Spatial and temporal diversity of positive selection on shared haplotypes at the *PSCA* locus among worldwide human populations

Risa L. Iwasaki[1,2] and Yoko Satta[1,2,*]

[1]Department of Evolutionary Studies of Biosystems, School of Advanced Science, SOKENDAI (The Graduate University for Advanced Studies), Hayama, Kanagawa 240-0193, Japan.

[2]Research Center for Integrative Evolutionary Science, SOKENDAI

[*]Corresponding Author: E-mail: <satta@soken.ac.jp>

Word count for text: 5512 words

Running title: Switching of selection target haplotype in *PSCA* locus

Abstract

Selection on standing genetic variation is important for rapid local genetic adaptation when the environment changes. We report that, for the prostate stem cell antigen (*PSCA*) gene, different populations have different target haplotypes, even though haplotypes are not unique to specific populations. The C-C-A haplotype, whereby the first C is located at rs2294008 of *PSCA* and is a low risk allele for gastric cancer, has become a target of positive selection mainly in Asian populations. Conversely, the C-A-G haplotype carrying the same C allele has become a selection target in African populations. However, Asian and African populations share both haplotypes, consistent with the haplotype divergence time (170 kya) prior to out-of-Africa. The frequency of C-C-A/C-A-G is 0.344/0.278 in Asia and 0.209/0.416 in Africa. 2D-SFS analysis revealed that the extent of intra-allelic

variability of the target haplotype is extremely small in each local population, suggesting that C-C-A or C-A-G is under ongoing hard sweeps in local populations. From the TMRCA of selected haplotypes, the estimated onset times of positive selection were recent (3-55 kya), concurrently with population subdivision from a common ancestor. Additionally, estimated selection coefficients from ABC analysis were up to ~3%, similar to those at other loci under recent positive selection on standing genetic variation. Phylogeny of local populations and TMRCA of selected haplotypes revealed that spatial and temporal switching of positive selection targets is a unique and novel feature of ongoing positive selection at *PSCA*. This switching may reflect potential of rapid adaptability to distinct environments.

Introduction

Selective sweeps can be classified into two cases, whereby the target allele is either single (classic hard sweep) or multiple (soft sweep) in a tested population (Rees et al. 2020). The former describes that a *de novo* mutation emerges and the allele soon after becomes advantageous. Many neutrality tests are designed to detect signatures of this type of classic hard sweep. In contrast, the latter, soft sweep, includes more complicated processes of positive selection. One of such processes includes a typical case where a *de novo* mutation emerges within a particular haplotype but is maintained under neutrality. Through genetic drift, the haplotype number increases and is then maintained as standing genetic variation. Environmental change may result in the onset of selection on multiple haplotypes. In humans, both *MCPH1* and *ADAM17* are such examples in East Asian populations (Satta et al.

2020). *MCPH1* is associated with brain size development (Evans et al. 2005); haplogroup D at *MCPH1* is composed of multiple haplotypes that carry the derived C allele at G37995C and is a selection target. Similarly, there are multiple target haplotypes at *ADAM17* (Satta et al. 2020). *ADAM17* is involved in the *NRG-ERBB1* signaling pathway, and variants in genes of this pathway are associated with the risk of schizophrenia and psychiatric phenotypes (Pickrell et al. 2009). Other cases of soft sweeps include repeated advantageous alleles that emerged from different genetic backgrounds (recurrent mutations), as well as gene conversion of adaptive mutations into a new genetic background (Jones and Wakeley 2008; Rees et al. 2020).

As one of the major causes of soft sweeps, standing genetic variation in a mother population is reported to contribute to rapid adaptation of a daughter population to a novel niche. In

sticklebacks, neutral polymorphism of genes in a mother population living in seawater became adaptive when a daughter population colonized from seawater to freshwater. For example, polymorphisms containing adaptive alleles at both *Eda* (Colosimo et al. 2005) and *KITLG* genes (Miller et al. 2007) associated with light skin pigmentation contributed to the adaptation to freshwater. Similarly, copy number variation in the *FADS2* gene (Ishikawa et al. 2019), which encodes an enzyme that synthesizes DHA, affects the survival rate in freshwater. Interestingly, orthologous *KITLG* in humans is also associated with light skin pigmentation. Standing genetic variation at this locus in Africa contributes to parallel adaptation for vitamin D synthesis due to the reduced UV light in both Asia and Europe (Miller et al. 2007). The common functional haplotype in both populations becomes a target of recent positive selection.

Schrider et al. (2017) scanned the human genome using a machine learning method S/HIC for detecting loci under completed positive selection; they revealed that most loci under selection are classified into soft sweeps (including standing genetic variation) rather than classic hard sweeps. Although Schrider et al. did not examine each case in detail, we expect many unexplored loci under completed/ongoing soft sweeps and that such loci may be associated with rapid adaptation to new environments.

Another approach for detecting signals of both soft and hard sweeps is the two-dimensional site frequency spectrum (2D-SFS) method. Summary statistics (F_c, G_{c0} and L_{c0}) in 2D-SFS (Fujito et al. 2018; Satta et al. 2020) evaluate the decay of intra-allelic variability (IAV) of haplotypes carrying a target site (D-group) due to selective sweep compared with that of haplotypes under neutrality (A-group). The 2D-SFS method can identify a

105 target site among different populations and distinguish signals of
106 ongoing selection from completed selection based on three
107 summary statistics. Combined with target haplotype TMRCA (t_D),
108 we can estimate the onset times of positive selection and
109 reconstruct the history of positive selection in individual
110 populations.

111 Humans migrated from Africa to most corners of the
112 Earth during a short period of at most 70 thousand years (Soares
113 et al. 2012; Haber et al. 2019) and subsequently have genetically
114 adapted to various novel environments, e.g. reduced UV light, low
115 temperature, high altitude, novel pathogens and dietary changes.
116 Many studies have been conducted to investigate associations
117 between signatures of positive selection and causal selective
118 pressures (reviewed in Rees et al. 2020). At some of these
119 signatures on target loci, onset times of positive selection are

120 inferred to be relatively recent (within ~50 thousand years). In
121 some cases, the same allele parallelly became a selection target in
122 genetically distinct populations (Xue et al. 2006; Yang et al. 2018).
123 We previously detected three shared haplotypes of the
124 prostate stem cell antigen (*PSCA*) gene in two East Asian
125 populations, Japanese in Tokyo (JPT) and Han Chinese in Beijing
126 (CHB) (Iwasaki et al. 2020). Two haplotypes carry a C allele at
127 rs2294008 (T/C) in the first position (C-C-A/C-A-G haplotype) and
128 are associated with low risk of gastric cancer (The Study Group of
129 Millennium Genome Project for Cancer 2008; Tanikawa et al.
130 2012), and the other haplotype carrying a T allele (T-C-A) is
131 associated with high risk of gastric cancer. Among the three,
132 C-C-A/C-A-G haplotypes were found to be under ongoing positive
133 selection. A signal of ongoing hard selective sweep was detected in
134 the C-C-A haplotype alone in JPT. In contrast, such signals were

detected in both C-C-A and C-A-G haplotypes in CHB, which is suggestive of an ongoing soft selective sweep in CHB. The three haplotypes are shared not only between JPT and CHB, but are also found worldwide including African populations. This supports that the two selected haplotypes, C-C-A and C-A-G, diverged before out-of-Africa. Indeed, the divergence time of the two haplotypes is estimated to be 170 kya (thousand years ago) (Albers and McVean 2020). We hypothesize that either or both of these haplotypes are under selection in different subpopulations in the 1000 genomes project (1KGP). To test this hypothesis, we apply the 2D-SFS method and examine the extent of IAV and t_D (TMRCA). Furthermore, using an ABC framework, we estimate the onset times of selection as well as selection coefficients. Based on these results, we discuss unique features of this selection.

Materials and Methods

Test populations

We retrieved variant call format (VCF) data of 26 subpopulations in five superpopulations (East Asian (EAS), European (EUR), South Asian (SAS), American (AMR), and African (AFR)) from 1KGP Phase 3 (The 1000 Genomes Project Consortium 2015). We followed each population code in the definition of 1KGP and used the GRCh37 genomic positions for each SNP. The ancestral/derived alleles are defined in 1KGP.

Neutrality test with Two-Dimensional Site Frequency Spectrum (2D-SFS) and TMRCA of target haplotype

We defined 2D-SFS with $\varphi_{i,j}$, which is the sum of the number of SNPs whose configuration is i derived alleles in the D-group and j derived alleles in the A-group (Fujito et al. 2018; Satta et al. 2020). We evaluated IAV with a particular frequency of

target allele by using $\varphi_{i,j}$. This method has the advantage of being able to detect incomplete and complete selective sweeps. We defined the core region for detection of selection by 2D-SFS in the following steps: we calculated r^2 with rs2294008 per SNP and the average r^2 in each 1000 bp non-overlapped sliding window, and then determined the boundaries of the core region based on SNPs with $r^2 > 0.75$. For one of the African subpopulations, MSL, the size of the core region determined by the above method was too short (2000 bp) for analysis; thus, we simply defined the boundaries of the core region using SNPs with $r^2 > 0.75$ without calculating average r^2.

