurasian ancestral branch, using Africans as the outgroup. We show the windows in the top 99.9% quantile of scores. Windows were merged together if they 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
17	58778100	58075800	59308200	549.661	HEATR6,CA4,USP32,C17orf64,APPBP2,PPM1D,BCAS3
10	22705100	22428900	22798800	500.245	EBLN1,COMMD3,COMMD3-BMI1,BMI1,SPAG6
4	41834200	41452400	42195900	496.603	LIMCH1,PHOX2B,TMEM33,DCAF4L1,SLC30A9,BEND4
18	67572500	67533400	67881500	493.711	CD226,RTTN
17	62870600	62655400	63061700	484.401	SMURF2,LRRC37A3,GNA13
7	99227800	98717700	99374100	446.801	ZSCAN25,CYP3A5,CYP3A7,CYP3A4,SMURF1,KPNA7,ARPC1A,ARPC1B,PDAP1,BUD31,PTCD1,ATP5J2
2	73545700	72370500	74117400	444.484	PTCD1,CPSF4,ATP5J2,ZNF789,ZNF394,ZKSCAN5,FAM200A,ZNF655
10	93143500	93049100	93325000	440.141	CYP26B1,EXOC6B,SPR,EMX1,SFXN5,RAB11FIP5,NOTO,SMYD5,PRADC1,CCT7,FBXO41,EGRA,ALMS1,NAT8,TPRKB,DUSP11,C2orf78,STAMBIP
1	230018000	229912000	230132000	437.859	HECTD2
2	22420000	22187700	22469200	436.745	-
17	61536300	60888200	61549600	435.653	TANC2,CYB561
20	54054100	53877600	54056600	429.932	-
9	90946300	90908100	91200000	426.111	SPIN1,NXNL2
8	30625000	30515900	30891400	416.16	GSR,PPP2CB,TEX15,PURG,WRN
11	39699100	39604100	39937900	413.515	-
6	10644100	10578900	10784300	408.918	GCNT2,C6orf52,PAK11P1,TMEM14C,TMEM14B,SYCP2L,MAK
3	188751000	188646000	188859000	400.376	TPRG1
1	64483200	64340800	64538400	398.325	ROR1
10	31863100	31479100	31908500	397.626	ZEB1
4	177625000	177608000	177889000	395.393	VEGFC
14	57824300	57607700	58048400	392.93	EXOC5,AP5M1,NAA30,C14orf105
10	66018600	65795200	66311900	389.492	-
11	19609000	19591300	19731200	387.843	NAV2
13	49136300	48726500	49293400	386.345	ITM2B,RB1,LPAR6,RCBTB2,CYSLTR2
4	13424000	13143500	13535500	379.283	RAB28
8	52617300	52362200	52930500	377.6	PXDNL,PCMTD1
6	3149410	3073260	3204820	376.252	RIPK1,BPHL,TUBB2A
1	25592800	25517600	25869600	372.377	SYF2,C1orf63,RHD,TMEM50A,RHCE,TMEM57
3	97346000	96453200	97364600	371.806	EPHA6
6	105946000	105800000	105954000	371.534	PREP
15	45332100	45094600	45436600	369.82	C15orf43,SORD,DUOX2,DUOXA2,DUOXA1,DUOX1
12	111447000	111331000	111655000	368.231	CCDC63,MYL2,CUX2
6	131952000	131736000	132060000	366.757	ARG1,MED23,ENPP3,OR2A4,CTAGE9
5	11741300	11640500	11850200	365.997	CTNND2
1	116880000	116723000	117028000	365.955	ATP1A1
8	43401200	42499600	49036000	364.188	CHRN3,CHRNA6,THAP1,RNF170,HOK3,FNTA,POMK,HGSNAT,SPIDR,CEBPD,MCM4,UBE2V2
7	30270500	30178800	30471600	360.954	MTURN,ZNRF2,NOD1
4	33576600	33301000	33643000	359.425	-

Table 5. Enriched GO categories in the European, East Asian and Eurasian branches. We tested for ontology enrichment among the regions in the 99% quantile of the 3P-CLR scores for each population branch ($P < 0.05$, $FDR < 0.2$).

Population Branch	Raw p-value	FDR	GO category
Europe	0.00001	0.009243333	protein localization to membrane
Europe	0.00001	0.009243333	cellular protein localization
Europe	0.00001	0.009243333	cellular macromolecule localization
Europe	0.00003	0.016986	single-organism cellular localization
Europe	0.00003	0.016986	single-organism localization
Europe	0.00005	0.019915556	NAD(P)H oxidase activity
Europe	0.00005	0.019915556	cellular localization
Europe	0.00006	0.019915556	thyroid hormone generation
Europe	0.00006	0.019915556	cuticle development
Europe	0.0001	0.029226923	protein localization
Europe	0.00012	0.029226923	thyroid hormone metabolic process
Europe	0.00013	0.029226923	antioxidant activity
Europe	0.00013	0.029226923	macromolecule localization
Europe	0.00018	0.041946429	cellular response to reactive oxygen species
Europe	0.00022	0.047034444	protein localization to plasma membrane
Europe	0.00024	0.047034444	hydrogen peroxide catabolic process
Europe	0.00026	0.047034444	oxidoreductase activity, acting on peroxide as acceptor
Europe	0.00026	0.047034444	peroxidase activity
Europe	0.00034	0.061767895	cellular component assembly involved in morphogenesis
Europe	0.0004	0.069824	cellular response to hydrogen peroxide
Europe	0.00052	0.085701905	bicarbonate transport
Europe	0.00055	0.087832273	oxidoreductase activity, acting on NAD(P)H, oxygen as acceptor
Europe	0.00073	0.112074348	cell-cell junction
Europe	0.00085	0.127015	recycling endosome membrane
Europe	0.00101	0.144204	MLL5-L complex
Europe	0.00108	0.146987037	single-organism membrane organization
Europe	0.00112	0.146987037	interleukin-2 biosynthetic process
Europe	0.00143	0.171870313	plasma membrane organization
Europe	0.00145	0.171870313	PTW/PP1 phosphatase complex
Europe	0.00148	0.171870313	localization
Europe	0.0015	0.171870313	response to reactive oxygen species
Europe	0.0015	0.171870313	determination of left/right symmetry
Europe	0.00174	0.192166471	cleavage furrow
Europe	0.00178	0.192166471	hydrogen peroxide metabolic process
East Asia	0.00004	0.11411	regulation of peroxisome proliferator activated receptor signaling pathway
East Asia	0.00013	0.151273333	pyruvate dehydrogenase (acetyl-transferring) kinase activity
East Asia	0.00025	0.151273333	negative regulation of intestinal cholesterol absorption
East Asia	0.00025	0.151273333	regulation of intestinal cholesterol absorption
East Asia	0.00025	0.151273333	negative regulation of intestinal phytosterol absorption
East Asia	0.00028	0.151273333	negative regulation of peroxisome proliferator activated receptor signaling pathway
Eurasia	0.00001	0.0272	RNA polymerase II core promoter proximal region sequence-specific DNA binding
Eurasia	0.00003	0.042205	transcription factor activity involved in negative regulation of transcription
Eurasia			RNA polymerase II transcription regulatory region sequence-specific DNA binding
Eurasia			transcription factor activity involved in negative regulation of transcription

Table 6. Top hits for 3P-CLR run on the ancestral branch to Eurasians and Africans, using archaic humans as the outgroup. We show the windows in the top 99.9% quantile of scores. Windows were merged together if they 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
21	34916200	34737300	35222100	868.062	IFNGR2, TMEM50B, DNAJC28, GART, SON, DONSON, CRYZL1, ITSN1
17	56595700	56373200	57404800	846.821	BZRAP1, SUPT4H1, RNF43, HSF5, MTMR4, SEPT4, C17orf47, TEX14, RAD51C, PPM1E, TRIM37, SKA2, PRR11, SMG8, GDDP1
12	79919700	79756800	80109400	840.098	SYT1, PAWR
14	71790600	71658900	72283600	835.723	SIPA1L1
12	116589000	116366000	116760000	826.65	MED13L
2	37989300	37917400	38021500	823.495	CDC42EP3
3	36941700	36836900	37517500	819.361	TRANK1, EPM2AIP1, MLH1, LRRFIP2, GOLGA4, C3orf35, ITGA9
14	29635300	29222200	29696100	813.978	FOXG1
5	86911000	86463700	87101400	811.741	RASA1, CCNH
1	213498000	213145000	213561000	810.015	VASH2, ANGEL2, RPS6KC1
2	156468000	155639000	156764000	809.474	KCNJ3
4	146155000	145355000	146222000	806.198	HHIP, ANAPC10, ABCE1, OTUD4
7	121700000	121620000	122369000	794.997	PTPRZ1, AASS, FEZF1, CADPS2, RNF133, RNF148
17	61237300	60906000	61544500	793.984	TANC2, CYB561
7	107229000	106639000	107308000	784.544	PRKAR2B, HBP1, COG5, GPR22, DUS4L, BCAP29, SLC26A4
3	50576000	50177100	51929300	778.276	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
5	93214400	92677500	93645500	772.879	NR2F1, FAM172A, POU5F2, KIAA0825
7	99167000	98722100	99375300	764.946	ZSCAN25, CYP3A5, CYP3A7, CYP3A4, SMURF1, KPNA7, ARPC1A, ARPC1B, PDAP1, BUD31, PTCD1, ATP5J2
21	36769600	36689700	36842100	764.846	PTCD1, CPSP4, ATP5J2, ZNF789, ZNF394, ZKSCAN5, FAM200A, ZNF655
13	96782600	96180700	97420500	759.296	RUNX1
1	176411000	175890000	176437000	758.886	CLDN10, DZIP1, DNAJC3, UGGT2, HS6ST3
15	49269100	49247500	50036600	758.119	RFWD2, PAPA2
22	40521100	40350900	41223300	757.637	SECISBP2L, COPS2, GALK2, FAM227B, FGF7, DTWD1, SHC4
9	125562000	125505000	126059000	755.506	GRAP2, FAM83F, TNRC6B, ADSL, SGSM3, MKL1, MCHR1, SLC25A17, ST13
5	89579400	89413800	89654700	755.236	ZBTB26, RABGAP1, GPR21, STRBP, OR1L6, OR5C1, PDCL, OR1K1, RC3H2, ZBTB6
19	19313800	19100800	19788800	755.007	-
2	63889800	62767900	64394800	753.877	SUGP2, ARMC6, SLC25A42, TMEM161A, MEF2B, MEF2B, MEF2B, MEF2B, RFXANK, NR2C2AP, NCAN, HAPLN4, TM6SF2, SUGP1, MAU2, GATAD2A, TSSK6, NDUFA13, YJEFN3, CILP2, PBX4, LPAR2, GMIP, ATP13A1, ZNF101
2	73506800	73482800	74054300	749.705	EHBP1, OTX1, WDPCP, MDH1, UGP2, VPS54, PELI1
1	66833100	66772600	66952600	745.941	FBXO41, EGR4, ALMS1, NAT8, TPRKB, DUSP11, C2orf78
4	46634200	46361400	46994100	743.893	PDE4B
3	110709000	110513000	110932000	739.819	GABRA2, COX7B2, GABRA4
5	50148300	44575900	50411900	737.793	PVRL3
2	145109000	144689000	145219000	736.476	MRPS30, HCN1, EMB, PARP8
13	44935700	44887400	45237200	736.039	CTDC1, ZEB2
11	65212900	65129200	65379700	734.241	SERP2, TSC22D1
8	79211300	78656900	79554100	734.157	KCNK7, MAP3K11, SLC25A45, FRMD8, SCYL1, LTBP3, SSSCA1, FAM89B, EHBP1L1
4	22923300	22826000	23196800	733.814	PKIA
14	76101700	76054700	76450300	733.155	-
4	13346700	13142500	13533100	732.271	FLVCR2, TTL5, C14orf1, IFT43, TGFB3
12	97334600	96828100	97424100	729.578	RAB28
5	135849000	135588000	136038000	724.441	NEDD1
1	21083000	21012100	21629100	724.414	TRPC7
10	74730800	74007800	75399900	723.785	KIF17, SH2D5, HP1BP3, EIF4G3, ECE1
2	201156000	200636000	201355000	720.454	DDIT4, DNAJB12, MICU1, MCU, OIT3, PLA2G12B, P4HA1, NUDT13, ECD, FAM149B1, DNAJC9, MRPS16, TTC18, ANXA7, MSS51, PPP3CB, USP54, MYOZ1
3	147223000	146941000	147360000	719.916	C2orf69, TYW5, C2orf47, SPATS2L, KCTD18
5	127571000	127254000	127734000	712.643	ZIC4, ZIC1
9	23839000	23717700	24073800	711.977	SLC12A2, FBN2
					ELAVL2

