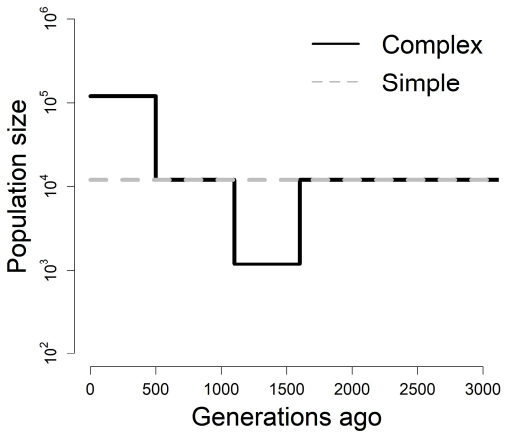
SUPPLEMENTARY TEXT

Prior settings and demographic conditions

We used a demographic model inferred for Europeans [1] that is characterized by three transitions (see a figure below): 1) a 10-fold population reduction from 12,000 to 1,200 individuals (*i.e.*, bottleneck) between 1,600 and 1,100 generations ago, 2) a recovery from the bottleneck 1,100 generations ago and 3) an instantaneous 10-fold population expansion from 12,000 to 120,000 individuals 500 generations ago. These transitions define four-time intervals with constant population size, namely, $0 < T \leq 500$, $500 < T \leq 1100$, $1100 < T \leq 1600$, and $1600 < T$. We number these intervals from 1 to 4, from the most recent to the oldest. Under all three models (*i.e.*, selection on standing variation, selection on new mutations and neutrality), we assume the prior distribution of T is a log-normal distribution with a mean of 1,600 and a variance equal to the square of the mean. We choose a log-normal distribution to avoid negative time values, but the choice of variance is arbitrary. Under the SNM and SSV models we assume $\log_{10} s$ has a uniform distribution between -3 and -0.5. For the SSV model we also require a prior distribution for f_T, the frequency of the previously neutral variant which is subsequently subject to selection. This prior distribution is the SFS of neutral variants and will change through time under our model due to the demographic transitions. We estimate this SFS at just four time points, 250, 800, 1350 and 2000 generations ago. These times correspond to the midpoints of the first three demographic intervals and an arbitrary recent time point in the oldest interval. We denote these estimated spectra by SFS_i, and these time points by T_i where $i = \{1, 2, 3, 4\}$ indexes the four time points corresponding to the four demographic intervals. We assume that the frequency spectrum at any point within a demographic interval is well approximated by our estimated SFS at the midpoint of that interval or at generation 2000 for times in oldest interval. To estimate SFS_i, we draw a random time for the origin of a new neutral mutation from a uniform distribution that has the range from T_i to $4N_0$ generations ago, where N_0 is the current population size of 120,000 in our demographic model. We then generated a neutral trajectory forward in time from our random time to T_i, and if the neutral variant has not been fixed or lost, we record its frequency. We repeated this process until we obtained 1 million replicates of allele frequency at T_i.

As an alternative demographic scenario, we also included a constant size model ($N = 12{,}000$) into our simulation study. The prior settings for T and s were the same as the complex demography. For f_T, we used $\mathrm{SFS}_{i=4}$ regardless of T_i as the expected SFS is not expected to depend on T under the constant size model.



Simulation scheme

Step 1: Simulating trajectory under the Write-Fisher model.

We simultaneously sampled values of the parameters from the priors and simulated trajectories under different models as described below (Figure S1).

1) Natural selection on a new mutation (SNM): Using the selection coefficient s starting T generations ago at $f_T = 1/2N_{\mathrm{e}}$, an allele frequency trajectory was generated forward in time under a simple additive selective model (*i.e.*, 1, $1 + s/2$, and $1 + s$ for *aa*, *Aa*, *AA*, where *A* is a derived and beneficial allele and *a* is an ancestral allele) with genetic drift scaled with the effective population size. The simulated trajectory was accepted based on the final frequency of the simulated allele with a probability proportional to the binomial probability based on allele frequency at present (f_0) and numbers of derived (n_{der}) and ancestral alleles (n_{anc}). If the simulated trajectory was rejected, a new set of s and T was sampled from the priors, so that the accepted samples of both s and T were conditional on f_0.