We calculated Q-values with Benjamini-Hochberg procedure (Benjamini and Hochberg 1995) to archive FDR controls for each of the summary statistics of 2D-SFS (F_c, G_{c0} and L_{c0}) and evaluated their significances (Q-value < 0.05). If there were more

than one significant summary statistic, we defined this as a signal of positive selection. We further classified this signal into completed or ongoing: significant values in G_{c0} and L_{c0} but not in F_c were interpreted as completed positive selection based on Satta et al. (2020), and otherwise classified as ongoing positive selection. Additionally, in the cases where only one summary statistic showed a significant value, we further calculated a combined P-value (Brown 1975) for the three summary statistics, and a highly significant combined P-value ($P < 0.01$) was also defined as positive selection.

Null distributions of the three summary statistics were generated using ms (Hudson 2002) under neutrality with demographic models based on Schaffner et al. (2005). From the simulated data, we chose at least 1000 replicates of simulated/pseudo target sites that had similar frequency to the

195 true/observed frequency of the target site.

196 Additionally, we estimated TMRCA (t_D) of the target
197 haplotype (D-group). The value of ut_D is the average number of
198 substitutions per lineage in the D-group that is $L_{\zeta 0}/m$, where $L_{\zeta 0} =$
199 $\sum_{k=1}^{m-1} k\varphi_{k,0}$, m is the number of target alleles or haplotypes and u is
200 the mutation rate per region per year (Satta et al. 2020). We
201 applied 0.5 x 10^{-9}/site/year for mutation rate (Scally and Durbin
202 2012) to calculate t_D. For measuring t_D in generations, the
203 generation time is 29 years (Fenner 2005).

204 A framework of ABC for inferring onset of positive selection and
205 selection coefficient (*s*)

206 We performed two simulation steps to obtain estimates of
207 onset times of positive selection and selection coefficients (s) for
208 three populations (JPT, YRI and CHB): (i) we generated samples
209 using mssel implemented with positive selection (dominance

coefficient (h) was set as 0.5) and accepted 10,000 replications with target site frequencies within ± 5% of observed target site frequency (Nakagome's ABC procedure (Nakagome et al. 2019)); (ii) of these 10,000 replications, we accepted replications with summary statistics (F_c, $\gamma(10)$, $\gamma^*(10)$ and imax) that were within ± 1% of observed values in a test population. Finally, we estimated 95% CIs of s and onset times of positive selection from accepted replications.

C-C-A/C-A-G haplotypes are shared among all subpopulations studied (including African populations); this suggests that frequencies of these haplotypes should be higher than 1/2N, where N is effective population size of each local population. Therefore, we used a SSV (standing variant) model of mssel, but the uniform distribution (the prior distribution of allele frequency at selection onset (ft)) in the original model was modified

225 with a simulated allele frequency distribution without
226 recombination (see below).

227 We simulated the prior distribution of ft with the
228 following procedure. We first defined the three haplotypes (C-C-A,
229 C-A-G and T-C-A) based on three SNPs (rs2294008 (T/C),
230 rs2976391 (C/A) and rs2978983 (A/G)). We then estimated the
231 order of emergence of the three haplotypes based on
232 derived/ancestral states and the divergence times of the three
233 SNPs. The divergence times of the SNPs were estimated from the
234 atlas of variant age (Albers and McVean 2020) (Fig. 1). C-C-A
235 diverged from T-C-A 1,065,141 ya (36,729 gen. ago) as a *de novo*
236 haplotype; then, C-A-G with two mutations (a derived A allele at
237 rs2976391 (168,432 ya) and a derived G allele at rs2978983
238 (171,999 ya)) diverged from C-C-A 168,432 ya (5,808 gen. ago).
239 Then, we simulated the frequency of a target haplotype for 10,000

240 replications, considering the above evolutionary process and
241 demography of each test population based on Schaffner et al.
242 (2005). The 10000 replications did not include cases where any
243 haplotypes were fixed or lost before the onset of positive selection.
244 We generated a distribution of target allele frequency at the onset
245 of positive selection respectively at $^{t_D}/_{29}$ generations ago. For
246 JPT and YRI, we generated a distribution of each single target
247 haplotype (C-C-A or C-A-G) frequency. In contrast, for CHB, we
248 treated C-C-A and C-A-G as a single C haplotype. Then, the
249 frequency of the C haplotype was set based on the summated
250 frequencies of both haplotypes. We used t_D of the C haplotype in
251 CHB as 28,225 ya. We divided haplotype frequency distributions
252 into bins from 0.0 to 1.0 with increments of 0.01 and drew a
253 histogram as the ft prior distribution. Replications did not accept
254 any case where ft exceeded observed target allele frequency in each

test population (Nakagome et al. 2018).

Results

Different selection target among populations genetically closely related

Reduction in genetic diversity in the D-group (with core region strongly linked with a target allele) is due to selective sweep. This trait is exploited to detect signatures of positive selection on putative target alleles with the 2D-SFS method (Satta et al. 2020). Summary statistics of 2D-SFS are robust against violation of genetic diversity reduction caused by recombination (Iwasaki et al. 2020; Satta et al. 2020), and thus they can detect older signals of positive selection compared to haplotype-based tests (e.g. nS_L) (Oleksyk et al. 2010; Ferrer-Admetlla et al. 2014). We applied the 2D-SFS method to 1KGP data to detect signals of positive selection

on C-C-A and C-A-G haplotypes in different subpopulations.

No signals are detected on the two haplotypes in most of the AMR and EUR subpopulations (Table1). The frequency of the C-C-A haplotype is too low (<10%) in all subpopulations to detect signals. Especially, the C-C-A haplotype in both FIN and GBR show no private alleles in the D-group. Therefore, we conclude that all subpopulations in AMR and EUR have no signals of selection acting on the C-C-A haplotype, even though the combined P-value is marginally significant (e.g. GBR for P = 0.017). The C-A-G haplotype in PEL shows low values in F_c (Q = 0.002) and combined P (P = 0.016); PEL allele configuration shows two SNPs (rs75914363: $\varphi_{87,0} = 1$; rs2920288: $\varphi_{92,0} = 1$) that have similar number of derived alleles as the C-A-G haplotype (93 haplotypes). Consequently, this results in a low F_c value but high G_{c0} and L_{c0} values (Satta et al. 2020) and a significant combined P value.

Therefore, the presence of $\varphi_{87,0}$ and $\varphi_{92,0}$ is considered to be due to missampling of recombinants, $\varphi_{87,j}$ and $\varphi_{92,j,}$ where j is a positive integer. We conclude that the signal from PEL in C-A-G is false positive. The C-A-G haplotype in PUR and CEU show significant values in L_{c0} alone, but the former shows a non-significant combined P value (P = 0.052); the latter shows a marginally significant combined P value (P = 0.046). Both signals in PUR and CEU can not be classified into any of the defined sweeps (Satta et al. 2020) (see Materials and Methods).

In non-AMR/EUR subpopulations, signals of selection are detected in either or both haplotypes. In all AFR subpopulations, ongoing positive selection on the C-A-G haplotype is detected. Values of imax, γ(10) and γ*(10) are supportive of hard selective sweeps in AFR. Additionally, ESN and YRI show significant L_{c0} values only in the C-C-A haplotype. ESN also shows a marginally

300 significant combined P value (P = 0.040). For the same reason as in
301 CEU, we do not regard this case as a positive selection signal. In
302 contrast, YRI shows a non-significant combined P value (P =
303 0.062).

304 In SAS subpopulations, we observe three patterns of
305 selection signals. First, a signal of ongoing positive selection on the
306 C-C-A haplotype only in BEB is detected. Values of imax, $\gamma(10)$ and
307 $\gamma^*(10)$ in BEB are indicative/supportive of a hard selective sweep.
308 Second, no signals for selection on any of the haplotypes in GIH
309 and PJL are detected. Third, both STU and ITU show
310 non-significant values of F_c but significant values of G_{c0} and L_{c0} in
311 the C-C-A haplotype. This observation may imply that completed
312 positive selection has acted on the C-C-A haplotype in both STU
313 and ITU (see Materials and Methods).

314 In contrast, EAS subpopulations show two patterns of

selection signals. We detect selection signals on C-C-A in JPT, CHS, KHV and CDX, and we find signals acting on both C-C-A and C-A-G haplotypes in CHB (Table 1). Based on the imax, $\gamma(10)$ and $\gamma^*(10)$ values, these signals are all indicative of hard selective sweeps. For the C-A-G haplotype in CHS, there is only a significant L_{c0} value ($Q = 0.042$) and a marginally significant combined P value ($P = 0.032$). For the same reasons as in CEU, we do not regard this as a positive selection signal. IAV of both haplotypes in CHB show signals of positive selection. Therefore, we conclude that the selective sweep acting on the two haplotypes in CHB appears to be soft. Variations in the pattern of selection signals are observed especially in Asian populations. This suggests that the target of selection can differ among subpopulations, even if they are genetically close and their habitat environments are seemingly similar.