Table 7. Enriched GO categories in the Modern Human branch. We tested for ontology enrichment among the regions in the 99.5% quantile of the 3P-CLR scores corresponding to the Modern Human branch after the split from archaic humans ($P < 0.05$, $FDR < 0.2$).

Population Branch	Raw p-value	FDR	GO category
Modern Human	0.00006	0.100843	B cell chemotaxis
Modern Human	0.00011	0.100843	histone H4 acetylation
Modern Human	0.00012	0.100843	transcription, DNA-templated
Modern Human	0.00015	0.100843	RNA biosynthetic process
Modern Human	0.00025	0.100843	heterocycle biosynthetic process
Modern Human	0.00028	0.100843	aromatic compound biosynthetic process
Modern Human	0.00028	0.100843	regulation of cell proliferation in bone marrow
Modern Human	0.00028	0.100843	positive regulation of cell proliferation in bone marrow
Modern Human	0.0003	0.100843	cellular nitrogen compound biosynthetic process
Modern Human	0.00031	0.100843	nucleobase-containing compound biosynthetic process
Modern Human	0.00049	0.138157692	organic cyclic compound biosynthetic process
Modern Human	0.00052	0.138157692	lymphocyte chemotaxis
Modern Human	0.00055	0.138157692	cellular macromolecule biosynthetic process
Modern Human	0.00063	0.144245625	RNA metabolic process
Modern Human	0.00069	0.144245625	regulation of cell cycle
Modern Human	0.0007	0.144245625	activation of Rap GTPase activity
Modern Human	0.00078	0.146167222	negative regulation of cell cycle process
Modern Human	0.0008	0.146167222	regulation of organelle organization
Modern Human	0.00099	0.149766452	germinal center formation
Modern Human	0.00103	0.149766452	regulation of cytoskeleton organization
Modern Human	0.00106	0.149766452	mitotic spindle assembly checkpoint
Modern Human	0.0011	0.149766452	nuclear division
Modern Human	0.00114	0.149766452	negative regulation of mitotic metaphase/anaphase transition
Modern Human	0.00114	0.149766452	negative regulation of metaphase/anaphase transition of cell cycle
Modern Human	0.00115	0.149766452	leucine zipper domain binding
Modern Human	0.00124	0.149766452	mitotic spindle checkpoint
Modern Human	0.00126	0.149766452	spindle assembly checkpoint
Modern Human	0.00135	0.149766452	cell cycle process
Modern Human	0.00139	0.149766452	negative regulation of molecular function
Modern Human	0.0014	0.149766452	anaphase-promoting complex
Modern Human	0.00142	0.149766452	mitosis
Modern Human	0.00151	0.1552175	spindle checkpoint
Modern Human	0.00162	0.158729118	lymphocyte migration
Modern Human	0.00162	0.158729118	intra-Golgi vesicle-mediated transport
Modern Human	0.00181	0.167352222	establishment or maintenance of monopolar cell polarity
Modern Human	0.00181	0.167352222	establishment of monopolar cell polarity
Modern Human	0.00192	0.171855135	mitotic anaphase
Modern Human	0.00208	0.177203721	mitotic cell cycle
Modern Human	0.00216	0.177203721	negative regulation of cell cycle
Modern Human	0.00222	0.177203721	establishment of protein localization to Golgi
Modern Human	0.00222	0.177203721	protein targeting to Golgi
Modern Human	0.00231	0.177203721	anaphase
Modern Human	0.00234	0.177203721	myeloid cell differentiation
Modern Human	0.00255	0.179768044	negative regulation of mitosis
Modern Human	0.00257	0.179768044	histone H4-K5 acetylation
Modern Human	0.00257	0.179768044	histone H4-K8 acetylation
Modern Human	0.00279	0.188911458	organelle fission
Modern Human	0.00284	0.188911458	nucleobase-containing compound metabolic process
Modern Human	0.003	0.1893844	regulation of metaphase/anaphase transition of cell cycle
Modern Human	0.003	0.1893844	regulation of mitotic metaphase/anaphase transition

Table 8. Overlap between GWAS catalog and catalog of modern human-specific high-frequency changes in the top modern human selected regions. 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
2	64279606	rs10171434	C	T	A/A,A/A	0.92	NA	regulatory	8.358	Suicide attempts in bipolar disorder	21041247
2	64279606	rs10171434	C	T	A/A,A/A	0.92	NA	regulatory	8.358	Urinary metabolites	21572414
2	144783214	rs16823411	T	C	A/A,A/A	0.93	GTDC1	intron	4.112	Body mass index	21701565
2	144783214	rs16823411	T	C	A/A,A/A	0.93	GTDC1	intron	4.112	Body mass index	21701565
2	145213638	rs731108	G	C	A/A,H/H	0.92	ZEB2	regulatory	10.31	Renal cell carcinoma	23184150
2	156506516	rs4407211	C	T	A/A,A/A	0.92	NA	intergenic	1.348	Alcohol consumption	23953852
3	51142359	rs4286453	T	C	A/A,A/A	0.91	DOCK3	intron	4.96	Multiple complex diseases	17554300
3	51824167	rs6796373	G	C	A/A,A/A	0.94	NA	intergenic	1.381	Response to taxane treatment (placlitaxel)	23006423
3	147200492	rs9876193	G	A	H/H,A/A	0.95	ZIC1	intron_nc	6.856	Type 2 diabetes	17463246
4	13325741	rs2867467	G	C	A/A,A/A	0.91	NA	intergenic	0.476	Obesity (extreme)	21935397
4	13328373	rs6842438	T	C	A/A,A/A	0.92	NA	intergenic	5.241	Obesity (extreme)	21935397
4	13330095	rs10019897	C	T	A/A,A/A	0.92	NA	upstream	1.472	Multiple complex diseases	17554300
4	13330095	rs10019897	C	T	A/A,A/A	0.92	NA	upstream	1.472	Obesity (extreme)	21935397
4	13333413	rs9996364	A	G	A/A,A/A	0.92	HSP90AB2P	upstream	5.865	Obesity (extreme)	21935397
4	13338465	rs11945340	C	T	A/A,A/A	0.92	HSP90AB2P	non coding exon	12.04	Obesity (extreme)	21935397
4	13340249	rs6839621	T	C	A/A,A/A	0.92	HSP90AB2P	non coding exon	0.074	Obesity (extreme)	21935397
4	13346602	rs11930614	C	T	A/A,A/A	0.92	NA	intergenic	0.587	Obesity (extreme)	21935397
4	13350973	rs10021881	T	C	A/A,A/A	0.92	NA	regulatory	3.032	Obesity (extreme)	21935397
4	13356393	rs16888596	G	A	A/A,A/A	0.94	NA	intergenic	2.344	Obesity (extreme)	21935397
4	13357274	rs11732938	A	G	A/A,A/A	0.94	NA	intergenic	15.45	Obesity (extreme)	21935397
4	13360622	rs11947529	T	A	A/A,A/A	0.93	RAB28	downstream	4.356	Obesity (extreme)	21935397
4	13363958	rs12331157	A	G	A/A,A/A	0.97	RAB28	intron	1.3	Obesity (extreme)	21935397
4	13363974	rs12332023	C	T	A/A,A/A	0.97	RAB28	intron	0.75	Obesity (extreme)	21935397
4	13366481	rs7673680	C	T	A/A,A/A	0.93	RAB28	downstream	4.16	Obesity (extreme)	21935397
4	13370308	rs10003958	T	C	A/A,A/A	0.93	RAB28	regulatory	16.58	Obesity (extreme)	21935397
4	13373583	rs9999851	C	T	A/A,A/A	0.97	RAB28	intron	1.305	Obesity (extreme)	21935397
4	13374462	rs9291610	G	A	A/A,A/A	0.93	RAB28	intron	3.264	Obesity (extreme)	21935397
4	13393897	rs9998914	A	T	A/A,A/A	0.96	RAB28	intron	0.414	Obesity (extreme)	21935397
4	13403855	rs11943295	G	A	A/A,A/A	0.94	RAB28	intron	1.702	Multiple complex diseases	17554300
4	13403855	rs11943295	G	A	A/A,A/A	0.94	RAB28	intron	1.702	Obesity (extreme)	21935397
4	13403998	rs11943330	G	A	A/A,A/A	0.93	RAB28	intron	3.295	Obesity (extreme)	21935397
4	13404130	rs7677336	G	T	A/A,A/A	0.94	RAB28	intron	0.752	Obesity (extreme)	21935397
4	13404717	rs7673732	A	C	A/A,A/A	0.93	RAB28	intron	0.702	Obesity (extreme)	21935397
4	13440031	rs11737264	C	G	A/A,A/A	0.93	RAB28	intron	1.159	Obesity (extreme)	21935397
4	13440271	rs11737360	C	T	A/A,A/A	0.94	RAB28	intron	2.745	Obesity (extreme)	21935397
4	13449532	rs16888654	A	C	A/A,A/A	0.94	RAB28	intron	0.46	Obesity (extreme)	21935397
4	13452022	rs16888661	C	A	A/A,A/A	0.91	RAB28	intron	5.359	Obesity (extreme)	21935397
4	13463991	rs11933841	T	C	A/A,A/A	0.93	RAB28	intron	4.193	Obesity (extreme)	21935397
4	13465710	rs11947665	T	A	A/A,A/A	0.93	RAB28	intron	4.41	Obesity (extreme)	21935397
4	23095293	rs6825402	C	T	A/A,A/A	0.96	NA	intergenic	2.599	Multiple complex diseases	17554300
5	45393261	rs6874279	G	A	A/A,A/A	0.93	HCN1	intron	1.47	Alcohol dependence	20201924
5	45393261	rs6874279	G	A	A/A,A/A	0.93	HCN1	intron	1.47	Alcoholism	pha002891
5	89540468	rs2935504	C	T	A/A,A/A	0.97	RP11-61G23.1	non coding exon	4.52	Multiple complex diseases	17554300
7	106720932	rs12154324	C	A	A/A,A/A	0.93	NA	regulatory	5.411	Multiple complex diseases	17554300
13	44978167	rs9525954	G	A	A/A,A/A	0.95	RP11-269C23.3	intron	2.731	Type 2 diabetes	17463246
13	45034814	rs9533862	G	C	A/A,A/A	0.93	FILIP1LP1	intron	2.026	Suicide attempts in bipolar disorder	21041247
13	45055091	rs17065868	T	C	A/A,A/A	0.92	FILIP1LP1	intron	3.214	Antineutrophil cytoplasmic antibody-associated vasculitis	22808956