2) Natural selection on a standing variant (SSV): The simulation scheme for the selection phase was the same as the SNM model, with the exception that $f_T \sim \mathrm{SFS}_i$. For the neutral phase, we started simulation at time T with frequency f_T and generated trajectory backwards in time with genetic drift only. We accepted the trajectory if the allele went extinct in the backward direction.

3) Neutral: A frequency path was conditional only on f_0. We simulated the trajectory backwards in time without any selection parameters, so the change of allele frequency was dependent on genetic drift scaled with the effective population size. The trajectory was accepted if the allele went extinct. Otherwise, we re-started the simulation at f_0.

Step 2: Simulating neutral variation under the coalescent model.

Next, for each accepted trajectory, neutral variation surrounding the selected site was generated by coalescent simulation conditional on the trajectory. To model the uncertainty in mutation rate (μ), μ was sampled from a truncated normal distribution with mean equal to the divergence-based μ estimate (from the comparison of human with the other primates) and a variance equal to the square of the mean; values less than a half of μ and greater than 1.5 times μ were excluded. The above procedure was repeated until 1 million sets of data and parameter values (*i.e.*, T and s under the SNM model; T, s, and f_T under the SSV model) were generated.

Our simulation iterates the procedures as follows:

for i = 1 to n (*i.e.*, the total number of simulations) **do**

if $m = 0$ **then** # **SSV model**

1a. Sample t, s, and f_T from the priors.

2a. Simulate trajectory at the selection phase, $traj_{i,\text{selection}}$.

3a. Accept $traj_{i,\text{selection}}$ if the final allele frequency at present matches f_0, with probability proportional to the binomial probability of f_0. Otherwise go back to **1a**.

4a. Simulate trajectory at the neutral phase, $traj_{i,\text{neutral}}$

5a. Accept $traj_{i,\text{neutral}}$ if the allele goes extinct. Otherwise go back to **4a**.

6a. Merge $traj_{i,\text{selection}}$ and $traj_{i,\text{neutral}}$ to generate $traj_{i,0}$

else if $m = 1$ **then** # **SNM model**

1b. Sample t and s from the priors.

2b. Simulate trajectory, $traj_{i,1}$.

3b. Accept $traj_{i,1}$ with the probability proportional to the binomial probability of f_0. Otherwise go back to **1b**.

else if $m = 2$ **then** # **Neutral model**

1c. Simulate trajectory, $traj_{i,2}$.

2c. Accept $traj_{i,2}$ if the allele goes extinct. Otherwise go back to **1c**.

Simulate neutral variation, $G_{i,m}$, with $traj_{i,m}$ and μ_i that is sampled from the truncated normal distribution.

Calculate summary statistics, $S(G_{i,m})$.

Additional condition at Step 1: Incorporating ancient DNA data into simulation.

To incorporate information from ancient DNA data we modified the procedure by retaining only trajectories with a specified allele frequency, $f_{T_{\text{past}}}$, at an additional time point, T_{past}. The values of these additional parameters are obtained from published ancient DNA data from specimens with dating information and where the numbers of sequence reads carrying reference and alternative alleles were provided. We estimated $f_{T_{\text{past}}}$ at a given time using a likelihood function described in [2]. To allow for uncertainty in the estimate due to the limited number of samples, we accepted trajectories in which the allele frequency at T_{past} fell within a range ($f_{T_{\text{past}}} \pm 5.0\%$). All other conditions were the same as those without ancient DNA as described above. Because all frequency paths have to pass through $f_{T_{\text{past}}}$ at T_{past} and reach a frequency close to f_0 at present, accepted samples of the parameters from this procedure were conditional on $f_{T_{\text{past}}}$, as well as f_0. However, if the difference between $f_{T_{\text{past}}}$ and f_0 is large, it may be difficult to generate trajectories under the neutral model. In turn, this implies that a given change of allele frequency is unlikely to be explained by neutrality. Even though the allele frequency path would include all information regarding selection, the path captured only by two time points (*i.e.*, T_{past} and T_0) may not be sufficient to infer the strength of