Inference of positive selection onset times

We estimate TMRCA of the D-group ($t_D \pm SD$) of the C-C-A or C-A-G haplotype (Fig. 2 and Table 1). TMRCA is the time for all D-group sequences to coalesce to a common ancestor (Satta et al. 2020). When the target haplotype is novel within a test population (hard selective sweep), the estimated t_D is equivalent to the onset time of positive selection or earlier due to the waiting time of the *de novo* mutation (Satta et al. 2020). The values of t_D of C-C-A haplotypes are limited within 8-30 kya (BEB: 8,130 ya for the minimum, CHB: 30,063 ya for the maximum) and those of C-A-G haplotypes are limited within 20-40 kya (YRI: 19,540 ya for the minimum, LWK: 38,596 ya for the maximum) in all subpopulations. The values of $t_D \pm SD$ are limited within 3-42 kya for C-C-A and 13-55 kya for C-A-G. We also estimate the onset time for JPT/CHB/YRI using the ABC framework with samples generated

345 with mssel (Nakagome et al. 2019) implemented with positive
346 selection and compare them with the above t_D values.

347 Mean onset times estimated by the ABC framework (ABC
348 onsets) are 7,334 ya (C-C-A haplotype in JPT) and 13,845 ya
349 (C-A-G haplotype in YRI). In CHB, two haplotypes have very
350 similar t_D values and they show significant low F_c and L_{c0} values in
351 2D-SFS (Fig. 2 and Table 1); therefore, we regard both C-C-A and
352 C-A-G as a single haplotype (C haplotype) carrying a derived (C)
353 allele at rs2294008. The mean ABC onset in CHB is estimated to
354 be 18,174 ya (Table 2 and Fig. 4A). All $t_D \pm$ SD values are within
355 the range of the 95% CI of ABC onsets.

356 Additionally, we infer selection coefficients using the ABC
357 framework. Estimated means of s are 1.89% for C-C-A in JPT,
358 2.90% for C-A-G in YRI and 1.94% for the C haplotype in CHB; all s
359 values are lower than 3% (Table 2 and Fig. 4B).

Discussion

Transition process of target haplotypes of positive selection

We estimate the number of transitions of target haplotypes along the human lineage in a parsimonious manner, considering t_D values and current target haplotypes in subpopulations together with the phylogeny based on pairwise F_{ST} values among 1KGP subpopulations (The 1000 Genomes Project Consortium 2015) (Fig. 3). We assume that positive selection did not act on any haplotypes in the human common ancestor (before out-of-Africa), as t_D values of selected haplotypes are more recent than out-of-Africa. The estimation results suggest that positive selection on target haplotypes would cease/relax and operate several times within each subpopulation or lineage, leading to the extant subpopulations after population splits. For example, for

375 EUR and AMR, although they possess the two haplotypes, no
376 signatures of positive selection in *PSCA* are detected. In contrast,
377 the C-A-G haplotype in African subpopulations has been selected
378 independently four times (Fig. 3), and positive selection target
379 switched between C-A-G and C-C-A haplotypes multiple times in
380 Asian subpopulations (see below).

381 In EAS subpopulations, target haplotypes are not the
382 same between JPT and CHB, despite these populations being
383 genetically close. The maximum t_D of EAS is 30063 ya for the
384 C-C-A haplotype and 27,027 ya for the C-A-G haplotype in CHB
385 (Fig. 2). The Jomon population is one of the ancestral populations
386 to the Japanese that split off from the stem lineage leading to the
387 extant East Asian populations around 18-38 kya (Wang et al. 2018;
388 Kanzawa-Kiriyama et al. 2019; Gakuhari et al. 2020). This period
389 overlaps with the $t_D \pm$ SD (13-42 kya) of the C-C-A and C-A-G

haplotypes in CHB, suggesting that target haplotypes have been maintained from an East Asian common ancestor. In contrast, we do not detect any signals of positive selection on the C-A-G haplotype in JPT/CHS/CDX/KHV (Fig. 3 and Table1). This suggests that target haplotypes have transitioned from two haplotypes in a common ancestor (C-C-A and C-A-G) to a single haplotype (C-C-A) in all four populations. The phylogenetic relationship among CDX/KHV, CHS and JPT implies transitions are independent events in the three lineages.

Different from EAS subpopulations, only the C-C-A haplotype has been a target of selection in SAS. We detect a signal of ongoing positive selection in BEB and signals of completed positive selection in ITU and STU, whereas no selection signals are found in PJL and GIH. Parsimoniously considering these observations with the topology of the phylogeny (Fig. 3), selection

405 on the C-C-A haplotype may have started to act in a common
406 ancestor of BEB/ITU/STU. Then, positive selection was
407 relaxed/ceased in a common ancestor of ITU and STU. In contrast,
408 selection is still present in the extant BEB. The inferred
409 divergence times of these SAS subpopulations have not been
410 determined so far, thus making it difficult to discuss transition
411 events of target haplotypes in detail.

412 In all AFR subpopulations, C-A-G is a common target
413 haplotype. African demography is complex (Busby et al. 2016) and
414 the examined African subpopulations share either or both western
415 or/and eastern African ancestry with each other (Lachance et al.
416 2012; Excoffier et al. 2013; The 1000 Genomes Project Consortium
417 2015; Fan et al. 2019). One of the western African populations, YRI,
418 and an admixed population of western African and eastern African,
419 LWK (Henn et al. 2011; Excoffier et al. 2013; Gurdasani et al. 2015),

420 diverged 100-300 kya (Terhorst et al. 2017). Although 1KGP did
421 not cover subpopulations in eastern/southern Africa and there
422 missing data of divergence times of some African subpopulations,
423 the divergence time between LWK and YRI indicates that African
424 subpopulations diverged earlier than the onset timings of positive
425 selection (t_D ± SD: 13-55 kya). The divergence time of the two
426 subpopulations and the timing of positive selection parsimoniously
427 suggest that positive selection on C-A-G in African subpopulations
428 may have proceeded in parallel (Fig. 3).

429 Target haplotypes are different among subpopulations,
430 revealing that adaptive alleles in each subpopulation are not
431 always the same. Phylogenetic and t_D analyses suggest that an
432 adaptive allele in a subpopulation may change temporally and also
433 differ spatially to other subpopulations. This change may follow
434 some form of environmental transition. In both AFR

435 subpopulations and CHB, the C-A-G haplotype is the target of
436 selection and the values of t_D are limited to more recent times,
437 after out-of-Africa and subsequent population subdivision. This
438 suggests that global environmental transitions may have caused a
439 switch of selection operation among distinct subpopulations. In
440 order to elucidate the global and local selective forces and to
441 reconstruct the process of target haplotype transition, we need to
442 survey selection signals in additional populations in the future. In
443 particular, to determine the target haplotypes under selection in
444 other populations and when target changed in other populations
445 considering the EAS and Jomon split in the phylogeny, EAS
446 subpopulations not included in 1KGP should be examined.
447 Furthermore, we need to examine the association between global
448 environmental change and the onset timing of positive selection.

449 The relationships between t_D and inferred onsets of positive selection

450 Using the approach by Satta et al. (2020), the timing of
451 positive selection onset by calculating TMRCA of the D-group ($t_D \pm$
452 SD) is estimated as 10,811 ± 8,058, 28,266 ± 9,300 and 19,540 ±
453 5,747 for C-C-A in JPT, the C haplotype in CHB and C-A-G in YRI,
454 respectively (Fig. 2). We compare t_D ± SD values with inferred
455 onsets of positive selection from the ABC framework. Given the ft
456 (allele frequency at selection onset) and current observed
457 frequency, the ABC framework can estimate ABC onsets (i.e. the
458 timing of positive selection operation) through simulations
459 (Nakagome et al. 2019). As a result, for JPT, CHB and YRI, ranges
460 of ABC onset 95% CI are wide enough to cover the respective t_D ±
461 SD values (Table 2 and Fig. 4A). We also compare t_D ± SD with
462 ABC onset 95% CI from Nakagome's original ABC framework
463 based on uniform distribution of ft (see Materials and Methods).

Although ranges of the ABC onset 95% CI cover the respective $t_D \pm$ SD values (data not shown), difference between the range of ABC onset 95% CI and $t_D \pm$ SD is reduced if we use a prior ft distribution that is generated from simulation of ft rather than a uniform distribution. It is concluded that $t_D \pm$ SD is supported by ABC onset 95% CI regardless of prior distributions of ft, suggesting that the ranges of $t_D \pm$ SD are robust.

Comparison of selection coefficients with those of other genes under selective sweeps

Selection coefficients (s) of the two haplotypes range from 1.9-2.9% in JPT/CHB/YRI (Table2). Since the t_D of both haplotypes implies that positive selection likely acted recently, we first compare these s values to that of two representative genes under recent positive selection characterized to be involved in gene-culture coevolution: *LCT*, which is associated with lactase

479 persistence among people of European ancestry (Gerbault et al.
480 2009; Itan et al. 2009; Peter et al. 2012), and *ADH1B*, which is
481 associated with alcohol metabolism in CHB (Peter et al. 2012). The
482 s of the C-C-A/C-A-G haplotypes are found to be equivalent to or
483 smaller than s of *LCT* ($s \leq 9.6\%$) and *ADH1B* ($s = 3.6\%$).