670 **Figures**

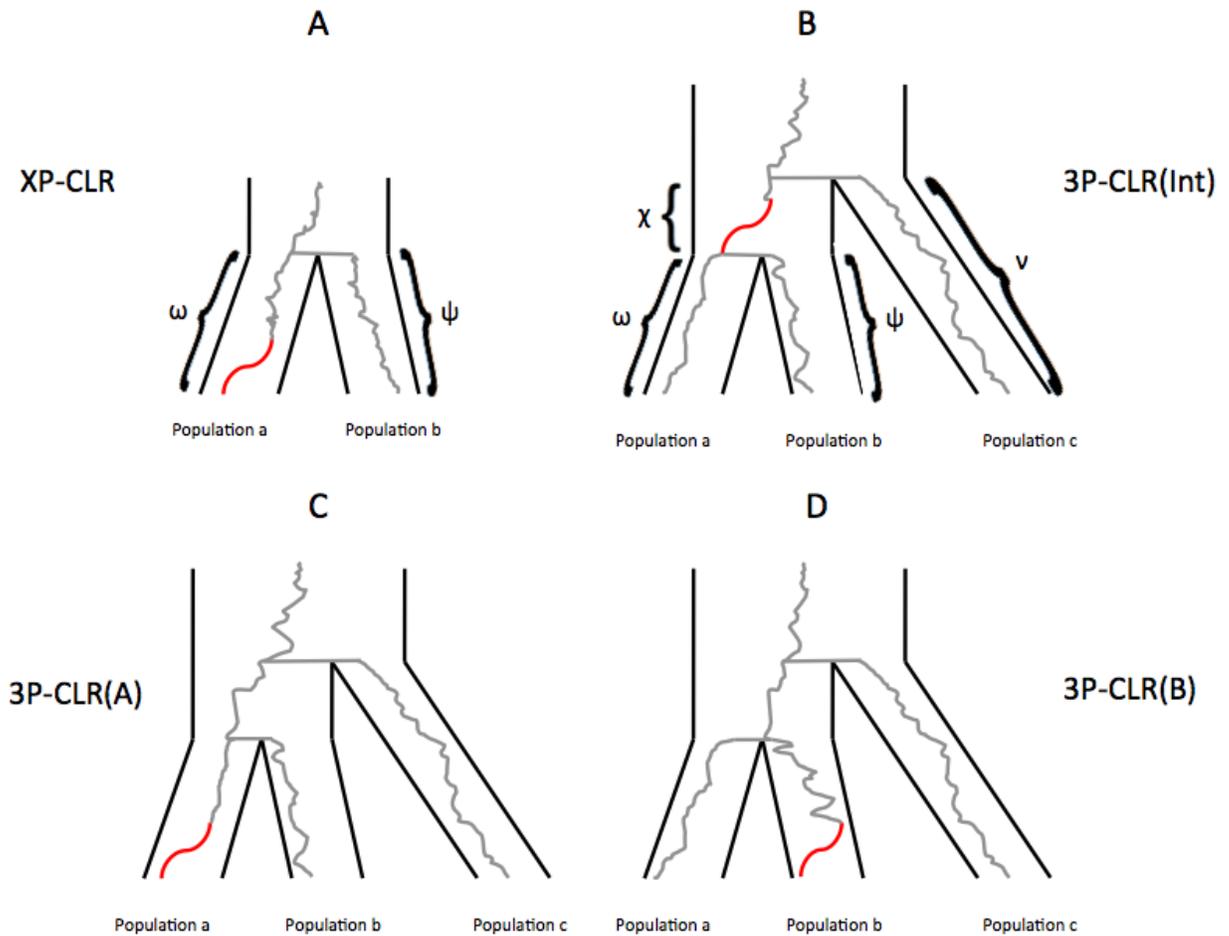


Figure 1. Schematic tree of selective sweeps detected by XP-CLR and 3P-CLR. While XP-CLR can only use two populations (an outgroup and a test) to detect selection (panel A), 3P-CLR can detect selection in the ancestral branch of two populations (3P-CLR(Int), panel B) or on the branches specific to each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

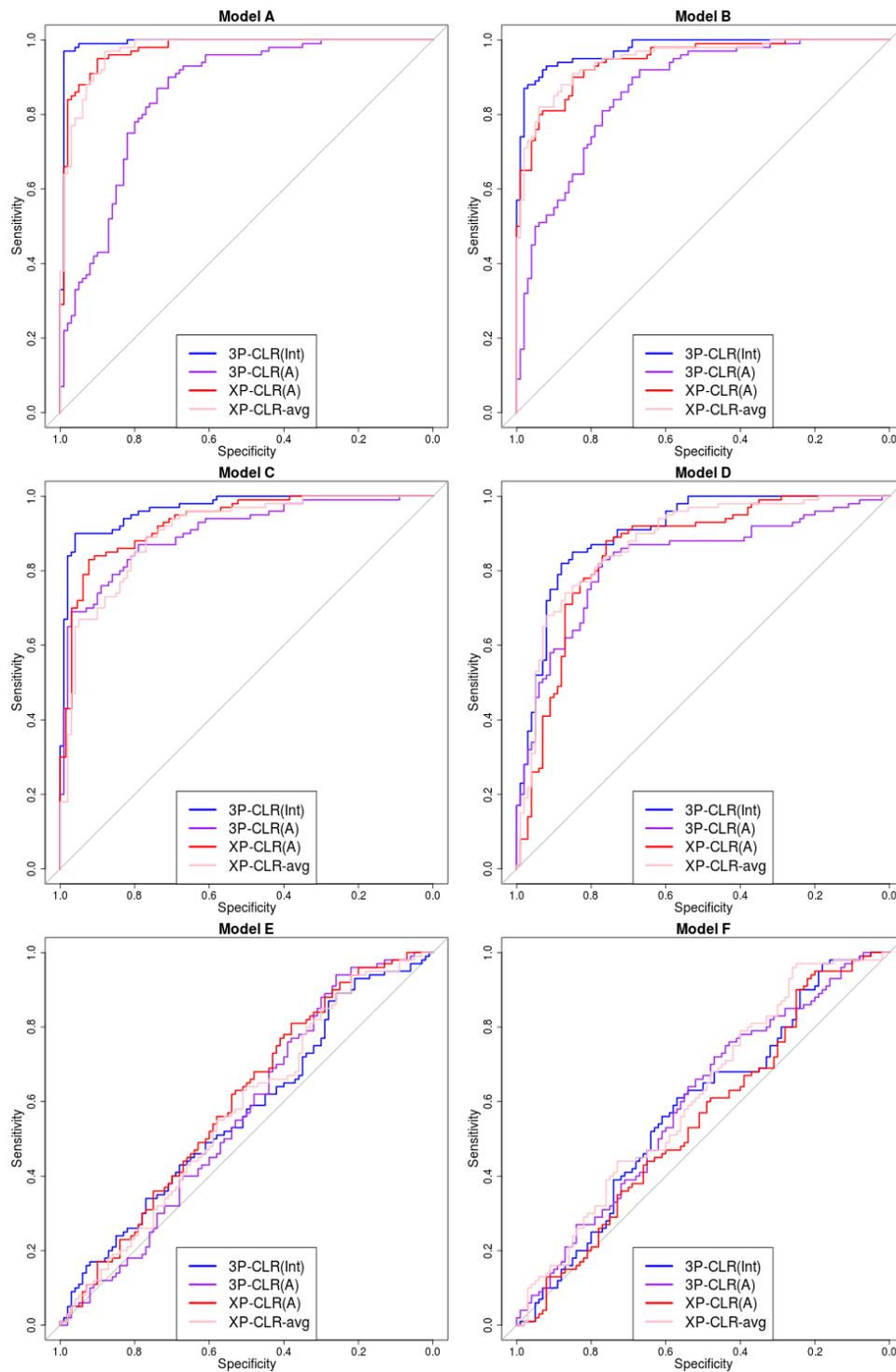


Figure 2. ROC curves for performance of 3P-CLR(Int), 3P-CLR(A) and two variants of XP-CLR in detecting selective sweeps that occurred before the split of two populations *a* and *b*, under different demographic models. In this case, the outgroup panel from population *c* contained 100 haploid genomes.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