selection. Therefore, we simulated trajectories with the conditions of $f_{T_{\text{past}}}$ and f_0 under the SNM and SSV models with the goal of distinguishing between the types of selection rather than to test the deviation from neutral model.

The simulation procedures are almost the same as those without ancient DNA, except for "3a" or "5a" under the SSV model and "3b" under the SNM model. Depending on T sampled from the prior, the condition on $f_{T_{\text{past}}}$ was given for the trajectory at the selection phase or the neutral phase in the SSV model. If $T < T_{\text{past}}$ or $T > T_{\text{past}}$, **3a** or **5a** was replaced with **3a'** or **5a'** as follows:

3a'. Accept $traj_{i,\text{selection}}$ if the final allele frequency at present matches f_0, with probability proportional to the binomial probability of f_0, and the frequency at T_{past} is within $f_{T_{\text{past}}} \pm 5.0\%$. Otherwise go back to **1a**.

5a'. Accept $traj_{i,\text{neutral}}$ if the allele goes extinct and the frequency at T_{past} is within $f_{T_{\text{past}}} \pm 5.0\%$. Otherwise go back to **4a**.

On the other hand, no condition was given for the trajectory in the SNM model if T is younger than T_{past}. Therefore, **3b** was replaced with **3b'** only if $T < T_{\text{past}}$:

3b'. Accept $traj_{i,1}$ with the probability proportional to the binomial probability of f_0 and the frequency at T_{past} is within $f_{T_{\text{past}}} \pm 5.0\%$. Otherwise go back to **1b**.

Summarizing patterns of neutral variation in to a set of statistics

The rapid increase in allele frequency due to selection is expected to result in a skewed SFS and increased linkage disequilibrium (LD) [3-5]. Commonly used summaries of the SFS include the number of segregating sites, nucleotide diversity, Tajima's D, Fu and Li's D and F, and Fay and Wu's H [6-9], which are not independent. In contrast, using the full SFS, which covers the entire range of allele frequency from singletons to mutations shared in all chromosomes, provides complete and non-redundant information on the local reduction in nucleotide diversity and on the distortion of the SFS in a population. The spatial distribution of mutations within a genomic region can also provide information on the strength of selection as stronger selection will affect patterns of variation over longer genetic distances. To develop our ABC approach, we focused on a 300kb genomic region centered on a focal site and calculated the full SFS separately for sequences carrying derived (*i.e.*, beneficial) and ancestral alleles at the focal site. Examples of the full SFS obtained by simulations under selection and neutral models are shown in Figure S2.

Another aspect of sequence variation that is impacted by natural selection is the extent of LD. The extended haplotype homozygosity (EHH) is defined as a probability that two randomly chosen haplotypes are homozygous at all SNPs within a given distance from the focal SNP site, taking a value in between 0 (*i.e.*, all haplotypes are different) and 1 (*i.e.*, all haplotypes are identical) [10]. To

fully capture the haplotype structure in a genomic region containing a beneficial allele, we recorded the physical positions of the variable sites on each side of the focal site where the EHH value decreased from 0.9 to 0.1, in steps of 0.1, for the derived and ancestral alleles separately. This is similar to integrated haplotype scores that have been widely used for selection scans [11]. However, instead of summarizing the decay of EHH into a single statistic by integration, we generated the vector of physical positions as additional summary statistics to characterize in detail the breakdown of LD with the distance from the focal site (Figure S2).