484 Next, we compare the s values with those of seven
485 representative genes involved in responses to the outside
486 environment and related to population differentiation. For *TRPM8*,
487 a gene where selection acts in parallel among non-African
488 populations (Key et al. 2018) and is associated with an endogenous
489 response to cold temperature, s is estimated to be ~1.5%; this is
490 equivalent to the s of the two haplotypes. For the genes associated
491 with skin pigmentation or morphology, most of the s values are
492 equivalent to those of the two haplotypes: 1-2% for *KITLG*
493 independently in East Asian and European populations (Beleza et

494 al. 2013), 2-3% for *TYRP1* in European populations (Beleza et al.
495 2013), 4-5% for *SLC45A2* in European populations (Beleza et al.
496 2013) and 2.9% for *ASPM* in European populations (Peter et al.
497 2012). In contrast, s values at *EDAR* ($s \leq 14\%$; associated with
498 hair and teeth morphology in East Asian populations (Kamberov et
499 al. 2013; Peter et al. 2013)) and *SLC24A5* (s = 8-16%; associated
500 with light skin pigmentation in European populations (Beleza et al.
501 2013)) show much larger values than those of the two haplotypes.

502 Finally, we assume that immune genes, which are
503 associated with susceptibility to diseases, are under strong
504 selective pressures, and thus, compare the s values of four
505 representative immune genes with those of the two haplotypes.
506 The s values of the two haplotypes are higher than most of those of
507 *MHC* (s = 0.03-4%), which play a major role in immune responses
508 (Yasukochi et al. 2013). A *CASP12* variant that is resistant to

509 severe sepsis and under positive selection shows smaller values (s
510 = 0.5-1% (Wang et al. 2006; Xue et al. 2006)) in human populations
511 than that of the two haplotypes. In contrast, genes associated with
512 resistance to malaria and under positive selection in African
513 populations show larger s values (4.3% for *DARC* (McManus et al.
514 2017) and 20% > s > 10% (Saunders et al. 2005) or 4.4% (Tishkoff et
515 al. 2001) for *G6PD*) than those of the two haplotypes. Therefore,
516 the s values of the two haplotypes do not display conformation to
517 any specific group of genes with particular function. Nevertheless,
518 the s values of the two haplotypes (s = 1.9-2.9%), which emerged
519 from standing genetic variation and under recent selection, are
520 similar to s of another gene that also has haplotypes from standing
521 genetic variation and under recent selection at the same time in
522 distinct populations, *KITLG* (s = 1-2%) (Beleza et al. 2013).
523 Comparison of onsets with those of other genes under selective

sweep

Peter et al. (2012) assumed the C allele at rs2294008 as a single origin target allele and estimated onset age of selection and *s* in YRI. Their estimates for the onset age as 8000 ya (95%CI: 1000-54900) and *s* = 3.5% (95%CI: 0.4-15%) are in concordance with our estimations of the C-A-G haplotype in YRI.

Smith et al. (2017) also estimated the onset age of selection of target alleles at five genes that are under recent ongoing positive selection but are not fixed in any populations of 1KGP. The study concluded that positive selection has acted in most human populations 50 kya or more recently. The values of t_D ± SD (3-55 kya) for the selected C-C-A/C-A-G haplotypes overlap with the range of estimations by Smith et al. (2017). This suggests that the two haplotypes are also associated with recent differentiation among human populations.

539 Temporal and spatial switching of target haplotype at *PSCA*

540 We consider that both C-C-A and C-A-G haplotypes each
541 have a single origin and then spread by human expansion. The two
542 haplotypes are locally under hard sweep in most subpopulations in
543 EAS/SAS and AFR or soft sweep in CHB, although they are shared
544 among all subpopulations tested.

545 Standing genetic variation contributes to rapid
546 adaptation to environmental changes (Colosimo et al. 2005; Miller
547 et al. 2007; Ishikawa et al. 2019). In *PSCA,* the two haplotypes are
548 shared among all subpopulations tested. The observed major
549 target allele in AFR is the C-A-G haplotype, whereas that in
550 EAS/SAS is the C-C-A haplotype. Haplotypes derived from
551 standing genetic variation respond differently and rapidly to the
552 environments that each population encounters. This suggests that
553 a particular local haplotype is not always adaptive among global

554 human populations and such local adaptive haplotype(s) may not
555 always be the same or only adaptive haplotype(s) in other
556 subpopulations.

557 We also observe that target alleles can change within a
558 lineage. For example, a target allele was both C-C-A and C-A-G in
559 a common ancestor of EAS, but it changed to only C-C-A in
560 JPT/CHS/CDX/KHV. This suggests that local 'adaptive' alleles can
561 differ temporally and spatially.

562 In general, the same alleles emerged from standing
563 genetic variation (e.g. *KITLG* in European/East Asian populations
564 (Yang et al. 2018)) or independent *de novo* adaptive alleles in
565 distinct populations (e.g. *LCT* in African/European populations
566 (Tishkoff et al. 2007; Schlebusch et al. 2013)) are positively
567 selected in many cases. In such cases, the selective force may be
568 common among populations. We report that target haplotypes,

569 C-C-A and C-A-G, switched temporally and spatially. Switching of
570 target haplotypes may occur in future adaptations; therefore, these
571 haplotypes may contribute as genetic sources of adaptation. Such
572 target allele switching would provide a chance for individuals in a
573 population to respond to a rapidly changing environment. We
574 unveiled a unique and novel feature of positive selection in the
575 *PSCA* locus and this study contributes to uncovering genetic
576 diversity from the viewpoint of selection targets.
577

578 ## Acknowledgements

579 The authors would like to thank Drs. Shigeki Nakagome, Naoki
580 Osada, Hie Lim Kim, Tatsuya Ota and Jun Gojobori for valuable
581 advice and discussion in this project, as well as colleagues in our
582 University. Furthermore, we also thank Dr. Quintin Lau for
583 English editing. This work has been supported in part by the

584 Graduate University for Advanced Studies, SOKENDAI.

585

586 Author contribution statement

587 RLI was responsible for reviewing the literature, conducting the
588 research, extracting and analyzing data, interpreting results and
589 writing the manuscript. YS was responsible for extracting and
590 analyzing data, interpretation and discussion of results and
591 amending the manuscript.

592

593 Conflict of interest

594 The authors declare no conflict of interest.

595

596 References

597 The 1000 Genomes Project Consortium (2015). A global reference
598 for human genetic variation. *Nature* **526**:68-74.

599 Albers PK, McVean G (2020). Dating genomic variants and shared
600 ancestry in population-scale sequencing data. *PLoS Biol*
601 **18**:e3000586.
602 Beleza S, Santos AM, McEvoy B, Alves I, Martinho C, Cameron E
603 et al. (2013). The timing of pigmentation lightening in
604 Europeans. *Mol Biol Evol* **30**:24-35.
605 Benjamini Y, Hochberg Y (1995). Controlling the false discovery
606 rate: a practical and powerful approach to multiple testing.
607 *J Roy Stat Soc Ser* **B57**:289-300.
608 Brown MB (1975). 400: A Method for Combining Non-Independent,
609 One-Sided Tests of Significance. *Biometrics* **31**:987-92.
610 Busby GB, Band G, Si Le Q, Jallow M, Bougama E, Mangano VD
611 et al. (2016). Malaria Genomic Epidemiology Network.
612 Admixture into and within sub-Saharan Africa. *Elife*
613 **5**:e15266.

614 Colosimo PF, Hosemann KE, Balabhadra S, Villarreal GJ, Dickson
615 M, Grimwood J et al. (2005). Widespread parallel
616 evolution in sticklebacks by repeated fixation of
617 Ectodysplasin alleles. *Science* **307**:1928-33.

618 Evans PD, Gilbert SL, Mekel-Bobrov N, Vallender EJ, Anderson
619 JR, Vaez-Azizi LM et al. (2005). Microcephalin, a gene
620 regulating brain size, continues to evolve adaptively in
621 humans. *Science*, **309**:1717–1720.

622 Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M
623 (2013). Robust demographic inference from genomic and
624 SNP data. *PLoS Genet* **9**:e1003905.

625 Fan S, Kelly DE, Beltrame MH, Hansen MEB, Mallick S, Ranciaro
626 A et al. (2019). African evolutionary history inferred from
627 whole genome sequence data of 44 indigenous African
628 populations. *Genome Biol* **20**:82.

629 Fenner JN (2005). Cross-cultural estimation of the human
630 generation interval for use in genetics-based population
631 divergence studies. *American journal of physical*
632 *anthropology* **128**:415–423.

633 Ferrer-Admetlla A., Liang M, Korneliussen T, Nielsen R (2014).
634 On detecting incomplete soft or hard selective sweeps
635 using haplotype structure. *Mol Biol Evol* **31**:1275–1291.

636 Fujito NT, Satta Y, Hayakawa T, Takahata N (2018). A new
637 inference method for detecting an ongoing selective sweep.
638 *Genes Genet Syst* **93**:149-161.

639 Gakuhari T, Nakagome S, Rasmussen S Allentoft ME, Sato T,
640 Korneliussen T et al. (2020). Ancient Jomon genome
641 sequence analysis sheds light on migration patterns of
642 early East Asian populations. *Commun Biol* **3**: 437.