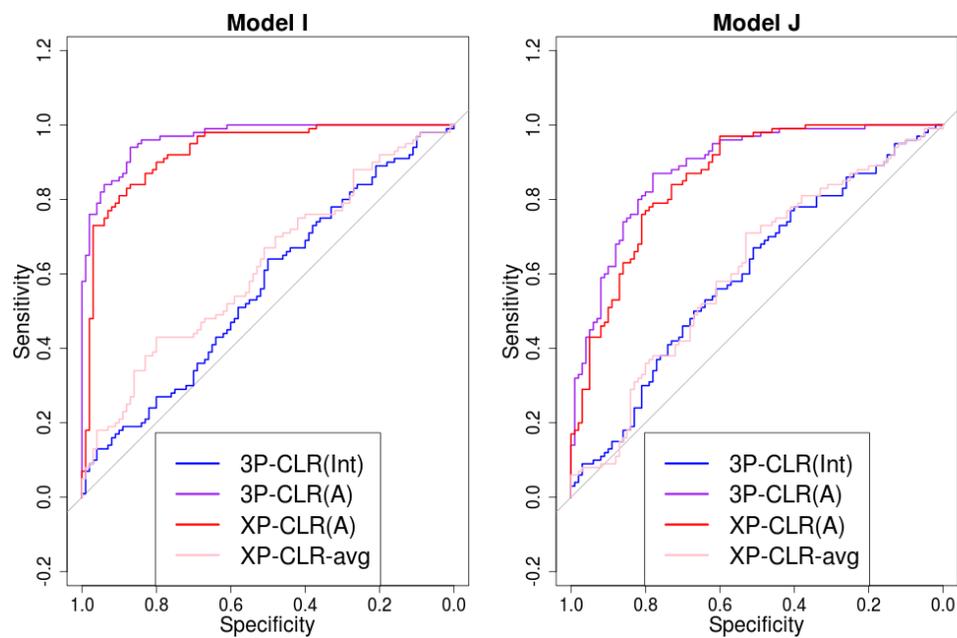


Figure 3. 3P-CLR(Int) is tailored to detect selective events that happened before the split t_{ab} , so it is largely insensitive to sweeps that occurred after the split. ROC curves show performance of 3P-CLR(Int) and two variants of XP-CLR for models where selection occurred in population a after its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

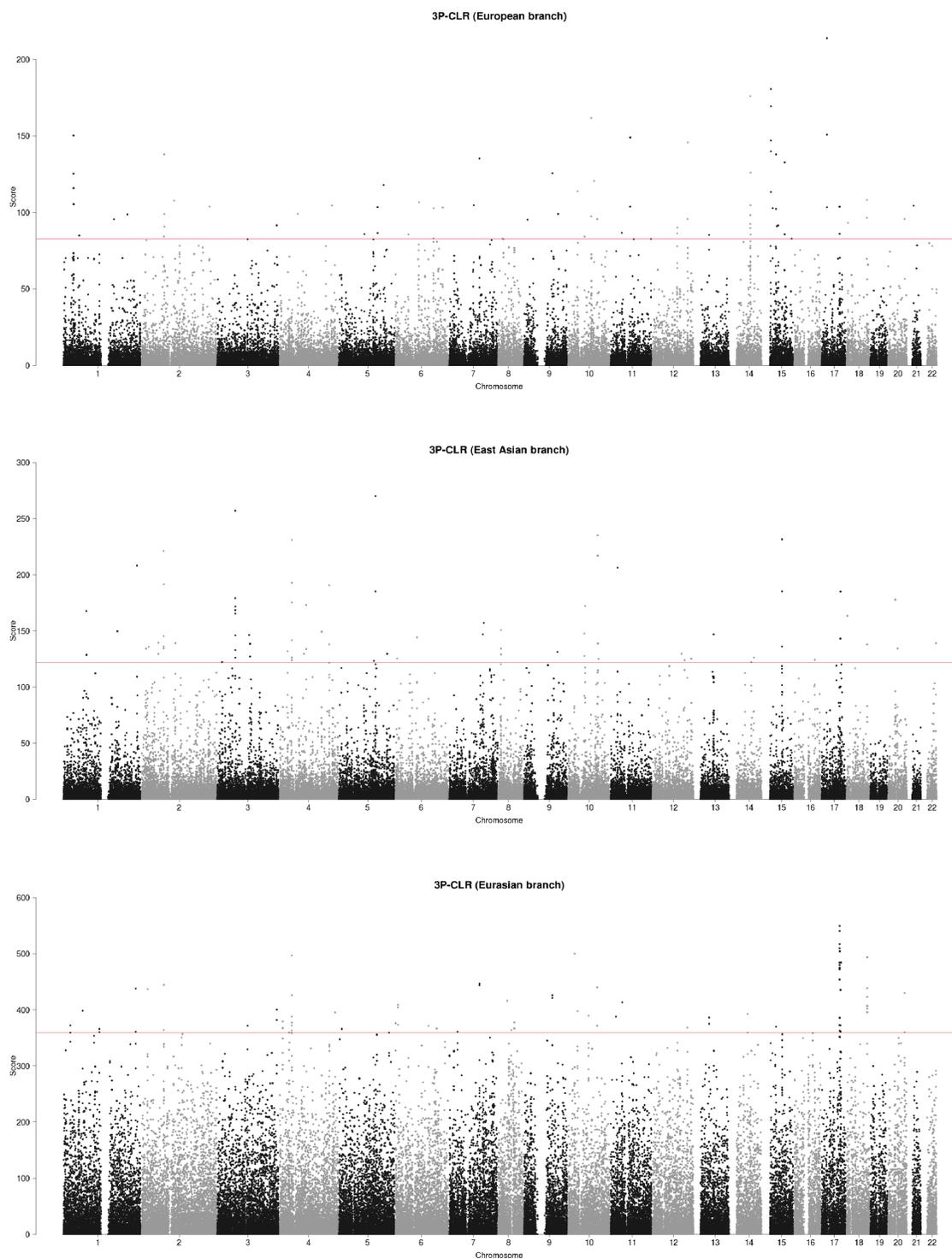


Figure 4. 3P-CLR scan of Europeans (upper panel), East Asians (middle panel) and the ancestral population to Europeans and East Asians (lower panel), using Africans as the outgroup in all 3 cases. The red line denotes the 99.9% quantile cutoff.

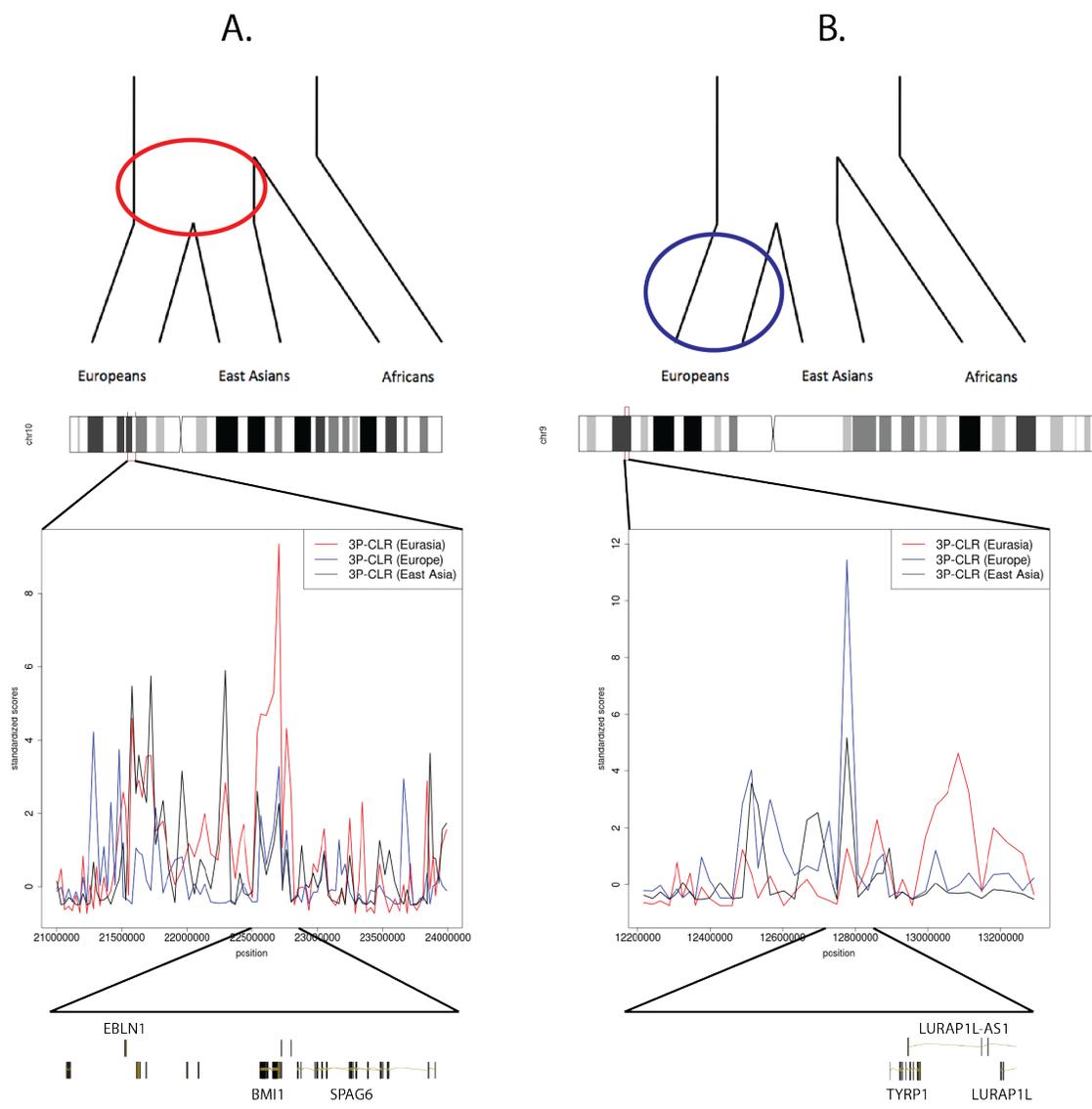


Figure 5. 3P-CLR scan of Europeans (blue), East Asians (black) and the ancestral Eurasian population (red) reveals regions under selection in different branches of the population tree. To make a fair comparison, all 3P-CLR scores were standardized by subtracting the chromosome-wide mean from each window and dividing the resulting score by the chromosome-wide standard deviation. A) The region containing genes *SPAG6* and *BMI1* is a candidate for selection in the ancestral population of Europeans and East Asians. B) The region containing *TYRP1* is a candidate for selection in the European population. The image was built using the GenomeGraphs package in Bioconductor.