Implementation of ABC for the inference on onsets of natural selection

Our ABC framework is designed to first choose a model the best fit to the observed data and then estimate the parameters of interest under the model. To take advantage of using high-dimensional summaries in the inference [12], we employed a specific ABC method that incorporates kernel methods into its framework to achieve a better approximation [13-15]. Kernel methods, which has been widely used in machine learning, provide systematic ways of analysing high-dimensional data by transforming the data into high or potentially infinite space, which is called reproducing kernel Hilbert space [16]. In this space, the similarity of summary statistics is evaluated as the inner product between pairs of data points, which is simply computed by a kernel function such as the Gaussian radial base function kernel.

Kernel density estimation (KDE) is a smoothing method commonly used to estimate probability density at a given data point from random samples. This method has been extended into ABC model selection with high-dimensional summaries to calculate approximate marginal likelihoods (aMLs), $P(S(\mathcal{D})|M)$, where $S(\mathcal{D})$ is a set of summary data and M is a given model [15]. The basic form of ABC is a rejection-based method, which gives an estimator of aML as an acceptance rate under a model. If a set of summaries is high-dimensional or continuous, however, it becomes hard to find an exact match between observed and simulated summaries. The model selection with KDE provides a simple way to overcome this limitation by replacing an indicator function to accept or reject simulation data with a kernel function to estimate the density from all simulated data. We used 1 million data sets to calculate aMLs under the SNM, SSV, and neutral models. A bandwidth of a Gaussian radial base function kernel was chosen based on the 10-fold cross-validation with 30,000 simulated samples [15]. Fitting to the models was then evaluated with approximate Bayes factors (aBFs), a ratio of aML in one model to that in the other, to test the deviation from neutrality and to distinguish the two selection models.

Kernel ABC provides an algorithm to estimate posterior means with a large dimensional summary of the data, taking advantage of ridge regression integrated into kernel methods to account for varying degrees of correlation among the summary statistics and for non-linear relationships between the summaries and parameters of interest [13, 14]. Simulated samples are used as training data to

calculate weights, w, which provides an estimator of the posterior mean as $E[\boldsymbol{\theta}|S(\mathcal{D})] = \sum_{i=1}^{n} w_i \boldsymbol{\theta}_i$, where, n is the total number of simulations, and $\boldsymbol{\theta}$ is a set of the parameters, $\boldsymbol{\theta} = \{T, s\}$ for SNM or $\boldsymbol{\theta} = \{T, s, f_T\}$ for SSV models. The weight, w_i, is given by $w_i = \sum_{j=1}^{n} (G_S + n\varepsilon I)_{ij}^{-1} k\left(S(\mathcal{D}_j), S(\mathcal{D})\right)$, where G_S is a Gram matrix consisting of $\left(k\left(S(\mathcal{D}_i), S(\mathcal{D}_j)\right)\right)_{i,j=1}^{n}$, ε is a regularization parameter, and I is the $n \times n$ identity matrix. Importantly, $k\left(S(\mathcal{D}_j), S(\mathcal{D})\right)$ gives the similarity between the observed and each simulated data. The posterior mean is thus estimated as the sum of priors re-weighted with the similarity. We chose a bandwidth of the kernel function and ε by the 10-fold cross-validation [13, 14]. The weights are further used to estimate credible intervals by calculating the density of parameter at any given point, θ_0; $p(\theta_0) = \sum_{i=1}^{n} w_i J_h(\theta_i - \theta_0)$, where $J_h(\theta_i - \theta_0)$ is a smoothing kernel and given by a Gaussian kernel, $\frac{1}{\pi^{1/2}h}\exp(-\|\theta_i - \theta_0\|^2/h^2)$, and h is a bandwidth [17]. Because w_i is not positive-definite and may take a negative value, the density is sometimes estimated to be negative. Therefore, we only took account of positive density to calculate 95% credible intervals.