643 Gerbault P, Moret C, Currat M, Sanchez-Mazas A (2009). Impact of

644 selection and demography on the diffusion of lactase
645 persistence. *PLoS One* **4**:e6369.

646 Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L,
647 Tachmazidou I, Hatzikotoulas K et al. (2015). The African
648 Genome Variation Project shapes medical genetics in
649 Africa. *Nature* **517**:327-32.

650 Haber M, Jones AL, Connell BA, Asan, Arciero E, Yang H et al.
651 (2019). A Rare Deep-Rooting D0 African Y-Chromosomal
652 Haplogroup and Its Implications for the Expansion of
653 Modern Humans Out of Africa. *Genetics* **212**:1421-1428.

654 Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM,
655 Kidd JM et al. (2011). Hunter-gatherer genomic diversity
656 suggests a southern African origin for modern humans.
657 *Proc Natl Acad Sci U S A* **108**:5154-62.

658 Hudson RR (2002). Generating samples under a Wright-Fisher

neutral model of genetic variation. *Bioinfomatics* **18**:337-338.

Ishikawa A, Kabeya N, Ikeya K, Kakioka R, Cech JN, Osada N et al. (2019). A key metabolic gene for recurrent freshwater colonization and radiation in fishes. *Science* **364**:886-889.

Itan, Y., Powell, A., Beaumont, M. A., Burger, J., & Thomas, M. G (2009). The origins of lactase persistence in Europe. *PLoS computational biology* **5**:e1000491.

Iwasaki RL, Ishiya K, Kanzawa-Kiriyama H, Kawai Y, Gojobori J, Satta Y (2020). Evolutionary History of the Risk of SNPs for Diffuse-Type Gastric Cancer in the Japanese Population. *Genes (Basel).* **11**:775.

Jones DA, Wakeley J (2008). The influence of gene conversion on linkage disequilibrium around a selective sweep. *Genetics* **180**:1251–1259.

674 Kamberov YG, Wang S, Tan J, Gerbault P, Wark A, Tan L (2013).
675 Modeling recent human evolution in mice by expression of
676 a selected EDAR variant. *Cell* **152**:691–702.
677 Kanzawa-Kiriyama H, Jinam TA, Kawai Y, Sato T, Hosomichi, K,
678 Tajima, A et al. (2019). Late Jomon male and female
679 genome sequences from the Funadomari site in Hokkaido,
680 Japan. *Anthropol. Sci.* **127**:83-109.
681 Key FM, Abdul-Aziz MA, Mundry R, Peter BM, Sekar A, D'Amato
682 M et al. (2018). Human local adaptation of the TRPM8
683 cold receptor along a latitudinal cline. *PLoS Genet*
684 **14**:e1007298.
685 Lachance J, Vernot B, Elbers CC, Ferwerda B, Froment A, Bodo
686 JM et al. (2012). Evolutionary history and adaptation
687 from high-coverage whole-genome sequences of diverse
688 African hunter-gatherers. *Cell* **150**:457-69.

McManus KF, Taravella AM, Henn BM, Bustamante CD, Sikora M, Cornejo OE (2017). Population genetic analysis of the DARC locus (Duffy) reveals adaptation from standing variation associated with malaria resistance in humans. *PLoS Genet* **13**:e1006560.

Miller CT, Beleza S, Pollen AA, Schluter D, Kittles RA, Shriver MD et al. (2007). cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* **131**:1179-89.

Nakagome S, Hudson RR, Di Rienzo, A (2019). Inferring the model and onset of natural selection under varying population size from the site frequency spectrum and haplotype structure. *Proceedings Biological sciences* **286**:20182541.

Oleksyk TK, Smith MW, O'Brien SJ (2010). Genome-wide scans for footprints of natural selection. *Philos Trans R Soc Lond B*

Biol Sci 365:185-205.

Peter BM, Huerta-Sanchez E, Nielsen R (2012). Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS Genet* **8**:e1003011.

Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D et al. (2009). Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* **19**:826-37.

Rees JS, Castellano S, Andrés, AM (2020). The Genomics of Human Local Adaptation. *Trends in genetics* **36**:415–428.

Satta Y, Zheng W, Nishiyama KV, Iwasaki RL, Hayakawa T, Fujito NT et al. (2020). Two-dimensional site frequency spectrum for detecting, classifying and dating incomplete selective sweeps. *Genes & genetic systems*, **94**:283–300.

Saunders MA, Slatkin M, Garner C, Hammer MF, Nachman MW

719 (2005). The extent of linkage disequilibrium caused by
720 selection on G6PD in humans. *Genetics* **171**:1219-29.
721 Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D
722 (2005). Calibrating a coalescent simulation of human
723 genome sequence variation. *Genome Res* **15**:1576-83.
724 Schlebusch CM, Sjödin P, Skoglund P, Jakobsson M (2012).
725 Stronger signal of recent selection for lactase persistence
726 in Maasai than in Europeans. *Eur J Hum Genet* 21:550-3.
727 Schrider DR, Kern AD (2017). Soft Sweeps Are the Dominant Mode
728 of Adaptation in the Human Genome. *Mol Biol Evol*
729 **34**:1863-1877.
730 Smith J, Coop G, Stephens M, Novembre J (2018). Estimating
731 Time to the Common Ancestor for a Beneficial Allele. *Mol*
732 *Biol Evol* **35**:1003-1017.
733 Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso

C et al.. (2012). The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol Biol Evol* **29**:915-27.

The Study Group of Millennium Genome Project for Cancer (2008). Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat Genet* **40**:730-40.

Tanikawa C, Urabe Y, Matsuo K, Kubo M, Takahashi A, Ito H et al. (2012). A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population. *Nat Genet* **44**:430-4.

Terhorst J, Kamm JA, Song YS (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat Genet* **49**:303-309.

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS et al. (2007). Convergent adaptation of

749 human lactase persistence in Africa and Europe. *Nat*
750 *Genet* **39**:31-40.

751 Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G,
752 Destro-Bisol G et al. (2001). Haplotype diversity and
753 linkage disequilibrium at human G6PD: recent origin of
754 alleles that confer malarial resistance. *Science*
755 **293**:455-62.

756 Wang X, Grus WE, Zhang J (2006). Gene losses during human
757 origins. *PLoS Biol* **4**:e52.

758 Wang Y, Lu D, Chung YJ, Xu S (2018). Genetic structure,
759 divergence and admixture of Han Chinese, Japanese and
760 Korean populations. *Hereditas.* **155**:19.

761 Xue Y, Daly A, Yngvadottir B, Liu M, Coop G, Kim Y (2006).
762 Spread of an inactive form of caspase-12 in humans is due
763 to recent positive selection. *Am J Hum Genet* **78**:659-70.

764 Yang Z, Shi H, Ma P, Zhao S, Kong Q, Bian T et al. (2018).
765 Darwinian Positive Selection on the Pleiotropic Effects of
766 KITLG Explain Skin Pigmentation and Winter
767 Temperature Adaptation in Eurasians. *Mol Biol Evol*
768 **35**:2272-2283.

769 Yasukochi Y, Satta Y (2013). Current perspectives on the intensity
770 of natural selection of MHC loci. *Immunogenetics*
771 **65**:479-83.

772

773 ## Figure legends

774 **Figure 1.** The relationships among haplotypes.
775 The C haplotype, with a C allele at the first position of three SNPs
776 (rs2294008, rs2976391 and rs2978983), diverged from the T
777 haplotype around 1 mya (Albers and McVean 2020). Within the C
778 haplotype, the C-A-G haplotype with two derived alleles at

rs2976391 and rs2978983 diverged from the C-C-A haplotype around 170 kya (Albers and McVean 2020).

Figure 2. Values of t_D ± SD in populations under positive selection are limited to within 3-55 kya.

Note that the C-C-A haplotype in ESN and the C-A-G haplotype in CEU/PEL/CHS are excluded because they have marginal/unusual signals (see main text). (A) Values of t_D ± SD for the C-C-A haplotype. (B) Values of t_D ± SD for the C-A-G haplotype. The X markers represent the t_D values and the double-sided arrows represent the ranges of t_D ± SD. Red, orange and green represent East Asian, South Asian and African subpopulations, respectively.

Figure 3. Transition of target alleles throughout human history.

Summary of positive selection signals detected on C-A-G and C-C-A haplotypes in 1KGP 26 subpopulations. Filled pink/blue, dashed and open pink/blue circles indicate ongoing positive

794 selection on C-C-A/C-A-G, completed positive selection on C-C-A
795 and no significant signals of selection on C-C-A/C-A-G, respectively.
796 The phylogeny of the 26 subpopulations was constructed using the
797 neighbor-joining method of pairwise F_{ST} values (The 1000
798 Genomes Project Consortium 2015). Transition process of selection
799 target haplotype is inferred from the extant target haplotype, their
800 t_D values as well as subpopulation divergence times. In cases with
801 no available divergence time data of subpopulations, we inferred
802 timing of transition in a parsimonious manner.