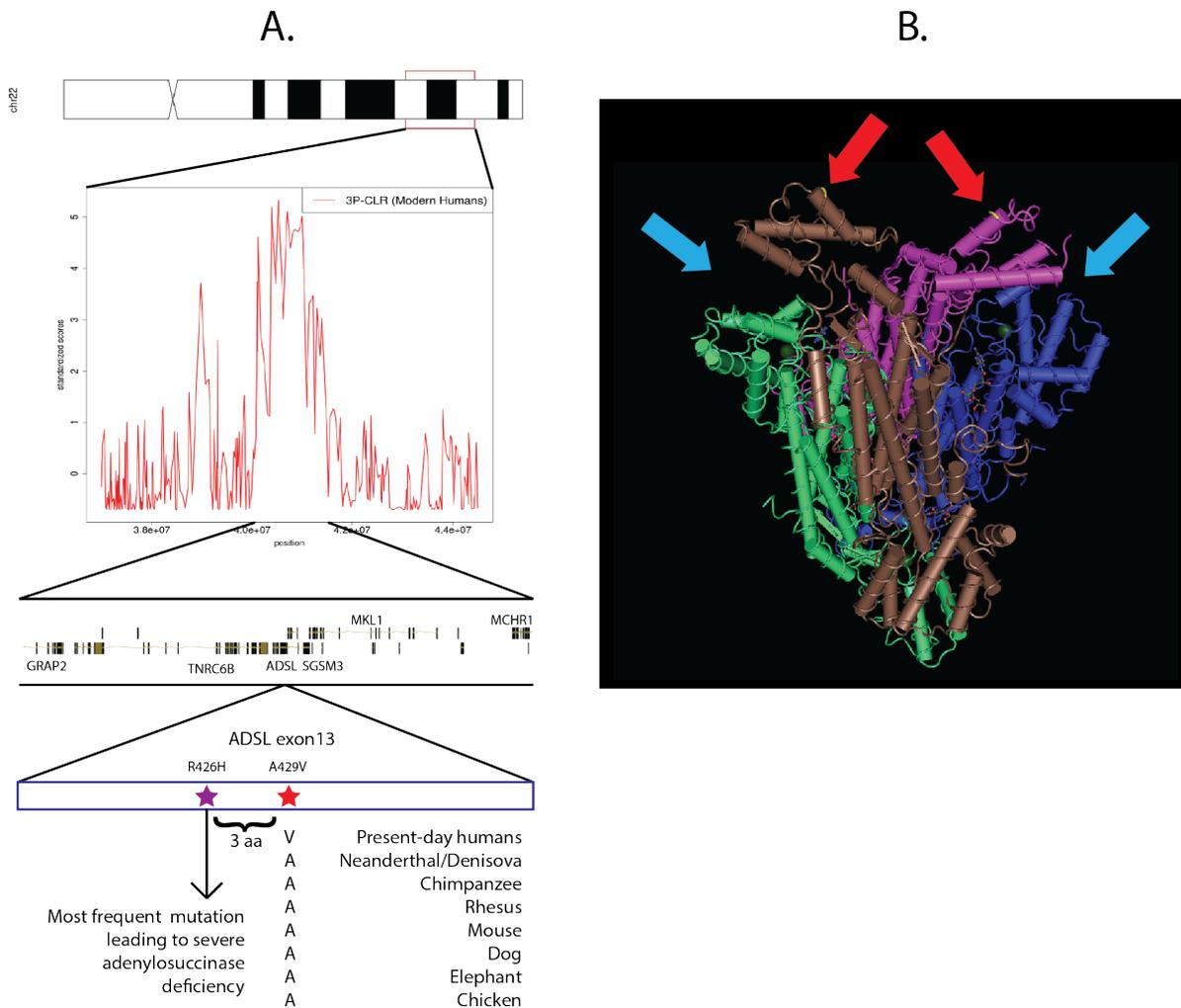


Figure 6. ADSL is a candidate for selection in the modern human lineage, after the split from Neanderthal and Denisova. A) One of the top-scoring regions when running 3P-CLR on the modern human lineage contains genes *TNRC6B*, *ADSL*, *MKL1*, *MCHR1*, *SGSM3* and *GRAP2*. The most disruptive nonsynonymous 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 3 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 tetramer (red arrows), which are near the active sites (light blue arrows). Scores in panel 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.

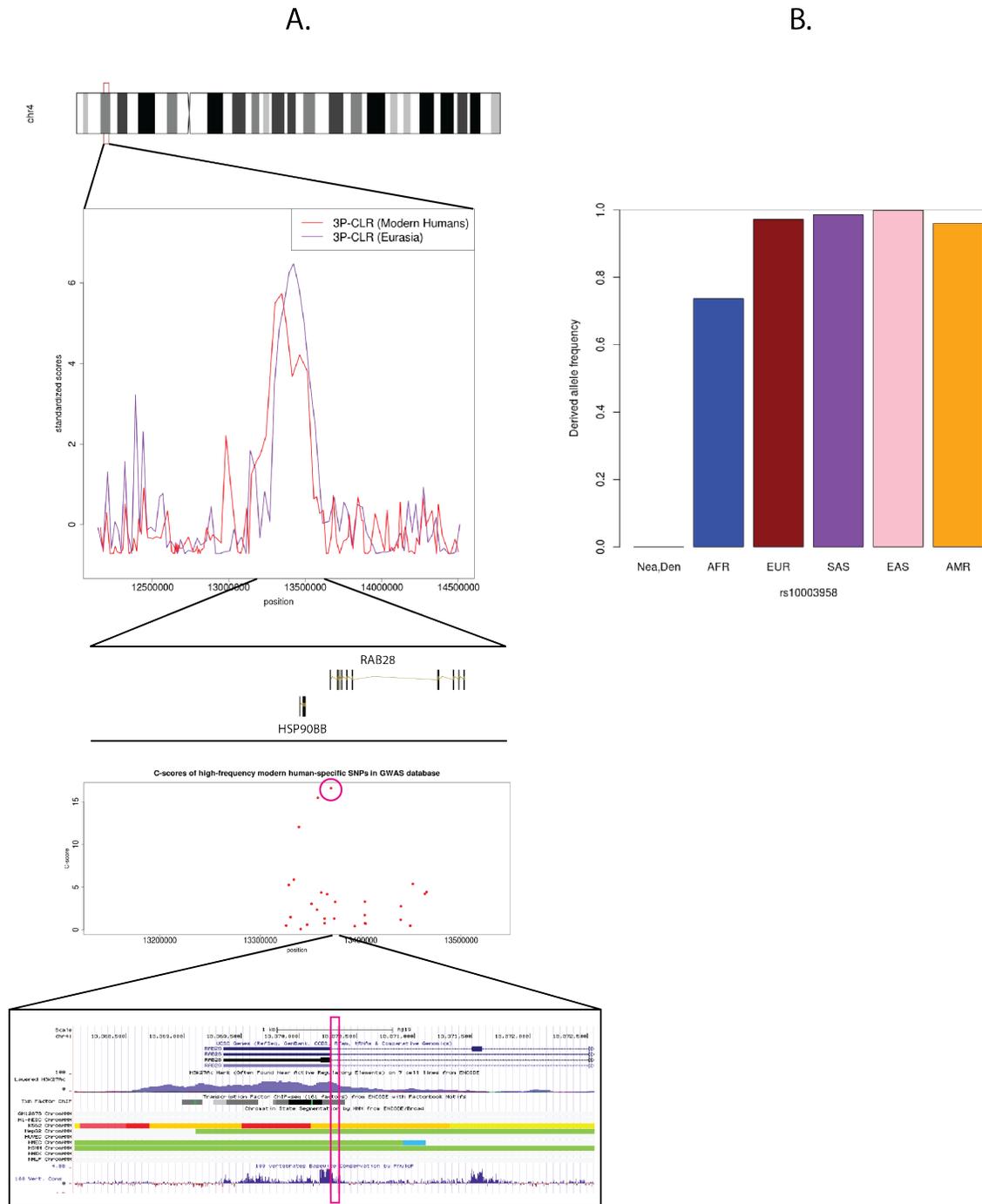


Figure 7. *RAB28* is a candidate for selection in both the Eurasian and the modern human ancestral lineages. A) The gene lies in the middle of a 3P-CLR peak for both ancestral populations. The putatively selected region also contains several SNPs that are significantly associated with obesity and that are high-frequency derived in present-day humans (> 93%) but ancestral in archaic humans (red dots). The SNP with the highest C-score among these (rs10003958, pink circle) lies in a highly conserved strong enhancer region adjacent to the last exon of the gene. Color code for ChromHMM segmentation regions in UCSC genome browser: red = promoter, orange = strong enhancer, yellow = weak enhancer, green = weak transcription, blue = insulator. The image was built using the GenomeGraphs package in Bioconductor and the UCSC Genome Browser. B) Derived allele frequencies of SNP rs10003958 in the Denisova and Neanderthal genomes, and in different 1000 Genomes continental populations. AFR = Africans. AMR = Native Americans. SAS = South Asians. EUR = Europeans. EAS = East Asians.