Analysis of genomic data from modern and ancient Europeans

We retrieved the Complete Genomics data for a population with European ancestry (noted as "CEU") including 32 trios from ftp.1000genomes.ebi.ac.uk:/vol1/ftp/release/20130502/supporting/cgi_variant_calls/filtered_calls/. First, we extracted a region spanning 2.5 Mb upstream and downstream to each of the target SNP sites. We then removed SNPs with missing genotypes or Mendelian errors. Third, we phased the data for the parents of each trio using the SHAPEIT2 using the pedigree information [18, 19], thus providing a total of 64 unrelated individuals. This strategy increases phasing accuracy by combining both transmission and LD information, and it mirrors the approach used to generate the high-quality HapMap CEU haplotypes [20]. We retrieved the sequence data for a region 500 kb upstream and downstream of the focal SNP site. In calculating the summary statistics for each SNP, we removed flanking regions if there are strong recombination hotspots (*i.e.*, a region with $> 20 \times$ the background recombination rate). The estimates for the population recombination rate parameter for each region were obtained from [21] inferred from HapMap data; because the genomic segments examined did not contain strong hot spots, these estimates were assumed to apply uniformly in our simulations. The features and summary statistics of each genomic region are summarized in Figure S9. The mean of the distribution for μ was set at the estimate based on the number of fixed derived alleles in the 1000 Genome data identified using the alignment of human, chimpanzee, orangutan, and rhesus macaque, assuming a divergence time between human and chimpanzee of 5 million years ago and an ancestral population size of 12,500 diploid individuals [22-24]. The parameter values used in the simulations are given in Table S2.

For ancient DNA data, we first selected sequence data from ancient European samples [2, 25-27] that were dated within $10-7$ thousand years ago (KYA) and then counted the numbers of sequence reads carrying derived and ancestral alleles at each SNP site. We only included individuals who had, at least, 3 sequence reads, and the numbers used for the estimation of $f_{T_{\text{past}}}$ are as follows: 46 at rs16891982, 22 at rs1426654, 10 at rs642742, and 39 at rs1042602. We used $f_{T_{\text{past}}}$ to condition trajectories in the simulation under the SNM and SSV models.

REFERENCES

[1] Li, H. & Durbin, R. 2011 Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-496. (DOI:10.1038/nature10231nature10231 [pii]).

[2] Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. 2015 Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503. (DOI:10.1038/nature16152nature16152 [pii]).

[3] Kaplan, N. L., Hudson, R. R. & Langley, C. H. 1989 The "hitchhiking effect" revisited. *Genetics* **123**, 887-899.

[4] Kim, Y. & Nielsen, R. 2004 Linkage disequilibrium as a signature of selective sweeps. *Genetics* **167**, 1513-1524. (DOI:10.1534/genetics.103.025387167/3/1513 [pii]).

[5] Maynard Smith, J. & Haigh, J. 1974 The hitchhiking effect of a favorable gene. *Genetical Research* **23**, 23-35.

[6] Fay, J. C. & Wu, C. I. 2000 Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405-1413.

[7] Fu, Y. X. 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915-925.

[8] Tajima, F. 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.

[9] Watterson, G. A. 1975 On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7**, 256-276. (DOI:0040-5809(75)90020-9 [pii]).

[10] Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., et al. 2002 Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832-837. (DOI:10.1038/nature01140nature01140 [pii]).

[11] Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. 2006 A map of recent positive selection in the human genome. *PLoS Biol* **4**, e72. (DOI:05-PLBI-RA-1239R2 [pii]10.1371/journal.pbio.0040072).

[12] Nakagome, S. 2015 On the use of kernel approximate Bayesian computation to infer population history. *Genes Genet Syst* **90**, 153-162. (DOI:10.1266/ggs.90.153).

[13] Fukumizu, K., Song, L. & Gretton, A. 2013 Kernel Bayes' Rule: Bayesian Inference with Positive Definite Kernels *Journal of Machine Learning Research* **14**, 3753-3783.