803 **Figure 4**. Inferred timing of onsets of positive selection and
804 selection coefficients in JPT/CHB/YRI.
805 (A) Violin plot of the timing of positive selection onsets estimated
806 from the ABC framework. Double sided arrows indicate $t_D \pm$ SD.
807 See Table 2 for means (and 95% CI) of timing onsets of positive
808 selection and Table 1 for ranges of $t_D \pm$ SD. (B) Violin plot of

809 selection coefficients. See Table 2 for means (and 95% CI) of

810 selection coefficients.

Table 1. 2D-SFS and t_0 values for 26 subpopulations.

Population	JPT		CHB		CHS		CDX		KHV	
Core region	143755915–143770914		143755876–143771875		143755915–143771914		143755915–143771914		143755915–143771914	
Targeting subhaplotype	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G
n	37	40	79	74	71	80	71	64	70	73
Frequency of tested allele	0.178	0.192	0.383	0.359	0.338	0.381	0.382	0.344	0.354	0.369
S	91		88		89		82		89	
F_c	0.0112 (0.035)*	0.0483 (0.296)	0.0196 (0.035)*	0.0156 (0.023)*	0.0094 (0.009)**	0.0230 (0.052)	0.0146 (0.015)*	0.0899 (0.547)	0.0080 (0.007)**	0.0302 (0.086)
G_{c0}	1.50 (0.180)	2.17 (0.287)	1.90 (0.055)	1.78 (0.057)	2.00 (0.115)	2.44 (0.137)	2.40 (0.171)	3.43 (0.313)	1.50 (0.039)*	2.73 (0.186)
L_{c0}	0.0006 (0.046)*	0.0024 (0.273)	0.0031 (0.020)*	0.0026 (0.023)*	0.0016 (0.007)**	0.0035 (0.042)*	0.0021 (0.017)*	0.0042 (0.399)	0.0015 (0.007)**	0.0051 (0.117)
G^*_{c0}	2.00	8.00	4.00	8.00	6.00	3.17	3.33	6.67	2.00	4.80
$\gamma^*(10)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.20
I^*_{max}	0	0	1	2	1	5	2	1	2	1
i_{max}	2	8	5	8	6	5	4	10	2	10
combined P						Marginally significant (0.032, x2=7.83, df=2.44)				
selection operated?§	ongoing	No	ongoing	ongoing	ongoing	No	ongoing	No	ongoing	No
t_0 (years ago)	10811	43333	30063	27027	17606	34375	21127	46875	16071	51370
SD of t_0 (years ago)	8058	27689	12358	14333	11135	13073	10563	25615	6916	22062

Population	BEB		PJL		GIH		ITU		STU	
Core region	143754477–143778476		143756811–143778810		143755915–143778914		143754412–143778411		143754811–143778810	
Targeting subhaplotype	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G
n	41	68	34	70	29	76	55	67	66	72
Frequency of tested allele	0.238	0.395	0.177	0.365	0.141	0.369	0.270	0.328	0.324	0.353
S	123		129		107		121		123	
F_c	0.0194 (0.023)**	0.2379 (0.544)	0.0500 (0.247)	0.0592 (0.112)	0.0357 (0.309)	0.0673 (0.255)	0.0511 (0.074)	0.2443 (0.646)	0.0472 (0.053)	0.0286 (0.126)
G_{c0}	1.00 (0.017)*	2.45 (0.184)	1.43 (0.168)	1.94 (0.072)	1.50 (0.212)	2.85 (0.288)	1.33 (0.025)*	2.26 (0.174)	1.11 (0.008)**	2.75 (0.175)
L_{c0}	0.0006 (<0.002)**	0.0079 (0.148)	0.0017 (0.147)	0.0058 (0.067)	0.0009 (0.161)	0.0055 (0.095)	0.0011 (0.013)*	0.0061 (0.148)	0.0014 (0.008)**	0.0063 (0.126)
G^*_{c0}	NaN	5.83	2.00	3.43	3.00	4.00	3.00	4.43	2.00	5.67
$\gamma^*(10)$	NaN	0.33	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.17
I^*_{max}	0	8	1	5	0	10	1	6	1	9
i_{max}	1	12	2	7	3	9	3	12	2	16
combined P										
selection operated?§	ongoing	No	No	No	No	No	completed	No	completed	No
t_0 (years ago)	8130	60049	26738	45455	17991	42334	12121	53483	12626	50926
SD of t_0 (years ago)	4065	21888	10695	13927	10387	15049	5669	19067	4374	22203

Population	FIN		TSI		GBR		CEU		IBS	
Core region	143753801–143780800		143755915–143778914		143752095–143779094		143756573–143774572		143755915–143778914	
Targeting subhaplotype	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G
n	13	80	16	111	14	95	20	94	19	93
Frequency of tested allele	0.066	0.404	0.075	0.519	0.077	0.522	0.101	0.475	0.089	0.435
S	118		123		126		98		136	
F_c	0.0000 (0.053)	0.1284 (0.374)	0.0079 (0.128)	0.1660 (0.332)	0.0000 (0.033)*	0.1269 (0.356)	0.0133 (0.162)	0.1398 (0.327)	0.0298 (0.254)	0.1141 (0.291)
G_{c0}	0.00 (0.150)	5.92 (0.546)	1.00 (0.265)	4.00 (0.153)	0.00 (0.094)	6.28 (0.282)	1.00 (0.230)	3.29 (0.123)	1.25 (0.341)	3.76 (0.182)
L_{c0}	0.0000 (0.150)	0.0096 (0.143)	0.0001 (0.146)	0.0114 (0.053)	0.0000 (0.094)	0.0219 (0.190)	0.0002 (0.112)	0.0075 (0.027)*	0.0007 (0.364)	0.0086 (0.057)
G^*_{c0}	NaN	8.38	NaN	6.25	NaN	12.00	NaN	5.00	2.00	6.22
$\gamma^*(10)$	NaN	0.25	NaN	0.17	NaN	0.25	NaN	0.13	0.00	0.33
I^*_{max}	0	6	0	11	0	13	1	7	0	6
i_{max}	0	21	1	25	0	94	1	18	2	92
combined P					Marginally significant (0.017, x2=9.91, df=2.88)			Marginally significant (0.046, x2=7.38, df=2.65)		
selection operated?§	No	No	No	No	No	No	No	No	No	No
t_0 (years ago)	–	65741	5435	65805	–	122417	5556	54374	22883	59841
SD of t_0 (years ago)	–	26498	5435	24823	–	66405	5556	24167	12109	20400

Population	MXL		PUR		CLM		PEL	
Core region	143752915–143826914		143755891–143780890		143752389–143775388		143752891–143765890	
Targeting subhaplotype	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G
n	5	49	14	94	11	63	13	93
Frequency of tested allele	0.039	0.383	0.067	0.452	0.059	0.335	0.076	0.547
S	290		140		125		82	
F_c	0.0373 (0.635)	0.1221 (0.757)	0.0409 (0.521)	0.0957 (0.199)	0.0397 (0.545)	0.2778 (0.699)	0.0380 (0.704)	0.0086 (0.002)**
G_{c0}	0.83 (0.284)	26.21 (>0.999)	1.40 (0.640)	3.88 (0.189)	1.33 (0.726)	3.50 (0.312)	1.50 (0.548)	10.78 (0.478)
L_{c0}	0.0006 (0.704)	0.1747 (>0.999)	0.0010 (>0.999)	0.0086 (0.038)*	0.0006 (0.802)	0.0051 (0.062)	0.0007 (0.636)	0.0212 (0.073)
G^*_{c0}	0.00	36.66	3.00	5.60	2.00	4.57	2.00	23.00
$\gamma^*(10)$	0.00	0.78	0.00	0.20	0.00	0.00	0.00	0.50
I^*_{max}	0	19	1	13	1	15	0	1
i_{max}	4	48	3	18	2	9	2	92
combined P				Not significant (0.052, x2=7.16, df=2.66)				Marginally significant (0.016, x2=9.75, df=2.69)
selection operated?§	No	No	No	No	No	No	No	No
t_0 (years ago)	27027	838389	40000	52766	31621	48309	35503	160463
SD of t_0 (years ago)	5405	143970	20603	21191	19364	18875	26462	143215