671 **Supplementary Figures**

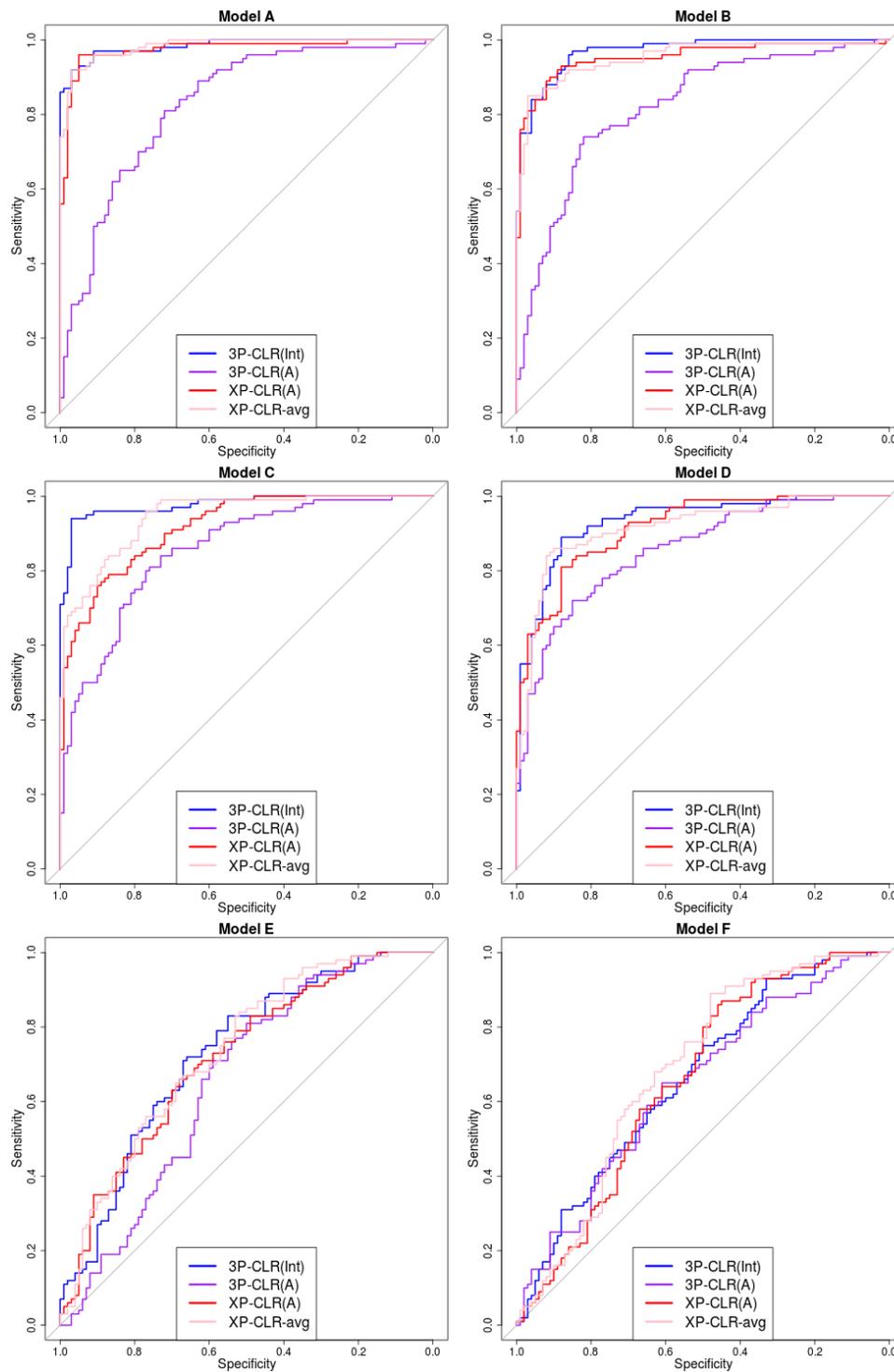


Figure S1. ROC curves for performance of 3P-CLR(Int), 3P-CLR(A) and two variants of XP-CLR in detecting selective sweeps that occurred before the split of two populations *a* and *b*, under different demographic models. In this case, the outgroup panel from population *c* contained 10 haploid genomes. The two sister population panels (from *a* and *b*) have 100 haploid genomes each.

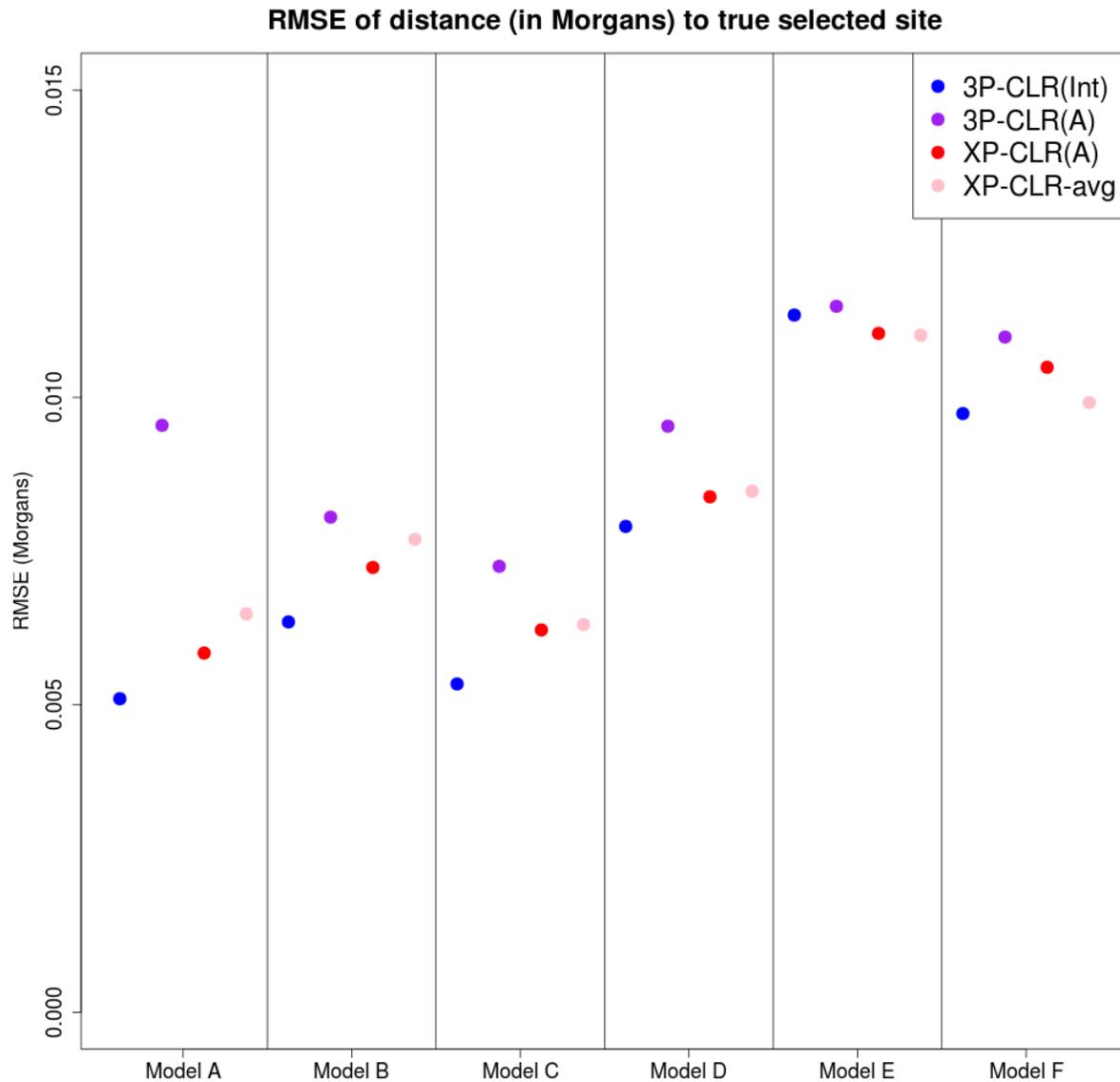


Figure S2. Root-mean squared error for the location of sweeps inferred by 3P-CLR(Int), 3P-CLR(A) and two variants of XP-CLR under different demographic scenarios, when the sweeps occurred before the split of populations a and b . In this case, the outgroup panel from population c contained 100 haploid genomes and the two sister population panels (from a and b) have 100 haploid genomes each.