[14] Nakagome, S., Fukumizu, K. & Mano, S. 2013 Kernel approximate Bayesian computation in population genetic inferences. *Stat Appl Genet Mol Biol* **12**, 667-678. (DOI:10.1515/sagmb-2012-0050

/j/sagmb.2013.12.issue-6/sagmb-2012-0050/sagmb-2012-0050.xml [pii]/j/sagmb.ahead-of-print/sagmb-2012-0050/sagmb-2012-0050.xml [pii]).

[15] Osada, N., Nakagome, S., Mano, S., Kameoka, Y., Takahashi, I. & Terao, K. 2013 Finding the factors of reduced genetic diversity on X chromosomes of Macaca fascicularis: male-driven evolution, demography, and natural selection. *Genetics* **195**, 1027-1035. (DOI:10.1534/genetics.113.156703genetics.113.156703 [pii]).

[16] Hofmann, T., Scholkopf, B. & Smola, A. J. 2008 Kernel methods in machine learning. *Ann Stat* **36**, 1171-1220. (DOI:Doi 10.1214/009053607000000677).

[17] Kanagawa, M. & Fukumizu, K. 2014 Recovering Distributions from Gaussian RKHS Embeddings. *Proceedings of the 17th International Conference on Articial Intelligence and Statistics (AISTATS) 2014* **33**, 457-465.

[18] Delaneau, O., Marchini, J. & Zagury, J. F. 2011 A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-181. (DOI:10.1038/nmeth.1785nmeth.1785 [pii]).

[19] Delaneau, O., Zagury, J. F. & Marchini, J. 2013 Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5-6. (DOI:10.1038/nmeth.2307nmeth.2307 [pii]).

[20] Marchini, J., Cutler, D., Patterson, N., Stephens, M., Eskin, E., Halperin, E., Lin, S., Qin, Z. S., Munro, H. M., Abecasis, G. R., et al. 2006 A comparison of phasing algorithms for trios and unrelated individuals. *Am J Hum Genet* **78**, 437-450. (DOI:S0002-9297(07)62383-0 [pii]10.1086/500808).

[21] Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. 2005 A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321-324. (DOI:310/5746/321 [pii]10.1126/science.1117196).

[22] Patterson, N., Richter, D. J., Gnerre, S., Lander, E. S. & Reich, D. 2006 Genetic evidence for complex speciation of humans and chimpanzees. *Nature* **441**, 1103-1108. (DOI:nature04789 [pii]10.1038/nature04789).

[23] Hobolth, A., Christensen, O. F., Mailund, T. & Schierup, M. H. 2007 Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet* **3**, e7. (DOI:06-PLGE-RA-0277R3 [pii]10.1371/journal.pgen.0030007).

[24] Scally, A., Dutheil, J. Y., Hillier, L. W., Jordan, G. E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T., et al. 2012 Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**, 169-175. (DOI:10.1038/nature10842nature10842 [pii]).

[25] Allentoft, M. E., Sikora, M., Sjogren, K. G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P. B., Schroeder, H., Ahlstrom, T., Vinner, L., et al. 2015 Population genomics of Bronze Age Eurasia. *Nature* **522**, 167-172. (DOI:10.1038/nature14507nature14507 [pii]).

[26] Lazaridis, I. & Patterson, N. & Mittnik, A. & Renaud, G. & Mallick, S. & Kirsanow, K. & Sudmant, P. H. & Schraiber, J. G. & Castellano, S. & Lipson, M., et al. 2014 Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413. (DOI:10.1038/nature13673nature13673 [pii]).

[27] Skoglund, P., Malmstrom, H., Omrak, A., Raghavan, M., Valdiosera, C., Gunther, T., Hall, P., Tambets, K., Parik, J., Sjogren, K. G., et al. 2014 Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344**, 747-750. (DOI:10.1126/science.1253448science.1253448 [pii]).