Population	ASW		ACB		GWD		MSL		ESN		LWK		YRI	
Core region	143756015–143771014		143756072–143771071		143756352–143774351		143752994–143781058		143757203–143774202		143756072–143771071		143756108–143771107	
Targeting subhaplotype	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G	C–C–A	C–A–G
n	22	39	36	73	48	100	26	64	57	82	37	76	50	116
Frequency of tested allele	0.180	0.320	0.188	0.380	0.212	0.442	0.153	0.376	0.288	0.414	0.187	0.384	0.231	0.537
S	133		126		151		244		129		142		138	
F_c	0.0667 (0.246)	0.0301 (0.008)**	0.0719 (0.442)	0.0143 (<0.002)**	0.1114 (0.591)	0.0187 (<0.002)**	0.0459 (0.247)	0.0367 (0.002)**	0.1327 (0.293)	0.0173 (<0.002)**	0.0774 (0.341)	0.0364 (0.006)**	0.0723 (0.243)	0.0181 (<0.002)**
G_{c0}	2.40 (0.293)	1.14 (0.002)**	4.00 (0.306)	1.50 (<0.002)**	6.57 (0.550)	1.42 (<0.002)**	3.18 (0.280)	1.50 (<0.003)**	3.75 (0.152)	1.50 (<0.002)**	3.83 (0.412)	2.44 (0.006)**	4.43 (0.230)	1.31 (<0.002)**
L_{c0}	0.0034 (0.130)	0.0023 (<0.002)**	0.0037 (0.094)	0.0022 (<0.002)**	0.0064 (0.155)	0.0024 (<0.002)**	0.0051 (0.278)	0.0039 (<0.003)**	0.0056 (0.025)*	0.0033 (<0.002)**	0.0040 (0.093)	0.0038 (<0.002)**	0.0048 (0.047)*	0.0026 (<0.002)**
G^*_{c0}	3.33	2.00	4.00	2.33	10.75	2.67	4.00	3.25	6.50	2.20	6.67	3.60	5.00	2.00
$\gamma^*(10)$	0.00	0.00	0.00	0.00	0.25	0.00	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00
I^*_{max}	0	6	1	8	0	1	1	4	4	15	0	11	1	18
i_{max}	4	2	7	3	19	4	5	5	11	3	8	7	8	2
combined P									Marginally significant (0.040, x2=8.37, df=3.02)				Not significant (0.062, x2=7.11, df=2.86)	
selection operated?§	No	ongoing	No	ongoing	No	ongoing	No	ongoing	No	ongoing	No	ongoing	No	ongoing
t_0 (years ago)	72727	27350	74074	21918	106481	18889	95931	30064	61920	25825	82883	38596	82667	19540
SD of t_0 (years ago)	36364	10811	36289	8567	54681	6383	32315	8838	30193	8116	42486	16644	35075	5747

For F_c, G_{c0} and L_{c0}, q–value in the parentheses represents under null model. * represents q–value < 0.05 and ** represents q–value < 0.01.

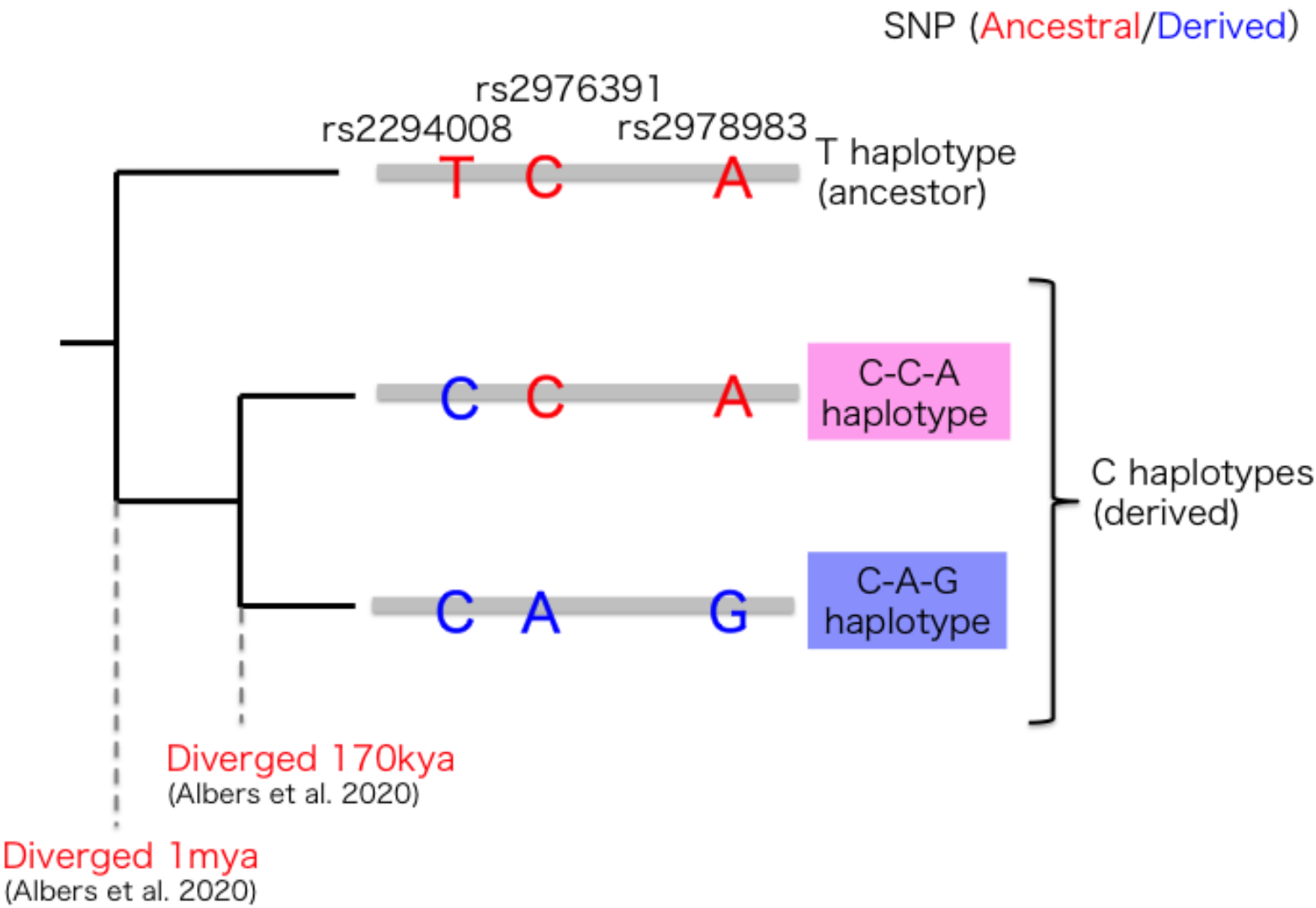
n represents the number of samples.

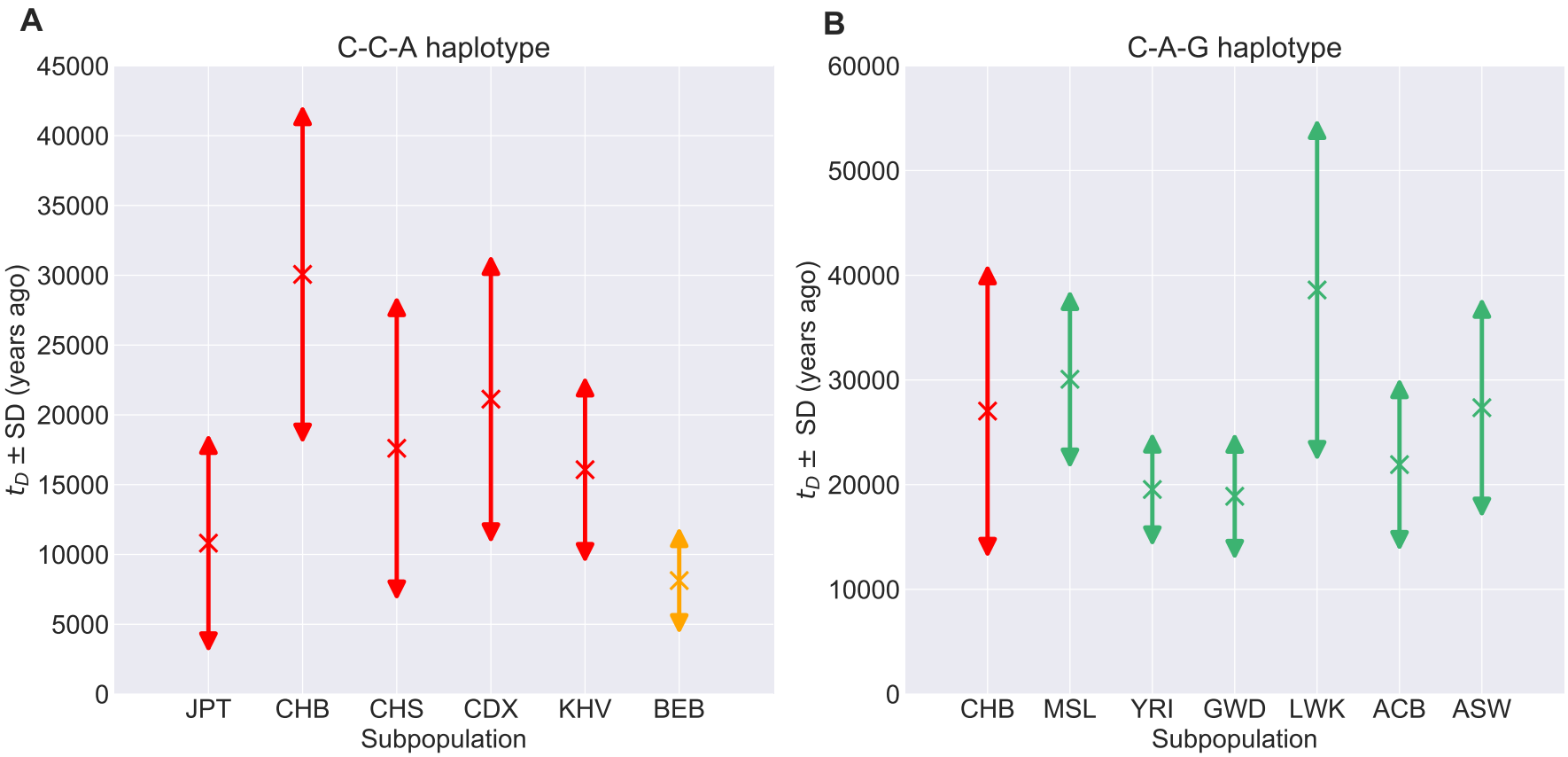
S represents the number of segregating sites in tested population.