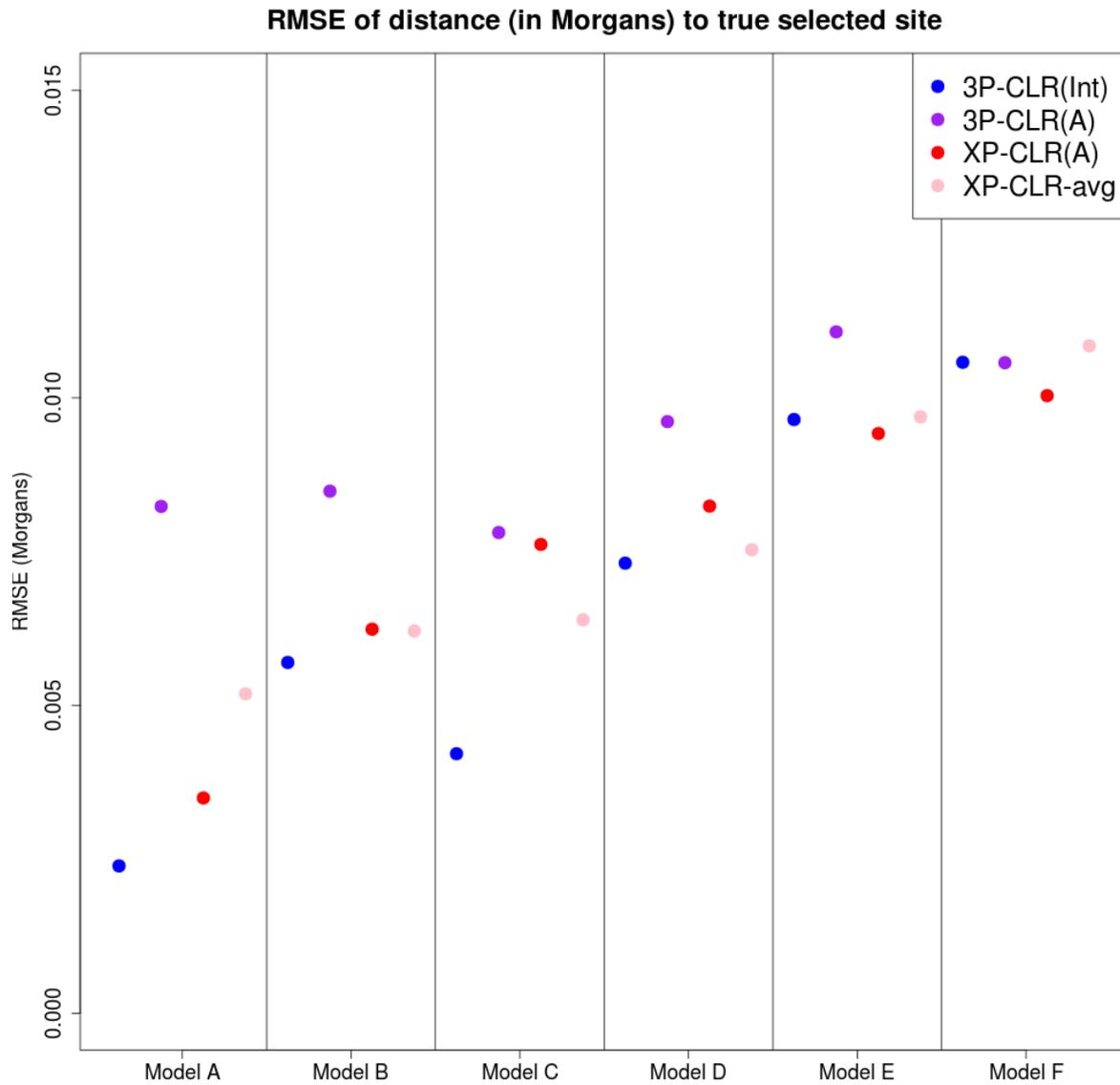


Figure S3. Root-mean squared error for the location of the sweep inferred by 3P-CLR(Int), 3P-CLR(A) and two variants of XP-CLR under different demographic scenarios, when the sweeps occurred before the split of populations *a* and *b*. the outgroup panel from population *c* contained 10 haploid genomes and the two sister population panels (from *a* and *b*) have 100 haploid genomes each.

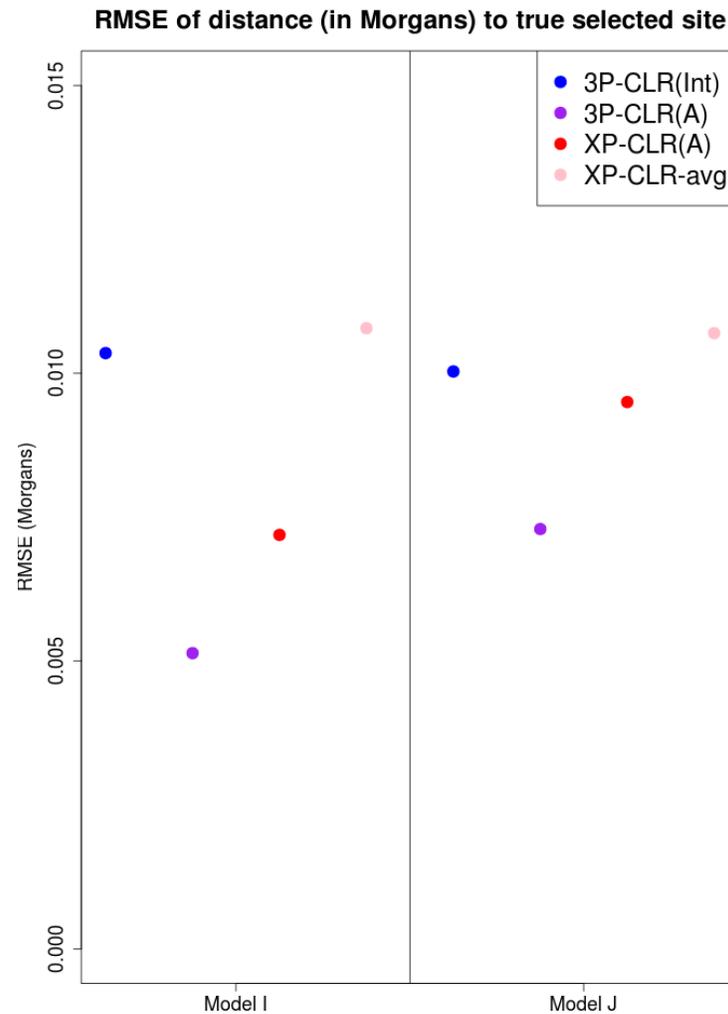


Figure S4. Root-mean squared error for the location of the sweep inferred by 3P-CLR(Int), 3P-CLR(A) and two variants of XP-CLR under different demographic scenarios, when the sweeps occurred in the terminal population branch leading to population *a*, after the split of populations *a* and *b*. In this 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.

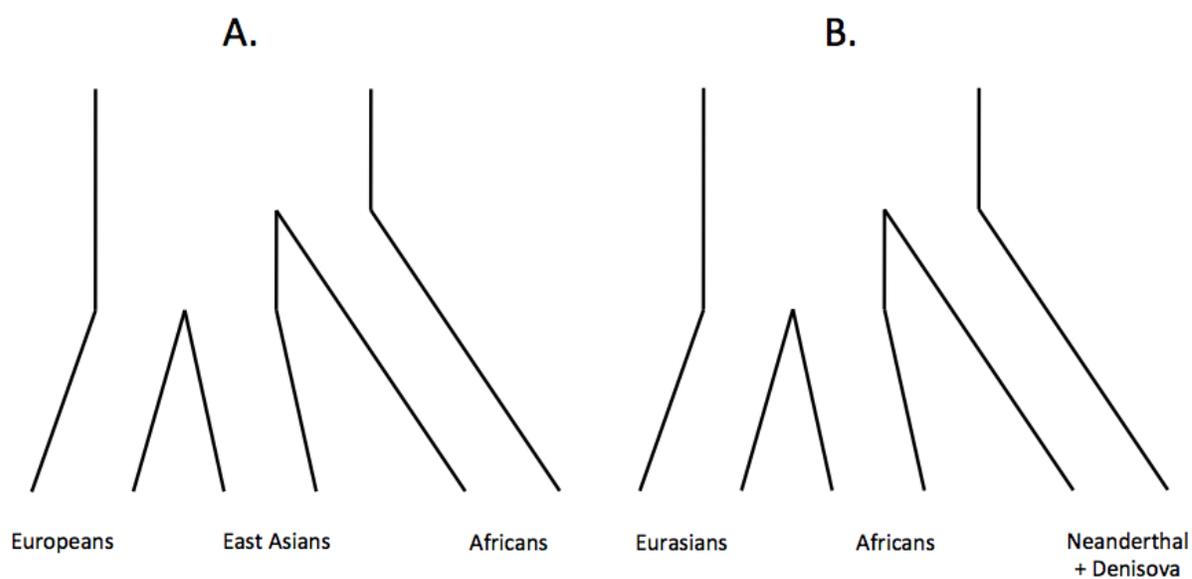


Figure S5. A. Three-population tree separating Europeans, East Asians and Africans. B. Three-population tree separating Eurasians, Africans and archaic humans (Neanderthal+Denisova).

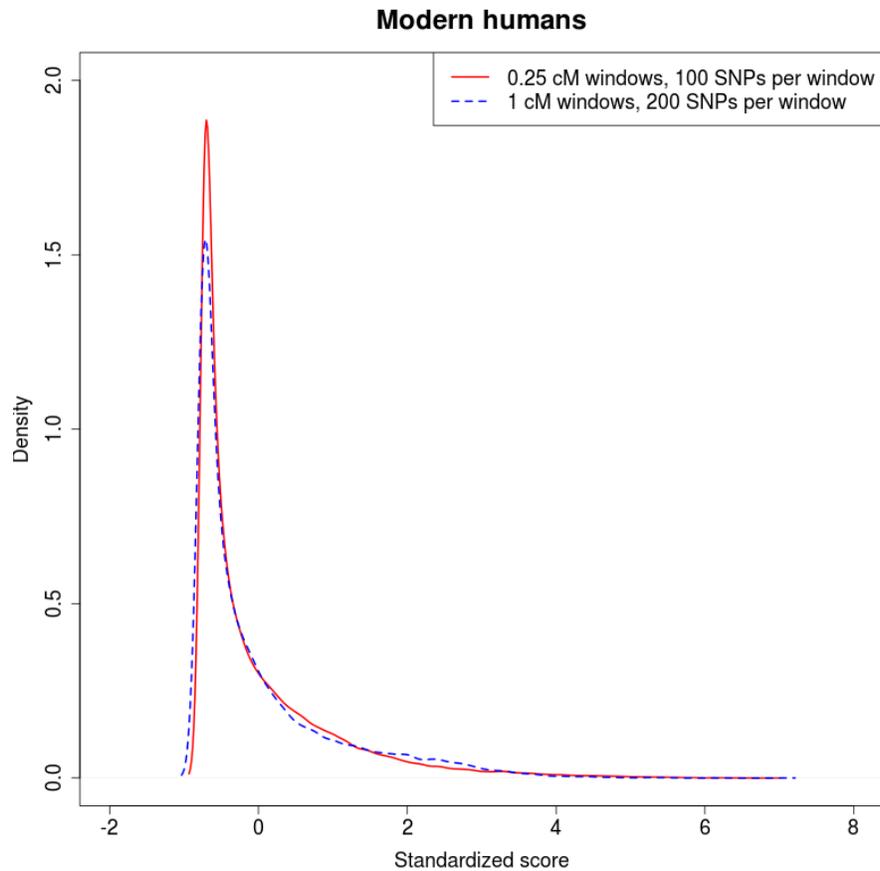


Figure S6. Comparison of 3P-CLR on the modern human ancestral branch under different window sizes and central SNP spacing. The red density is the density of standardized scores for 3P-CLR run using 0.25 cM windows, 100 SNPs per window and a spacing of 20 SNPs between each central SNP. The blue dashed density is the density of standardized scores for 3P-CLR run using 1 cM windows, 200 SNPs per window and a spacing of 80 SNPs between each central SNP.