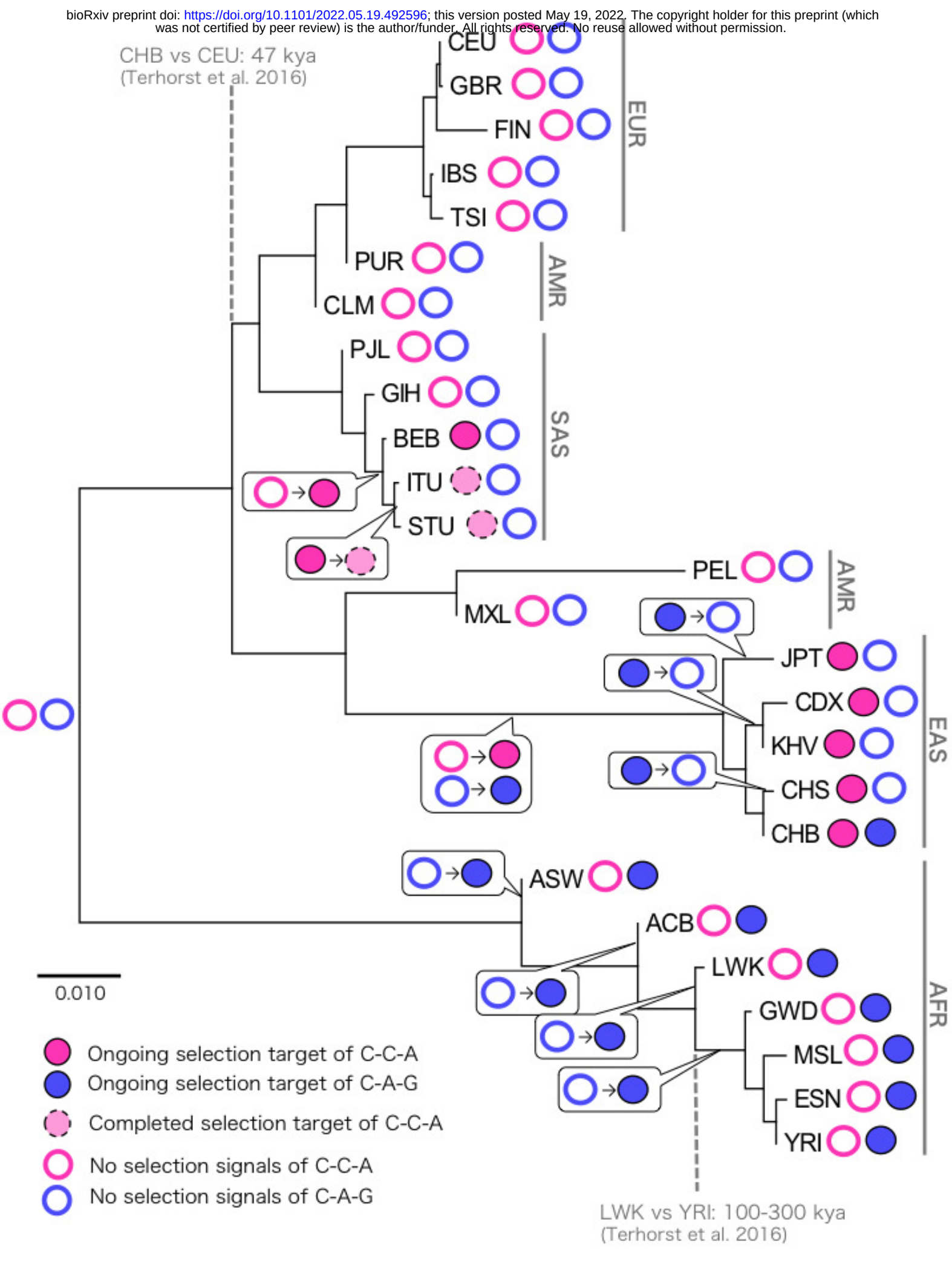
§Selection status marked based on three q–values of Fc, Gc0 and Lc0 and a combined P–value. See Materials and Methods for criteria.

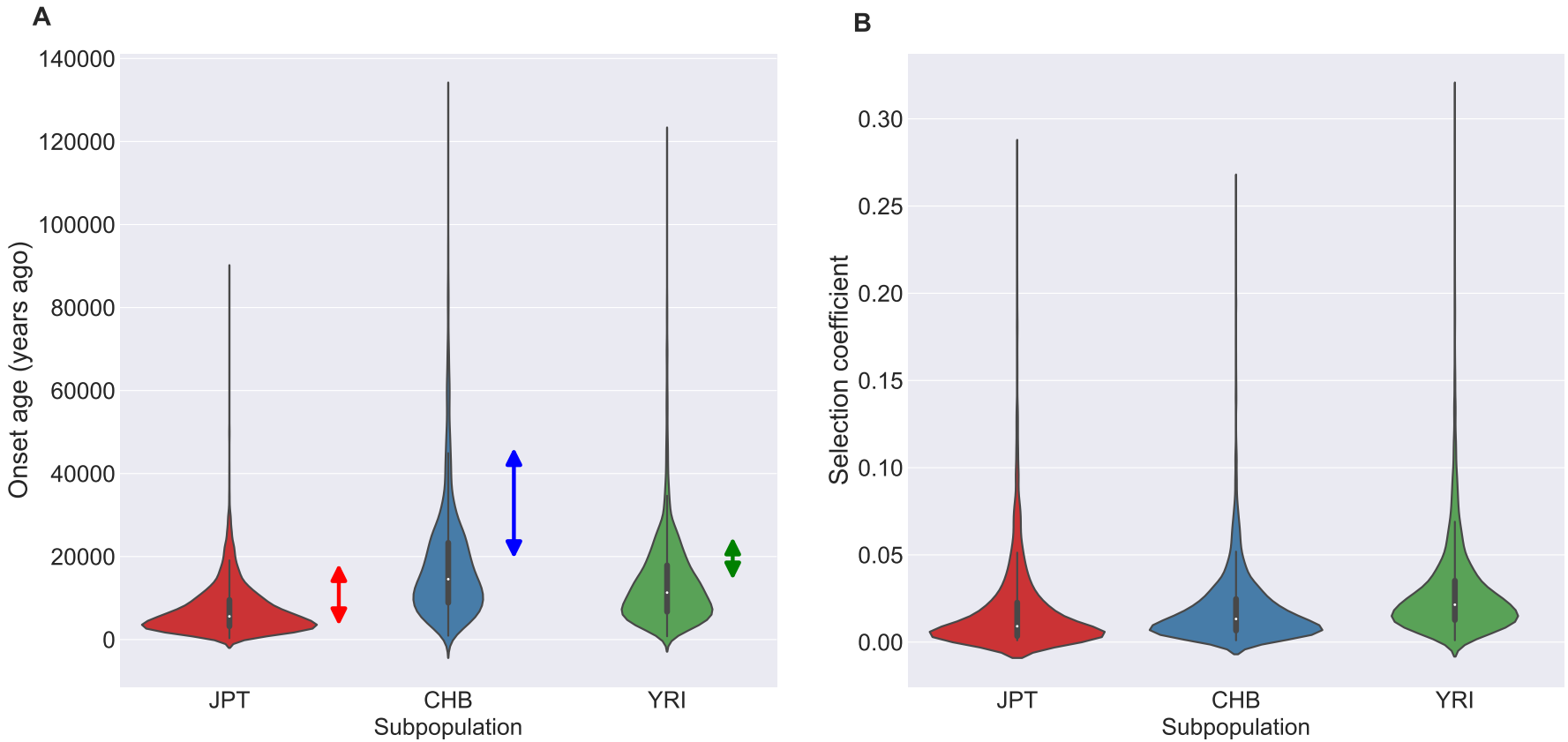
Table 2. The onset times and selection coefficients inferred with ABC method.

	Target haplotype	*s* (mean)	*s* (95% CI)	onset time (mean (years))	onset time (95% CI (years))
JPT	C-C-A	0.0189	0.0011-0.0958	7334	1137-23270
CHB	C haplotype	0.0194	0.0018-0.0718	18174	3213-59276
YRI	C-A-G	0.0290	0.0029-0.1038	13845	2390-42027




A
C-C-A haplotype
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
$t_D \pm$ SD (years ago)
JPT
CHB
CHS
CDX
KHV
BEB
Subpopulation
B
C-A-G haplotype
60000
50000
40000
30000
20000
10000
0
$t_D \pm$ SD (years ago)
CHB
MSL
YRI
GWD
LWK
ACB
ASW
Subpopulation




A
140000
120000
100000
80000
60000
40000
20000
0
Onset age (years ago)
JPT
CHB
YRI
Subpopulation
B
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Selection coefficient
JPT
CHB
YRI
Subpopulation