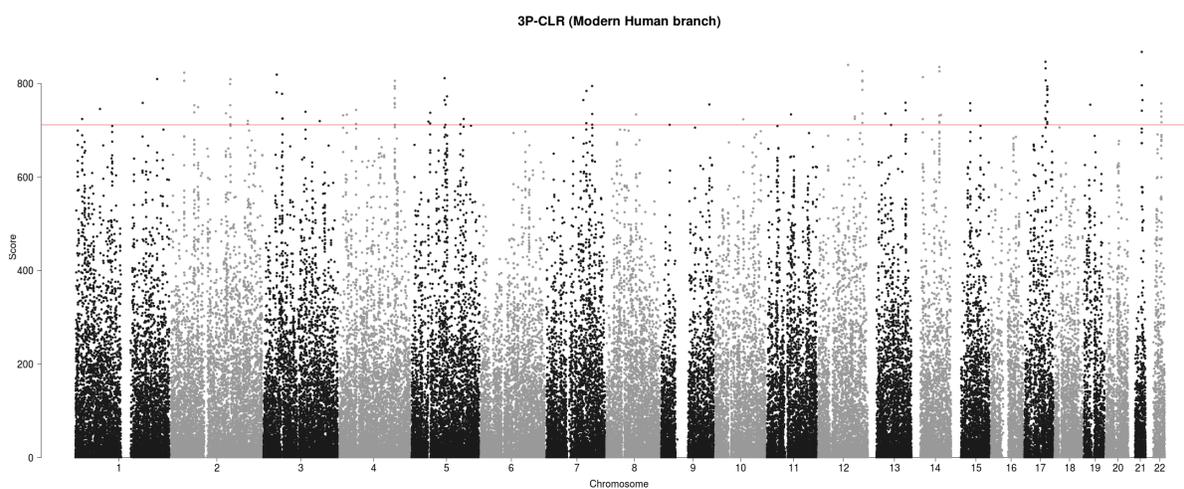


Figure S7. 3P-CLR scan of the ancestral branch to Africans and Eurasians, using the Denisovan and Neanderthal genomes as the outgroup. The red line denotes the 99.9% quantile cutoff.

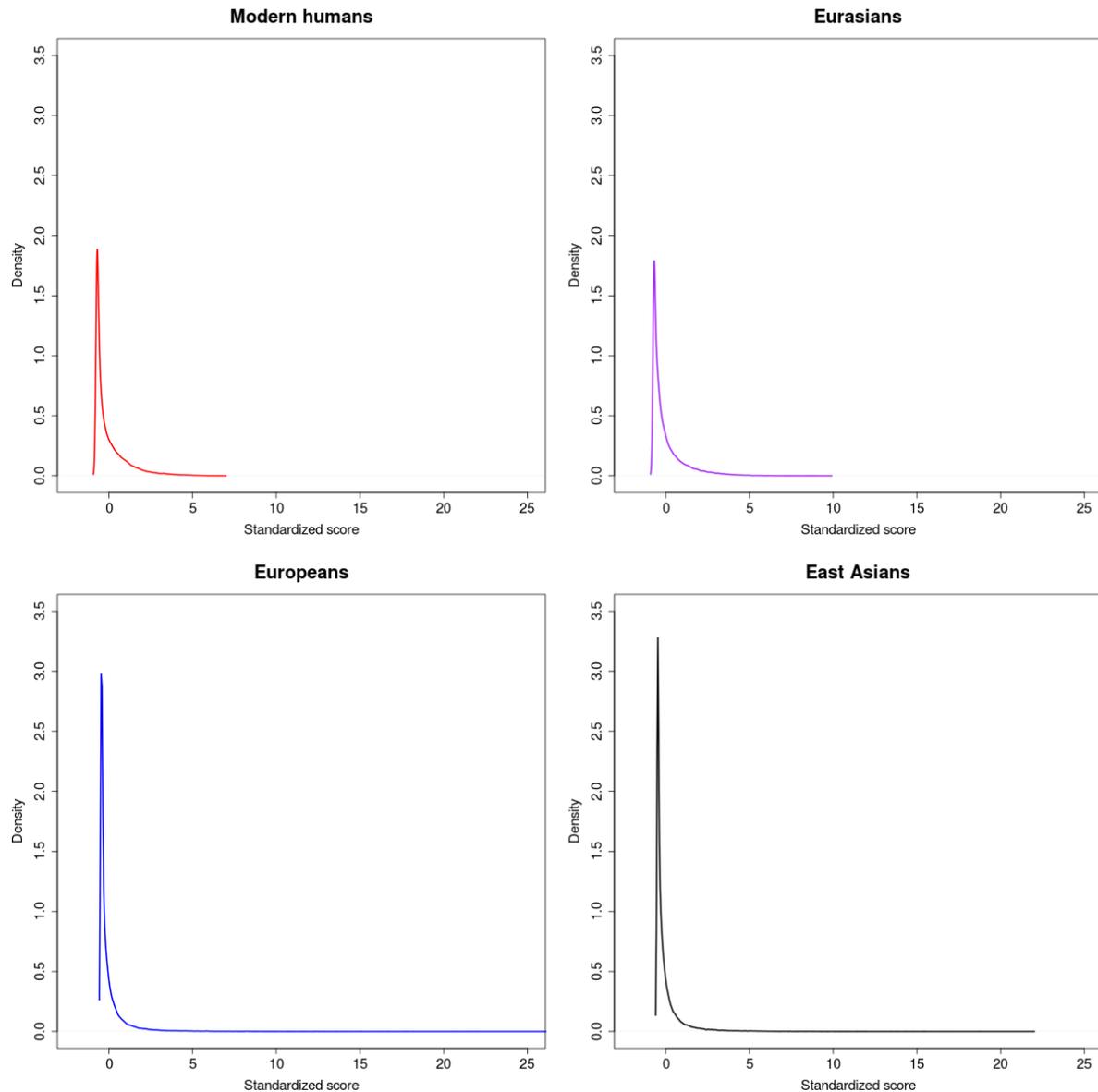


Figure S8. Genome-wide densities of each of the 3P-CLR scores described in this work. The distributions of scores testing for recent selection (Europeans and East Asians) have much longer tails than the distributions of scores testing for more ancient selection (Modern Humans and Eurasians). All scores were standardized using their genome-wide means and standard deviations.

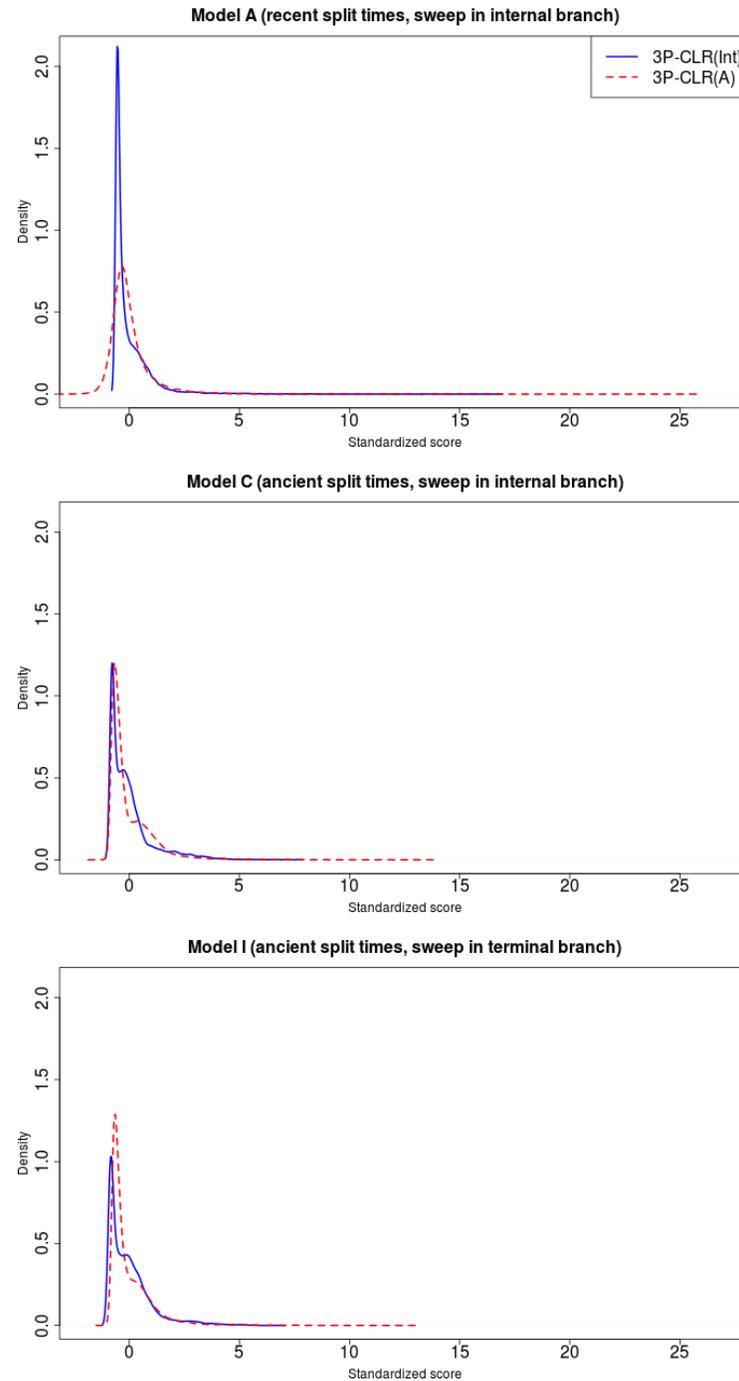


Figure S9. Distribution of 3P-CLR(Int) and 3P-CLR(A) scores under different demographic histories. We combined all scores obtained from 100 neutral simulations and 100 simulations with a selective sweep under different demographic and selection regimes. We then plotted the densities of the resulting scores. Top panel: Model A; Middle panel: Model C; Bottom panel: Model I. See Table 1 for details about each model.